Advances in Intelligent Systems and Computing 451

Ajith Abraham
Sergey Kovalev
Valery Tarassov
Václav Snášel *Editors*

Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)

Volume 2

Springer

Advances in Intelligent Systems and Computing

Volume 451

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

More information about this series at http://www.springer.com/series/11156

Ajith Abraham · Sergey Kovalev
Valery Tarassov · Václav Snášel
Editors

Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)

Volume 2

Springer

Editors
Ajith Abraham
Scientific Network for Innovation and Research Excellence
Machine Intelligence Research Labs
Auburn, WA
USA

Sergey Kovalev
Rostov State Transport University
Rostov-on-Don
Russia

Valery Tarassov
Bauman Moscow State Technical University
Moscow
Russia

Václav Snášel
VŠB-Technical University of Ostrava
Ostrava
Czech Republic

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-33815-6 ISBN 978-3-319-33816-3 (eBook)
DOI 10.1007/978-3-319-33816-3

Library of Congress Control Number: 2016937346

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

This volume of Advances in Intelligent Systems and Computing contains papers presented in the main track of IITI 2016, the First International Conference on Intelligent Information Technologies for Industry held in May 16–21 in Sochi, Russia. Sochi is a beautiful city in the south of Russia, a well-known resort center on the Black Sea and the capital of XXII Olympic Winter Games—2014. The conference was jointly co-organized by Rostov State Transport University (Russia) and VŠB-Technical University of Ostrava (Czech Republic) with the participation of Russian Association for Artificial Intelligence (RAAI) and Russian Association for Fuzzy Systems and Soft Computing (RAFSSC). These two Associations represent important Russian scientific communities incorporated into international activities: RAAI is a member society of the European Coordinating Committee for Artificial Intelligence (ECCAI) and RAFSSC is the member of International Fuzzy Systems Association (IFSA).

The First IITI Conference is not, of course, the first international conference on intelligent technologies and fuzzy systems in Russia. Among the most important international events (in English) organized by RAAI and RAFSSC in twenty-first century we may cite the Joint Conference on Knowledge-Based Software Engineering (KBSE), International Conference on Fuzzy Sets and Soft Computing in Economics and Finance (FSSCEF), Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC'2011). The XVth Russian National Conference in Artificial Intelligence with international participation (CAI-2016) will take place at Smolensk in October 2016.

The conferences above-mentioned were mainly focused on fundamental problems of artificial intelligence, fuzzy set theory, and soft computing. To differ from them, IITI'16 is devoted to practical models and industrial applications related to intelligent information systems. It is considered as a meeting point for researchers and practitioners to enable the implementation of advanced information technologies into various industries. Nevertheless, some theoretical talks concerning the state of the art in intelligent systems and soft computing were also included into proceedings.

There were 162 paper submissions from 10 countries. Each submission was reviewed by at least three Chairs or PC members. We accepted 92 regular papers (57 %). Unfortunately, due to limitations of conference topics and edited volumes, the Program Committee was forced to reject some interesting papers which did not satisfy these topics or publisher requirements. We would like to thank all authors and reviewers for their work and valuable contributions. The friendly and welcoming attitude of conference supporters and contributors made this event a success!

May 2016

Ajith Abraham
Sergey Kovalev
Valery Tarassov
Václav Snášel

Organization

Conference Chairs

Sergey Kovalev, Rostov State Transport University, Russia
Václav Snášel, VŠB-TU Ostrava, Czech Republic
Ajith Abraham, MIR Labs, USA

Conference Vice-Chair

Valery Tarassov, Bauman Moscow State Technical University, Russia

International Program Committee

Alexey N. Averkin, Dorodnicyn Computing Centre of Russian Academy of Sciences, Russia
Rafik A. Aliyev, Azerbaijan State Oil Academy, Azerbaijan
Mikołaj Bartłomiejczyk, Gdansk University of Technology, Poland
Ildar Z. Batyrshin, National Polytechnic Institute, Mexico
Anton Beláň, Slovak University of Technology in Bratislava, Slovakia
Arcady N. Borisov, Riga Technical University, Latvia
Jiří Bouchala, VŠB-TU Ostrava, Czech Republic, Czech Republic
Milan Dado, University of Žilina, Slovakia
Igor D. Dolgiy, Rostov State Transport University, Russia
Alexander P. Eremeev, Moscow Power Engineering Institute, Russia
Alena V. Fedotova, Bauman Moscow State Technical University, Russia
Bronislav Firago, Belarusian National Technical University, Belarus
Igor B. Fominykh, Moscow Power Engineering Institute, Russia
Vladimir V. Golenkov, Belarus State University of Informatics and Radioelectronics, Belarus
Jiří Hammerbauer, University of West Bohemia in Pilsen, Czech Republic

Stanislav Kocman, VŠB-TU Ostrava, Czech Republic, Czech Republic
Jaroslav Kultan, University of Economics in Bratislava, Slovakia
Viktor M. Kureychik, Southern Federal University, Russia
Igor Kurytnik, University of Bielsko-Biała, Poland
Oleg P. Kuznetsov, Institute of Control Sciences of Russian Academy of Sciences, Russia
Teresa Orłowska-Kowalska, Wrocław University of Technology, Poland
Gennady S. Osipov, Institute for Systems Analysis of Russian Academy of Sciences, Russia
Leszek Pawlaczyk, Wrocław University of Technology, Poland
Josef Paleček, VŠB-TU Ostrava, Czech Republic, Czech Republic
Zdenék Peroutka, University of West Bohemia in Pilsen, Czech Republic
Alexey B. Petrovsky, Institute for Systems Analysis of Russian Academy of Sciences, Russia
Stanislav Rusek, VŠB-TU Ostrava, Czech Republic, Czech Republic
Alexander V. Smirnov, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Russia
Vadim L. Stefanuk, Institute for Information Transmission of Russian Academy of Sciences, Russia
Vadim N. Vagin, Moscow Power Engineering Institute, Russia
Feodor Vainstein, Texas A&M University, USA
Yuri R. Valkman, International R&E Center of Information Technologies and Systems in Kiev, Ukraine
Vladimír Vašinek, VŠB-TU Ostrava, Czech Republic, Czech Republic
Ján Vittek, University of Žilina, Slovakia
Nadezhda G. Yarushkina, Ulyanovsk State Technical University, Russia
Alla V. Zaboleyeva-Zotova, Volgograd State Technical University, Russia

Organizing Committee Chair

Alexander N. Guda, Rostov State Transport University, Russia

Local Organizing Committee

Maria A. Butakova, Rostov State Transport University, Russia
Andrey V. Chernov, Rostov State Transport University, Russia
Vladislav S. Kovalev, JSC “NIIAS”, Russia
Pavel Krömer, VŠB-TU Ostrava, Czech Republic, Czech Republic
Vladimir V. Kureychik, Southern Federal University, Russia
Jan Platoš, VŠB-TU Ostrava, Czech Republic, Czech Republic
Vladimir L. Samsonov, Rostov State Transport University, Russia

Vitezslav Styskala, VŠB-TU Ostrava, Czech Republic, Czech Republic
Andrey V. Sukhanov, Rostov State Transport University, Russia
Ivan A. Yaitskov, Rostov State Transport University, Russia

Sponsoring Institutions

Rostov State Transport University, Russia
VŠB-Technical University of Ostrava, Czech Republic

Contents

Part VII Applied Systems

Part I
Cognitive Technologies on the Basis of Sensor and Neural Networks

The Data Transfer Development in MANET Networks on the Base of Chinese Remainder Theorem

Nikolay Ivanovich Chervyakov, Mikhail Grigorevich Babenko, Nikolay Nikolaevich Kucherov, Irina Sergeevna Krisina, Anastasiia Igorevna Garianina and Mariia Nikolaevna Shabalina

Abstract In the paper the data transmission scheme in MANET networks constructed on the threshold secret sharing scheme (SSS) possessing the computation capability similar to Asmuth-Blum SSS, however code with the speed of t times more is proposed. The proposed SSS constructed on residue number system (RNS) allows to detect errors in communication channels using the error correction codes in RNS. From results of simulation it is possible to draw the conclusion that in case of identical occupied space the time delay of the approximate method when using modules of size to 64 bits is 20 % less, and when using modules of size more than 64 bits the time delay is 30 % less in comparison with method of orthogonal bases. The use of SSS constructed on RNS allows to develop reliable data transfer protocols in MANET networks. The neural network implementation of the approximate method is 10 % more effective than classical implementation of the approximate method.

Keywords MANET · Secret sharing scheme · Residue number system · Error control codes · Chinese Remainder Theorem

N.I. Chervyakov (✉) · M.G. Babenko · N.N. Kucherov · I.S. Krisina · A.I. Garianina · M.N. Shabalina
North Caucasian Federal University NCFU, Stavropol, Russia
e-mail: k-fmf-primath@stavsu.ru; whbear@yandex.ru

M.G. Babenko
e-mail: mgbabenko@ncfu.ru

N.N. Kucherov
e-mail: nkucherov@ncfu.ru

I.S. Krisina
e-mail: irishechka.26@mail.ru

A.I. Garianina
e-mail: xrizoberil_09@mail.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_1

1 Introduction

Mobile MANET networks found broad application in information transfer infrastructure: during military operations for transmission of cards, satellite pictures and video instructions [1, 2], in emergency situations in case of video information transmission on a place of the tragedy and scales of corruptings [1, 3], on the city broadband wireless networks [4, 5], etc., satisfying to video information transmission quality provided by the DVB (Digital Video Broadcasting) standard [6]. However in MANET networks the probability of failures connected to their mobile nature is great. Failures appear as at any time any node can become unavailable or appear in a network [7–10].

An increase of devices computational capability according to Moores law, increase of the software requirements to computational capability of equipment according to Virt's law and increase in signal power for fall forward of the transmitted data on the wireless networks leads to bigger consumption of electrical energy by mobile devices, that leads to increase in battery capacity and reduction of time of independent operation. Use of one-time batteries allows to save energy on the long period, and use of accumulator leads to loss 90 % of the electric power within 100 days of their non-use [2] therefore it is inexpedient to use accumulator for production of the specialized rare application equipment.

Owing to everything said above, limited technical characteristics of devices, such as the accumulator capacity, computational capability, add vulnerability to devices and a network in general. It is proposed to use the threshold secret sharing schemes based on Chinese Remainder Theorem for providing the reliable and secure data transfer. This choice is made due to the fact that modular arithmetics allow to implement the effective algorithms of data protection, occupying small space, providing low energy consumption and high fault-tolerance. Also use of modular arithmetics guarantees transmission of the message in a case when the part of information is lost or distorted. Thus confidentiality, integrity and authenticity of the information transferred in mobile MANET networks is provided.

The main aim is the development of information transmission system satisfying to the following two conditions [7]:

1. Each node may perform necessary operations with the data for its secure transmission, but access to the information is provided only for those nodes to which it is intended.
2. In case of failure of one or more nodes, the message must be delivered without distortion with a probability close to one, in the conditions of the transfer in a hostile environment.

2 Mathematical Model of MANET Network

The multistep network is represented with the directed weighed multigraph $G(V, E)$ with the painted arches. A set of graph nodes V corresponds to the set of routers and a set of arcs $E \subseteq \{e_{ijk} : i,j \in V, k = 1,2,\ldots\}$—to the set of network links, and the color of an arch designated by an integer k corresponds to the frequency of channel used in a network. If the distance between nodes v_{ik} and v_{jk}, $i \neq j$ and $i, j \in Z_N$ is less than communication range, between them a communication channel exists, and nodes v_{ik}, v_{jk}, are called neighbors. Between two routers there can be some links (in different frequency channels) therefore the neighbor nodes in the graph can be connected with more than one arch. Each arc $e_{ijk} \in E$ the numbers is put in compliance: $(D_{ij}, q_{ijk}, d_{ik}, Cost_{ik}, n_{ik}, f_{ik})$.

1. distance D_{ij} between routers i and j;
2. probability $q_{ijk}(D_{ij})$ that attempt of a package transfer between routers i and j in the frequency channel k will be unsuccessful;
3. average delay d_{ik} of the first attempt of a package transfer the router i in the frequency channel k;
4. the cost $Cost_{ik}$ of one attempt of a package transfer the router i in the frequency channel k. We will notice that cost c_{ik} linearly depends on length of the message transferred in a package (that is if the message the length l bit transferred in a package by router i in frequency channel k costs c_{ik}, then the message the length rl bit transferred by router i in frequency channel k costs rc_{ik});
5. the number n_{ik} of repetitions of a package transfer the router i in the frequency channel k providing an appropriate level of reliability at information transfer;
6. the number of neighbours f_{ik} of the node v_{ik} in the frequency channel k (as in works [7, 10–16]).

The way $R(I, K)$ of an information package transferred from node $i_I = i_1$ to node $i_O = i_s$ lies through vertices $i_2, \ldots, i_{s-1}$, on communication links $k_1, \ldots, k_{s-1}$ respectively, and the cost of the package delivery on the way $R(I, K)$ equals

$$Car(i_l, i_O, R(I,K)) = \sum_{j=1}^{s-1} n_{i_j} \cdot Cost_{i_j}$$

The probability of the package delivery from node i_I to node i_O on the way $R(I, K)$ is:

$$P(i_I, i_O, R(I,K)) = \prod_{j=1}^{s-1} \left(1 - q_{i_j i_{j+1} k_{i_j}}^{n_j k_{i_j}}\right)$$

The average time of a time delay in case of information transfer from the node i_I to the node i_O along the route $R(I, K)$ is:

$$D(i_I, i_O, R(I, K)) = \sum_{j=1}^{s-1} d_{i_j}$$

For delivery of a package from a node i_I to a node i_O the route with the minimum cost and the maximum level of reliability is selected.

Let the maximum message, which can be transferred in a package, have length of L bits. For the tL bit length message S transmission in MANET networks we use the Secret Sharing Scheme (SSS) (t, r) constructed with use of the Residue Number System (RNS). We choose as the modules of RNS m_i pairwise coprimes satisfying a condition:

$$2^L < m_1 < m_2 < \cdots < m_r < 2^{L+1} \tag{1}$$

then $2^{tL} < \prod_{i=1}^{t} m_i$, therefore according to the Chinnese Remainder Theorem each number in the range $[0, 2^{tL}]$ can be represented in RNS unambiguously. We will calculate message S projections, using the following formula:

$$S_i = \bmod m_i \tag{2}$$

Knowing any t message projections, the receiver, using the Chinese Remainder Theorem, will be able to recover the message. If the receiver knows projections $(h > t)$, then using codes of detection and correction of errors in RNS, the receiver can find $h - t$ errors and correct the $h - t - 1$ errors. If the eavesdroper received $h(h < t)$ secret $S_{i_1}, S_{i_2}, \ldots, S_{i_h}$ projections, the power of a set of the secret S possible values is $|S| = \left\lceil \frac{2^{tL}}{\prod_{j=1}^{h} m_{i_j}} \right\rceil$. At worst, when the eavesdroper knows $h = t - 1$ secret projections, value of power of this set satisfies to an inequality:

$$2^{L-t+1} < |S| < 2^L$$

In case of a choice $L \geq t$ the secret sharing scheme constructed with use of RNS is threshold, and the set of possible values of a secret is more than 2 equally probable values of a secret.

Telling each projection S_i on the way $R_i(I_i, K_i)$ so that for any $i \neq j$ the condition $I_i \cap I_j = \{i_I, i_O\}$ is satisfied, that is $R_i(I_i, K_i)$ and $R_j(I_j, K_j)$—not crossed on peaks two simple ways, we will estimate the cost, probability and average time of a time delay of the message delivery:

$$Car(i_I, i_I) = \min_{I \in J}\left(\sum_{j \in I} Car\left(i_l, i_O, R_j\left(I_j, K_j\right)\right)\right)$$

$$p(i_I, i_I) = \sum_{I \in J} \prod_{j \in I} P\left(i_I, i_O, R_j\left(I_j, K_j\right)\right) \times \prod_{j \notin I}\left(1 - P\left(i_I, i_O, R_j\left(I_j, K_j\right)\right)\right)$$

$$D(i_I, i_I) = \min_{I \in J}\left(\max_{j \in I}\left(D\left(i_l, i_O, R_j\left(I_j, K_j\right)\right)\right)\right)$$

where I—a set of the numbers of the allowed set of participants, J—set of all possible allowed great number of participants for recovery of the message. The allowed great number of participants is meant as the number of participants sufficient for recovery of the message. In our case this is any set containing t or more secret projections.

3 The Method of Error Detection

3.1 The Error Detection Method Based on the Chinese Remainder Theorem

We will consider system with the bases $m_{i_1}, m_{i_2}, \ldots, m_{i_t}$ and the range $M = \prod_{j=1}^{t} m_{i_j}$. We will introduce the base $m_{i_{t+1}}$ coprime with any base m_{i_j} (on SSS condition). Also we will represent numbers from $i+1$ bases. It means that we will transfer numbers and we will make operations over the numbers lying in the range $[0, M)$, in broader range $[0, M')$ where $M' = \prod_{j=1}^{t+1} m_{i_j}$.

Theorem 1 *[11, 16] Let the RNS bases $m_{i_1}, m_{i_2}, \ldots, m_{i_t}, m_{i_{t+1}}$ satisfy a condition $m_{i_j} < m_{i_{j+1}}, j = 1, 2, \ldots, t$, and let $A = \left(a_{i_1}, a_{i_2}, \ldots, a_{i_j}, \ldots, a_{i_t}, a_{i_{t+1}}\right)$ be a correct number.*

Then number $\widetilde{A} = \left(a_{i_1}, a_{i_2}, \ldots, \tilde{a}_{i_j} \neq a_{i_j}, \ldots, a_{i_t}, a_{i_{t+1}}\right)$, where $j = 1, 2, \ldots, t, t+1$ is incorrect.

Follows from the theorem $\widetilde{A} > \frac{M}{m_{i_{t+1}}}$ takes place, i.e. $\widetilde{A}$ is the wrong number. Its worth to note that if among the bases of system there are such small bases $m_{i_1}, m_{i_2}, \ldots, m_{i_w}$, that

$$\prod_{j=1}^{w} m_{i_j} = M'' < m_{i_j}.$$

Then distortions in digits on several or even on all bases turn the correct number in wrong, and existence of distortions in all cases can be set.

To find existence or absence of an error in number A, it is necessary to compare it to the range M. Thus if $A \geq M$ appeared, that means the error at least in one digit takes place, if $A < M$ either there is no error or it has more difficult nature [16].

3.2 Modification of the Error Detection Method on the Basis of the Approximate Method

We will consider the method of number errors and number of the faulty residue number system channels determination which possesses high-speed performance and low hardware expenses from work [13].

Besides, the solution practically of any control task requires comparing in a necessary timepoint of a status of managed objects with the given statuses corresponding to algorithm of systems functioning. The purpose of comparing is detection of the fact of coincidence or mismatch of values, equalities or inequalities of numbers greater or smaller than some values.

We will apply the approximate method [13] which allows to implement correctly mentioned above functions to determinate dynamic range overflow and detect an error.

The essence of the approximate method is based on use of the relative value of numbers to the full range determined by the Chinese Remainder Theorem, which connects positional number A to its representation in residuals $(\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_t})$, where α_{i_j}—the smallest non-negative residues of number, concerning residue number system modules $m_{i_1}, m_{i_2}, \ldots, m_{i_t}$ the following expression

$$A = \left| \sum_{j=1}^{t} \frac{M}{m_{i_j}} \left| M_{i_j}^{-1} \right|_{m_{i_j}} \alpha_{i_j} \right|_M$$

where $M = \prod_{j=1}^{t} m_{i_j}$, m_{i_j}—RNS modules, $\left| M_{i_j}^{-1} \right|$—multiplicative inversion M_{i_j} relatively to m_{i_j}, and $M_{i_j} = \frac{M}{m_{i_j}} = m_{i_1}, m_{i_2}, \ldots, m_{i_{j-1}}, m_{i_{j+1}}, \ldots, m_{i_t}$.

If we divide the left and right part of expression (1) by the constant P corresponding to the range of numbers, we will receive approximate value

$$\frac{A}{M} = \left| \sum_{j=1}^{t} \frac{\left| M_{i_j}^{-1} \right|_{m_{i_j}}}{m_{i_j}} \alpha_{i_j} \right|_1 \approx \left| \sum_{j=1}^{t} k_{i_j} \alpha_{i_j} \right|_1$$

where $k_{i_j} = \frac{\left|M_{i_j}^{-1}\right|_{m_{i_j}}}{m_{i_j}}$—constants of the chosen system, α_{i_j}—digits of the number represented in RNS, thus value of each sum will be in an interval $[0, 1)$. The end result of the sum is being calculated after summing and discarding of an integer part of number on saving a fractional part of the sum. The fractional part can be written also as $A \bmod 1$, because $A = \lfloor A \rfloor + A \bmod 1$. The quantity of the number fractional part discharges is defined by the greatest possible difference between adjacent numbers. In need of exact comparing it is necessary to calculate value (2) which is equivalent of RNS to weighted number system conversion. For the solution of an objective task it is enough to know approximately the value of the used number A in relation to dynamic range M which is executed rather simply, but thus the ratio is defined truly $A = M$, $A > M$ or $A < M$ [13].

For determination of an error and overflow redundant RNS is used, possess t-work $r - t$-control bases, thus the excess range in case of two excess modules will equal $M_{over} = M \cdot m_{i_{t+1}} \cdot m_{i_{t+2}}$, where M working range. If values $\frac{A}{M_{over}} < \frac{M}{M_{over}}$, then number if faultless. Similarly also overflow of the range of the represented numbers is defined.

Finite process of dynamic range overflow detection and detection errors can be presented in the form of the generalized algorithm [12]:

The relative values $\frac{\overline{A}}{M_{over}} < \frac{\left|M_{i_j}^{-1}\right|}{m_{i_j}} \cdot a_{i_j}$ and $\frac{M}{M_{over}} < \frac{\left|M_{i_j}^{-1}\right|}{m_{i_j}} \cdot m_{i_j}$ are calculated, where: $\overline{A}$—fault number; α_{i_j}—digits of number $\overline{A}$; m_{i_j}—digits of number M. If $\frac{\overline{A}}{M_{over}} < \frac{M}{M_{over}}$, then there is no error, if $\frac{\overline{A}}{M_{over}} \geq \frac{M}{M_{over}}$ there is an error and dynamic range overflow.

4 The Effective Neural Network SSS Implementation

4.1 Finite Ring Neural Network

The Neural Network (NN) represents high-parallel dynamic system with topology of directed graph which can receive output information by means of response of its status to input influences. In NN processor elements and directed graphs are called as nodes [15, 17]. The structure of handling algorithm of data provided in residue number system, also as well as structure of NN possesses natural parallelism that allows to use NN as the formal device of the algorithm description.

NN architecture. The general interpretation of NN architecture is a mass and parallel interdependent network of simple elements and its hierarchical arrangement. The structure of NN somewhat models biological nervous system.

Power of NN is its ability to use the initial knowledge base for the solution of the existing problem. All neurons work is competitive, and direct computation is influenced by the knowledge ciphered in communications of a network [17–20].

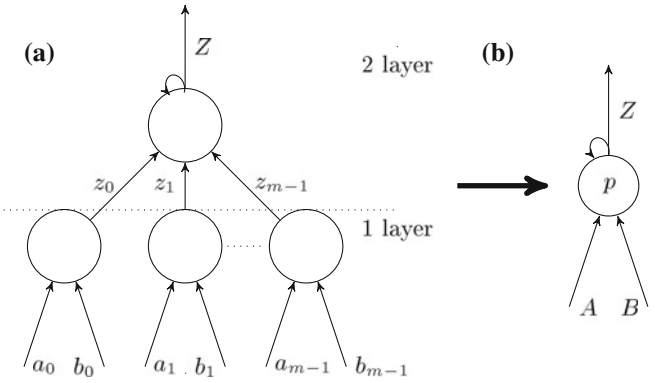


Fig. 1 Structure of the subnet (**a**) and its symbolic representation (**b**)

Interaction of neurons is considered in three-level hierarchy of the network consisting of layers [17, 19, 21]:

1. Display of parameter (this layer contains the residual connected to the weighed value of each computing discharge).
2. Display of bit computation (defines the function of a finite ring applied to each computing discharge).
3. Display of operation of a finite ring (defines the main operations which are used for arithmetics of a finite ring implementation).

We will consider architecture of FRNN from work [22]. Based on a computational model of a finite ring the main operator in which is the operator of extraction of separate discharges of the transformed number binary representation multi-layer subnets can be constructed. Structure of the subnet is shown in Fig. 1, where synoptic weights are $w_i = |2^i|$, $i = 0, 1, \ldots, n-1$.

1. Assembly (it is used for collection of the inputs belonging to one binary place of input sources). The result of binary to residual conversions, multiplication, addition operations calculated by means of FRNN is function of the amount of the weighed input discharges.
2. The computing. The result of computation is defined by the positive logic. The end result of FRNN will have the steady form.

4.2 Modeling

Simulation of the one FPGA higher described methods of errors detection, for which size of modules is carried out: size of 8, 16, 32, 64, 128, 192 and 256 bits. On an input in all methods the RNS bases m_i were moved and number

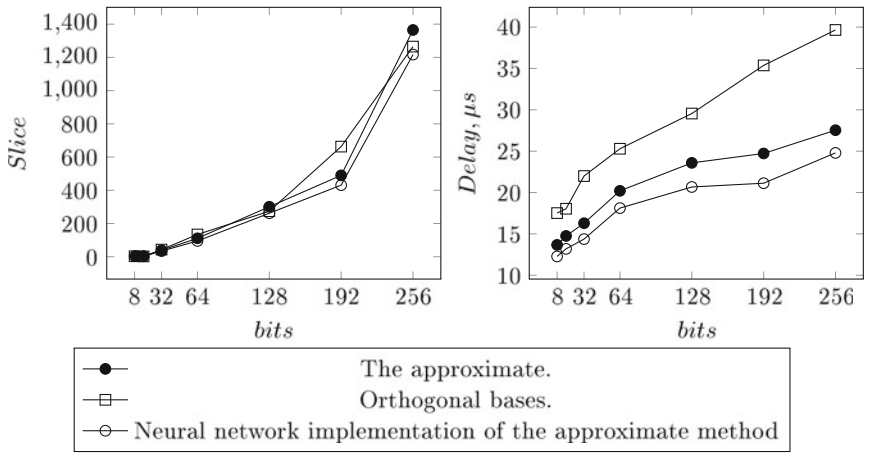


Fig. 2 Diagram of use of slice and temporal time delay units

$A = (\alpha_1, \alpha_2, \ldots, \alpha_t)$ which size changed in proportion to modules m_i. In a Fig. 2 results of simulation are shown.

From the results of simulation provided in Fig. 2 it is possible to draw the conclusion that in case of rather identical occupied space the time delay of the approximate method when using modules of size less or equal to 64 bits is 20 % less, and when using modules of size more than 64 bits the time delay is 30 % less in comparison to method of orthogonal bases. The neural network implementation of the approximate method is 10 % more effective than classical implementation of the approximate method.

5 Conclusion

In this paper the threshold secret sharing scheme (SSS) possessing the computation capability similar to SSS of Asmuth-Blum [23] is proposed, however the speed of proposed code scheme is t times more. The proposed SSS constructed on RNS allows to detect errors in communication links using RNS error correction codes. From the results of simulation provided in Fig. 2 it is possible to draw the conclusion that in case of rather identical occupied space the temporal time delay of the approximate method when using modules of size less or equal to 64 bits is 20 % less, and when using modules of size more than 64 bits the time delay is 30 % less in comparison to method of orthogonal bases. Use of SSS constructed on RNS allows to construct reliable transfer protocols of video data in the MANET networks. The neural network implementation of the approximate method is 10 % more effective than classical implementation of the approximate method.

Acknowledgment Current work was performed as a part of the State Assignment of Ministry of Education and Science (Russia) No. 2563.

References

1. Safonov, A., Lyakhov, A.I., Yurgenson, A., Sokolova, O.D.: Groupcast routing with the possibility of choosing a channel transmission method. Autom. Remote Control **74**(10), 1710–1723 (2013)
2. Szczodrak, M., Kim, J.: 4 g and manet, wireless network of future battlefield. In: Security and Management, pp. 282–290 (2007)
3. Andrews, J., Shakkottai, S., Heath, R., Jindal, N., Haenggi, M., Berry, R., Guo, D., Neely, M., Weber, S., Jafar, S., et al.: Rethinking information theory for mobile ad hoc networks. Commun. Mag. IEEE **46**(12), 94–101 (2008)
4. Kiryanov, A., Lyakhov, A., Nekrasov, P., Platov, D., Safonov, A., Feizkhanov, R., Khorov, E., Tsyganova, A.: Proximity-based groupcast in manet (gim). J. Commun. Technol. Electron. **57**(12), 1303–1313 (2012)
5. Vishnevskiy, V., Lyakhov, A., Portnoy, S., Shakhnovich, I.: Broadband wireless networks of information transfer. Technosphere (2005)
6. Reimers, U.: DVB: The Family of International Standards for Digital Video Broadcasting. Springer (2013)
7. Arbaugh, W.A.: Wireless security is different. Computer **36**(8), 99–101 (2003)
8. Mishra, A., Nadkarni, K.M.: Security in wireless ad hoc networks. In: The Handbook of Ad Hoc Wireless Networks, pp. 499–549. CRC Press Inc. (2003)
9. Papadimitratos, P., Haas, Z.J.: Securing mobile ad hoc networks. In: The Handbook of Ad Hoc Wireless Networks (2002)
10. Yang, H., Luo, H., Ye, F., Lu, S., Zhang, L.: Security in mobile ad hoc networks: challenges and solutions. Wireless Commun. IEEE **11**(1), 38–47 (2004)
11. Akushskii, I.J., Burcev, V.M., Pak, I.T.: A new positional characteristic of non-positional codes and its application. In: Amerbaev, V.M. (ed.) Coding Theory and the Optimization of Complex Systems. Nauka, Alma-Ata, Kazakh SSR (1977) (in Russian)
12. Chervyakov, N.I., Babenko, M.G., Deryabin, M.A., Kucherov, N.N., Kuchukova, N.N.: The ec sequences on points of an elliptic curve realization using neural networks. In: Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015. pp. 147–154. Springer (2016)
13. Chervyakov, N.I., Babenko, M.G., Kucherov, N.N., Kuchukov, V.A., Shabalina, M.N.: Researches of algorithm of prng on the basis of bilinear pairing on points of an elliptic curve with use of a neural network. In: Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015. pp. 167–173. Springer (2016)
14. Ertaul, L., Chavan, N.: Elliptic curve cryptography based threshold cryptography (ecc-tc) implementation for manets. IJCSNS **7**(4), 48 (2007)
15. Galushkin, A.I.: Theory of neural networks. M: IPRRGR (2000)
16. Omondi, A., Premkumar, B.: Residue Number Systems. World Scientific (2007)
17. Chervyakov, N., Sakhnyuk, P., Shaposhnikov, V., Makoha, A.: Neurocomputers in residue numbers. Radiotechnique (2003)
18. Chervyakov, N.I., Babenko, M.G., Kucherov, N.N., Garianina, A.I.: The effective neural network implementation of the secret sharing scheme with the use of matrix projections on fpga. In: Advances in Swarm and Computational Intelligence, pp. 3–10. Springer (2015)
19. Zhang, C.N., Shirazi, B., Yun, D.Y.: Parallel designs for chinese remainder conversion. In: ICPP, pp. 557–559 (1987)

20. Zhang, D., Jullien, G., Miller, W.: Vlsi implementations of neural-like networks for finite ring computations. In: Proceedings of the 32nd Midwest Symposium on Circuits and Systems, pp. 485–488. IEEE (1989)
21. Zhang, D., Jullien, G., Miller, W.: A neural-like network approach to finite ring computations. Circuits Syst. IEEE Trans. **37**(8), 1048–1052 (1990)
22. Chervyakov, N.I., Babenko, M.G., Krisina, I.S., Nazarov, A.S., Garianina, A.I.: Development of the protected data transfer protocol for the manet networks on the basis of residue number system. In: 2015 International Siberian Conference on Control and Communications (SIBCON), pp. 1–5. IEEE (2015)
23. Asmuth, C., Bloom, J.: A modular approach to key safeguarding. IEEE Trans. Inf. Theory **30** (2), 208–210 (1983)

Using of a Convolutional Neural Network with Changing Receptive Fields in the Tasks of Image Recognition

R.M. Nemkov, O.S. Mezentseva and D. Mezentsev

Abstract In the article synthesis procedure of mathematical model features of convolution neural network (CNN) is described. In order to improve the generalization capability of the network the training set is generated by adding a distorted image with changing of CNN receptive fields. This fact differs given procedure from the known procedures. We propose the reduction algorithm of an extended training set and the synthesis algorithm of features for CNN with non-standard receptive fields. The experiments results of the developed algorithms were shown in the article in order to assess of generalization capability changes of the convolution neural network. The experiments were performed with the hardware platform of the "Mechatronics" stand (SPA "Android techniques", Russia).

Keywords Convolutional neural network · Pattern recognition · Training with noise

1 Introduction

For invariant pattern recognition there are various methods: potential function, Bayesian networks, Markov networks, artificial neural networks, different types of associative memory and others. Problem of invariant images recognition is still not solved in general [3, 4].

Nowadays, the subclass of artificial neural networks (ANN) also known as convolution neural network (CNN) dominates in image recognition [1]. CNN

R.M. Nemkov (✉) · O.S. Mezentseva · D. Mezentsev
Department of Information Systems & Technologies, North-Caucasus Federal University, 2, Kulakov Prospect, Stavropol, Russian Federation
e-mail: nemkov.roman@yandex.ru

O.S. Mezentseva
e-mail: omezentceva@ncfu.ru

D. Mezentsev
e-mail: terrakoto@me.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_2

shows results more better than conventional ANN at 10–15 % [2]. However, training of ANN is an incorrect inverse problem [3]. Incorrect problem means that even large data set can't give us full information content about current task. Therefore in the synthesis of ANN mathematical model parameters key part belongs to training data. Making representative training data is one of the most difficult problems in machine learning [3].

The complex results of theoretical and experimental researches of a new type of CNN with changing receptive fields (RP) were presented in the article. With a help of such networks we can create distortions in relation to the current pattern (through the changing of pattern perception) by changing the internal parameters of the hierarchical model of ANN (receptive fields), thereby we obtain new patterns and expanding the training set [6, 7, 9].

2 Mathematical Modeling of CNN with Extended Training Set

Well-known generic synthesis parameters algorithm for CNN includes the following steps:

1. Forward propagation implementation, i.e. calculating signals propagation from the input to the output layer. For convolutional layer neurons values were obtained by the formula (1)

$$y_{m,n} = C^i_{m,n} = \varphi(p) = \varphi\left(b + \sum_{q \in Q_i} \sum_{k=0}^{K_C-1} \sum_{l=0}^{K_C-1} X^q_{m+k,n+l} * Z^q_{k,l}\right), \tag{1}$$

   where $C^i_{m,n}$—neuron output located at the ith card of C-layer in m, n, $\varphi(\cdot) = A * tanh(B * p)$—position where A = 1.7159, B = 2/3, p—weighted sum, b—bias, Q_i—set of indices cards of previous layer associated with the card C^i, K_C—size of square RP for the neuron $C^i_{m,n}$, $Z^q_{k,l}$—qth part of the adjustable features, which is responsible for interaction with qth map of previous layer.
2. Back propagation implementation, i.e. calculating signals propagation from the output layer to the input. For convolutional layer (2, 3)

$$\delta^{\lambda}_{m,n} = \sum_{i \in D} \delta^{\lambda+1}_i * w^{\lambda+1}_i[m,n] * \varphi'(p^{\lambda}_{m,n}), \tag{2}$$

   where D—is the set of neurons on subsequent map ($\lambda + 1$ layer) associated with neuron n, m, $w^{\lambda+1}_i[m,n]$—$u^{\lambda+1}_i$ index of the S-layer card, which is connected with the card of C-layer, $\varphi'(\cdot)$ is a derivative of $\varphi(\cdot)$, $\delta^i_{m,n}$—a residual was gathered for a neuron with coordinates m, n within the map of layer λ.

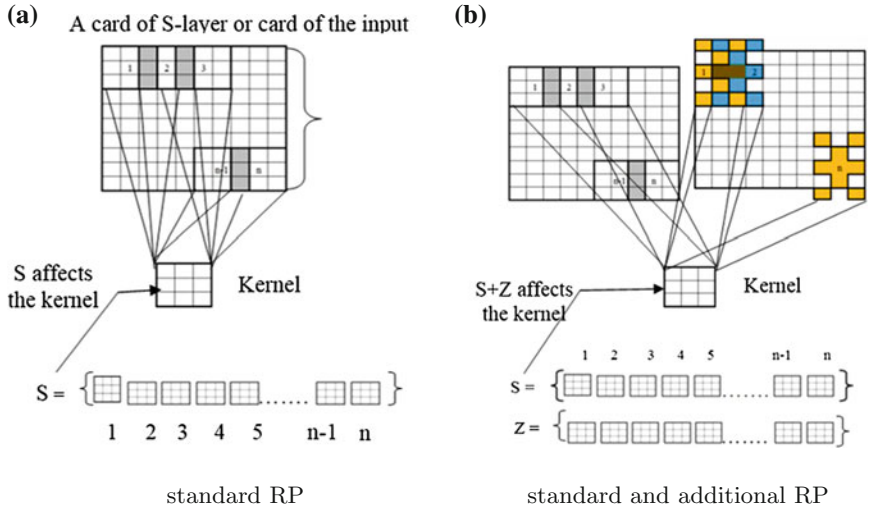


Fig. 1 **a** Standard RP, **b** standard and additional RP

$$\frac{\partial E}{\partial (Z_{k,l}^{\lambda})^{q}} = \sum_{m=0}^{SizeC} \sum_{n=0}^{SizeC} \delta_{m,n}^{\lambda} * y_{m+k,n+l}^{\lambda-1}, \tag{3}$$

where q—the part of the kernel of configurable features for which we receive the components of the gradient, SizeC—the size of the C-layer card.

The same pattern can be perceived by the network in different ways if for some layers of CNN we change set of RP. Due to this, we can extend the training set. The algorithm, which changes RP of neurons lying on any combination of S-layers before feeding pattern, was developed.

It is known, that the canonical form of RP in CNN is a square [11]. We offer to use a template for obtaining a non-standard RP, which elements are indices indicating their neighbors within two discrete steps from them on matrix of pixels. If you change all RP which lying on a map, a lot of information will impact on the configurable parameters that will lead to obtaining a better invariant (Fig. 1).

The algorithm which changes RP neurons lying on any combination of C-layers before feeding pattern is offered.

Each C-layer can have several types of RP: $SetRP_i \in \{RP_1, RP_2, \ldots, RP_n\}$, where $SetRP_i$—many types of RP for the ith C-layer, RP_1—a square RP, n_i—the quantity of RP types for the ith C-layer.

We offer to use two algorithms in order to assign neurons RP for ith C-layer. With the help of algorithm $Alg_1(RP_t \in SetRP_i; C_i)$ to all neurons of layer we assign C_i to RP with index RP_t, t = 1…T—measure number, $T = n_i$. With the help of algorithm $Alg_2(RP \in SetRP_i; C_i)$ to each neuron of layer we assign C_i random field RP of $SetRP_i$. These algorithms can ensure invariance to affine transformations.

Invariance to scale, texture background, a position of an object, luminance levels is provided with the original training sample.

Marking algorithms of a layer can be used by marking algorithms of all C-layers because of any C-layers may have their RP. Here suggests two algorithms:

$$Strategy_1(Alg_1(\cdot), Q_j) = \begin{cases} Alg_1(RP_1; C_i), & if\ x_i = 0 \\ Alg_1(RP_t; C_i), & if\ x_i = 1 \end{cases}, \tag{4}$$

$$Strategy_2(Alg_1(\cdot), Alg_2(\cdot), Q_j) = \begin{cases} Alg_1(RP_1; C_i), & if\ x_i = 0 \\ Alg_2(RP \in SetRP_i; C_i), & if\ x_i = 1 \end{cases}, \tag{5}$$

where i = 1 is the number of C-layers, $x_i \in Q_j, 1 \le t \le T = n_i, Q_j$—specific combination of C-layers, which has or hasn't non-standard shape RP.

The method of synthesis of CNN mathematical model parameters with extended training set based on the proposed algorithm changes of a shape of RP for neurons on an arbitrary combination of C-layers is proposed. This method adapts the back propagation algorithm to the non-standard form of RP.

It includes the following steps:

1. Change RP of neurons at the required combination of C-layers, using the algorithm of RP forms changes (Fig. 2), in the process of training before feeding a next pattern to the input of CNN.
2. Obtain neurons output values of C-layer during forward propagation according to 6 instead of 1.

$$y_{m,n} = \varphi\left(b + \sum_{q \in Q_i} \sum_{k=0}^{K_C-1} \sum_{l=0}^{K_C-1} Z^q_{k,l} X^q_{m+k+F_i(RP_{m,n};k;l),n+l+F_j(RP_{m,n};k;l)}\right), \tag{6}$$

where $F_i(RP_{m,n}; k; l), F_j(RP_{m,n}; k; l)$—the functions returns a shifts of row and column of RP template for neuron m, n at position k, l in this template. $index_{k,l}$ is an element of a template $RP_{m,n}$ at the position k, l, $index_{k,l} = 0 \ldots 24$. These functions are defined as:

$$F_i(\cdot) = \begin{cases} 0; & index_{k,l} \in \{0, 4, 5, 16, 17\} \\ 1; & index_{k,l} \in \{6, 7, 8, 18, 19\} \\ 2; & index_{k,l} \in \{20, 21, 22, 23, 24\} \\ -1; & index_{k,l} \in \{1, 2, 3, 14, 15\} \\ -2; & index_{k,l} \in \{9, 10, 11, 12, 13\} \end{cases}$$

$$F_j(\cdot) = \begin{cases} 0; & index_{k,l} \in \{0, 2, 7, 11, 22\} \\ 1; & index_{k,l} \in \{3, 5, 8, 12, 23\} \\ 2; & index_{k,l} \in \{13, 15, 17, 19, 24\} \\ -1; & index_{k,l} \in \{1, 4, 6, 10, 21\} \\ -2; & index_{k,l} \in \{9, 14, 16, 18, 20\} \end{cases}, \tag{7}$$

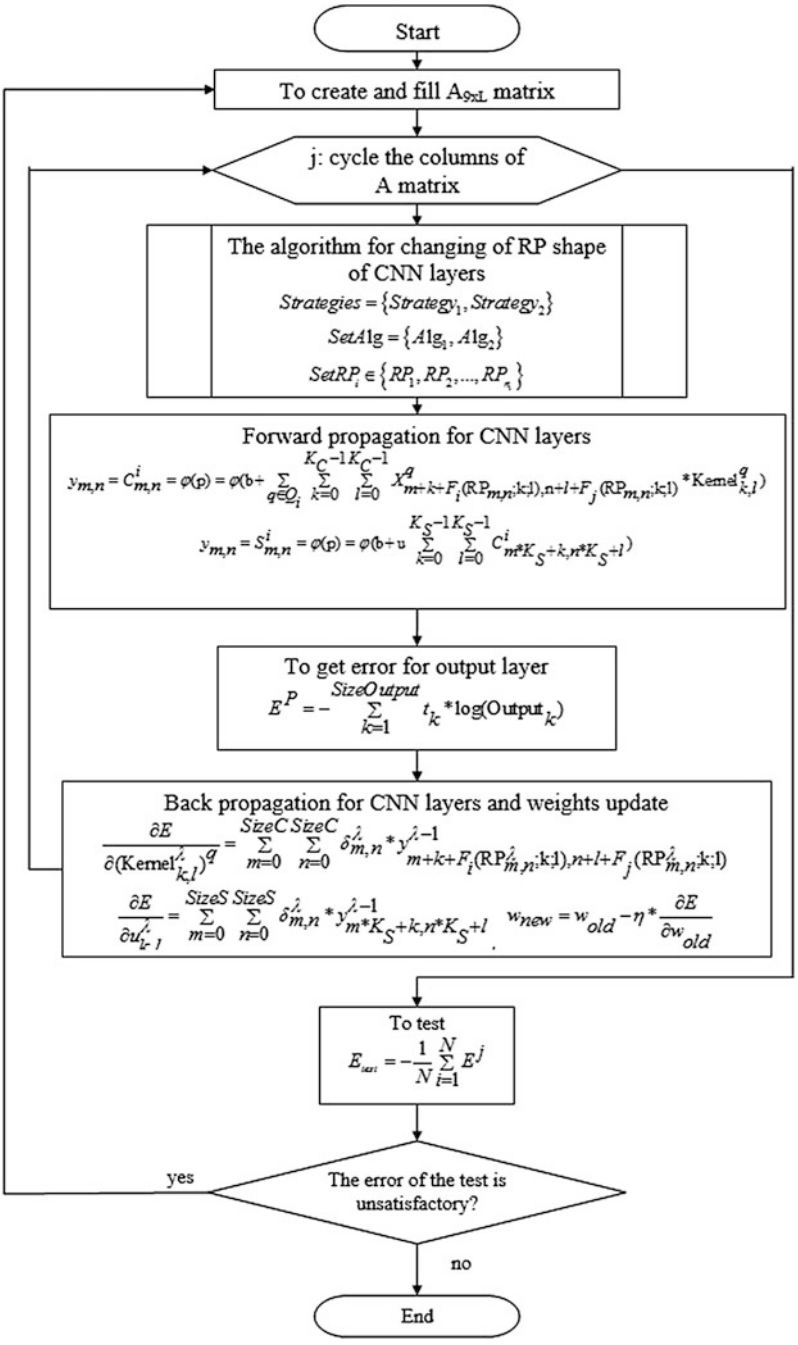


Fig. 2 The generalized synthesis features algorithm of CNN mathematical model with extended training set generated by the change of its RP

3. Obtain a local gradient for C-layer according to 8 instead of 3 in the process of back propagation.

$$\frac{\partial E}{\partial (Z_{k,l}^{\lambda})^{q}} = \sum_{m=0}^{SizeC} \sum_{n=0}^{SizeC} \delta_{m,n}^{\lambda} * y_{m+k+F_i(RP_{m,n}^{\lambda};k;l),n+l+F_j(RP_{m,n}^{\lambda}k;l)}^{\lambda-1}, \quad (8)$$

where $C_{m,n}^{i}$—output of the neuron located on the ith card C-layer in position m, n, $\varphi(\cdot) = A * tanh(B * p)$ when A = 1.7159, B = 2/3, p—weighted sum, b—bias, Q_i—set of cards indices of the previous layer associated with the C^i card, K_C—size of square RP for $C_{m,n}^{i}$ neuron, $X_{m+k,n+l}^{q}$—input to the $C_{m,n}^{i}$ neuron, $Z_{k,l}^{q}$—qth part of the adjustable parameters, which is responsible for interaction with the qth card of the previous layer.

Features synthesis algorithm of CNN mathematical model with extended training set (Fig. 2).

Numerical method for reduction of the extended training set generated by the change of CNN receptive fields was developed [9]. With increasing degree of network training more and more of patterns have weak contagion on the network weights adjustments. So to speed up the training without loss of quality, statistical information about the correct recognition of distorted patterns accumulated in the previous period is used. On the basis of this information, you can skip the part forward propagations, resulting in a reduction in training time without loss of quality recognition.

Matrix A reduction is carried at the beginning of the next period basis on convolution of comparing patterns list to be removed. At the end of period the list is created. During the period statistical information about patterns recognition is accumulated. For this purpose two matrices X_{N_0X7} and Y_{N_0X7} are used, where $x_{i,j}, y_{i,j}$—natural numbers. If the pattern distortions is correctly recognized, then the value of the matrix X is increased by 1in the row i (which is equal to the index of the donor-pattern, on which distortions are superimposed) and in column j (which is equal to the index of a particular combination of C-layers with non-standard RP). If it is recognized incorrectly, the value increases in the Y matrix at the same positions.

Realization algorithm of this method is shown in Fig. 3.

3 Experimental Research of CNN Generalization Capability with Changing Receptive Fields

Experiments for evaluating generalization capability of the proposed CNN mathematical model and quality of objects different classes recognition were carried out. Structure diagram of design techniques objects recognition software from the mobile robot camera is shown.

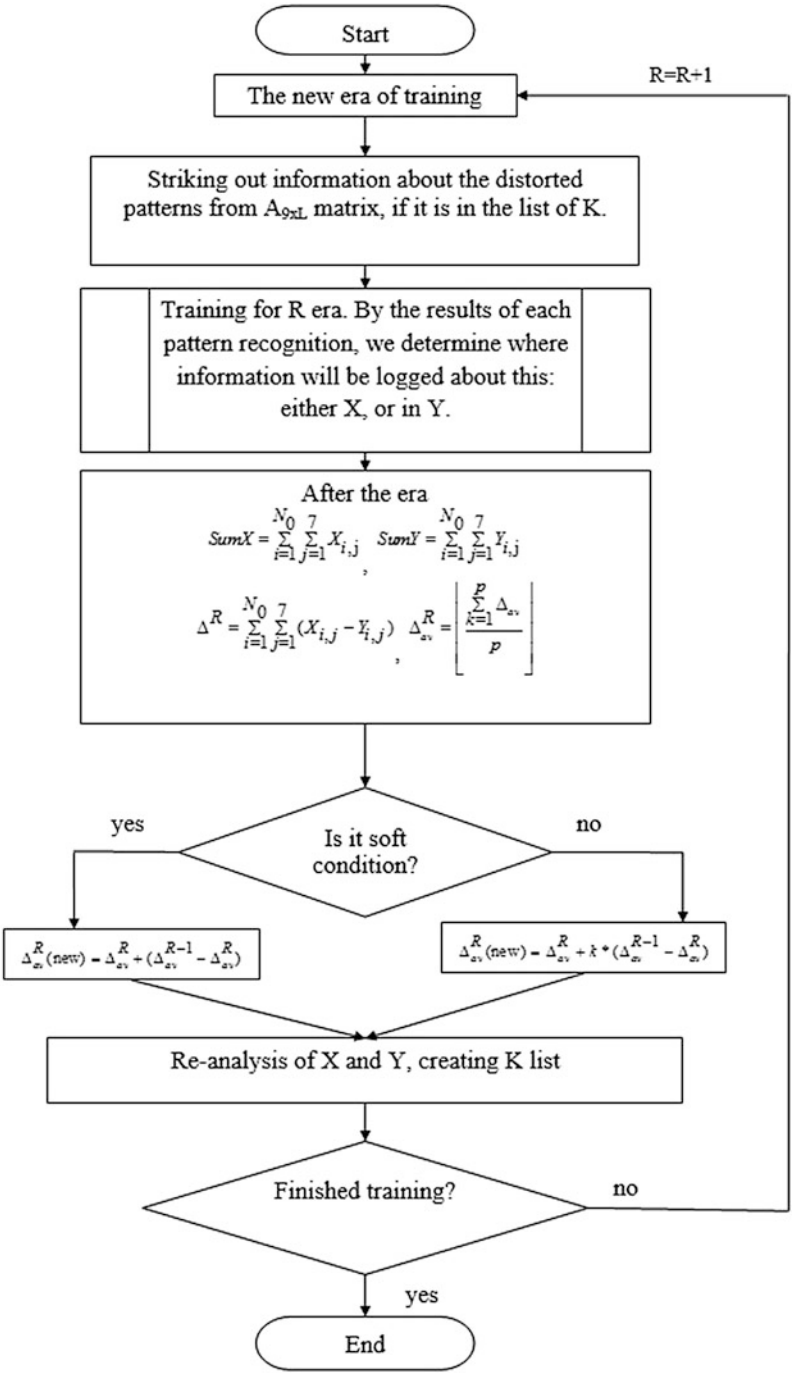


Fig. 3 The generalized extended training set reduction algorithm

Table 1 The results of computational experiments to assess the generalization capability of the learning algorithm of CNN with the help of proposed method and the impact of the reduction in the learning process

Generalization error with using of the proposed method of synthesis features mathematical model of CNN, %	Generalization error without using of the proposed method of synthesis features mathematical model of CNN, %	Error generalizations of the best analogue, %	Training time without reduction, h	Training time with reduction, h
MNIST				
0.6	2.8	0.23	46.15	31.2
Small NORB				
4.3	8.4	6.6	15.6	9.8
Training set for "Mechatronics" stand				
0.3	1.8	–	17.3	10.1

Experiments for evaluating generalization capability of the proposed CNN mathematical model with extended training set were carried out using three training sets: MNIST (handwritten digits) [5], Small NORB (five classes of objects) [8], the training set is created for recognition of 10 classes of objects from the "Mechatronics" stand (vision system: AXIS M1054 camera, 320 × 240 pixels (HDTV), horizontal viewing angle 80, the allowable exposure range from 0.9 to 105 lx, frame rate 30 fps (H. 264/M-JPEG)) and providing scale invariance, object position on the plane, the texture backgrounds and the illumination.

In Table 1 the results of evaluating experiments are presented. Training of CNN with a help of mathematical model improves average generalizing capability of the network to 2.5 % compared with the results without it and to 1.25 % in comparison with the best results achieved analogues as follows from the table [10]. The numerical algorithm of extended training set reduction allows to reduce training

Table 2 Percent of correct recognition for the experiments at "Mechatronics"—stand when using the proposed method of training of CNN and without it

The object/the distance from camera	50 cm (%)	1 m (%)	1.5 m (%)	2 m (%)
Bike	91 (85)	87 (79)	88 (74)	75 (69)
Car	89 (81)	85 (77)	82 (71)	68 (67)
Flashlight	94 (82)	90 (71)	92 (65)	72 (59)
Frog	95 (79)	91 (78)	89 (60)	69 (59)
Plane	91 (84)	90 (72)	87 (67)	74 (65)
Player	90 (79)	89 (77)	90 (63)	75 (58)
Pumpkin	85 (81)	83 (71)	80 (68)	63 (63)
Soldier	96 (85)	92 (76)	90 (64)	73 (60)
Stapler	92 (78)	89 (64)	88 (62)	71 (59)
Empty class	98	98	98	96

time by an average of 9 h, which is 37 % of the initial training time without reduction [9].

There was developed software which allows to recognize objects with the camera of mobile robot for objects recognition from "Mechatronics" stand.

The experiments for objects recognition on "Mechatronics" stand for different distances from the camera (AXIS M1054) are shown in Table 2.

According to results we see that the recognition accuracy for objects up to 96 % when using the method of synthesis of mathematical model features CNN with extended training set. When distance to the camera is increase the recognition accuracy is decrease due to inferior quality of input image. In this article presented the results of applying of convolutional neural network with changed receptive fields towards image recognition tasks.

References

1. Bengio, Y.: Learning deep architectures for AI. Found. Trends Mach. Learn. **2**(1), 1–127 (2009)
2. Bengio, Y., LeCun, Y.: Scaling learning algorithms towards AI. In: Large Scale Kernel Machines, MIT Press (2007)
3. Bishop, C.M.: Neural Networks for Pattern Recognition, pp. 482. Great Clarendon Street, USA (1995)
4. Cruz, V., Cristobal, G., Michaux, T., Barquin, S.: Invariant image recognition using a multi-network neural model. In: Proceedings International Joint Conference Neural Networks, vol. II, pp. 17–21 (1989). Electronic Neurocomputers
5. Modified National Institute of Standards and Technology (MNIST) (30.01.2016). http://yann.lecun.com/exdb/mnist/
6. Nemkov, R.: Dynamical change of the perceiving properties of neural networks as training with noise and its impact on pattern recognition. In: Young Scientists—International Workshop on Trends in Information Processing (YSIP) (2014) (30.01.2016). http://ceur-ws.org/Vol-1145/paper4.pdf/
7. Nemkov., R., Mezentseva, O.: The use of convolutional neural networks with non-specific receptive fields. In: The 4th International Scientific Conference: Applied Natural Sciences, pp. 284–289. Novy Smokovec (2013)
8. NYU Object Recognition Benchmark (NORB) (30.01.2016). www.cs.nyu.edu/ylclab/data/norb-v1.0/
9. Nemkov, R.M., Mezentseva, O.S.: Dynamical change of the perceiving properties of convolutional neural networks and its impact on generalization. Neurocomputers: Development and Application, vol. 2, pp. 12–18 (2015)
10. Nemkov, R.M.: Synthesis method of mathematical model parameters of the convolutional neural network with extended training set (30.01.2016). http://www.science-education.ru/125-19867/
11. Simard, P., Steinkraus, D., Platt, J.C.: Best practices for convolutional neural networks applied to visual document analysis. In: Proceeding of the Seventh International Conference on Document Analysis and Recognition (ICDAR03), vol. 2, pp. 958–964 (2003)

Application CUDA for Optimization ANN in Forecasting Electricity on Industrial Enterprise

Roman Taranov

Abstract Technological progress in the manufacturing sector is characterized by an increase in energy consumption and, consequently, an increase in electricity consumption. It's necessary to carry out electricities economical consumption to meet the growing demand for electricity. The problem of forecasting of energy consumption is a complex multi-factor problem with nonlinear dependencies. Due to the complexity of the calculations for the solution of this problem requires large computational resources. Therefore there is a need of optimization algorithms to improve the quality of the forecast. This article describes the use of parallel computing on the GPU algorithm neural network training based on CUDA technology, to optimize the energy consumption prediction process in an industrial plant. According to the results of the experiments presented in this paper, the parallel algorithm has reached the required prediction accuracy for a shorter period of time. Applying the proposed algorithm can enable enterprises to get a more accurate prognosis and reduce the costs associated with payment of electricity.

Keywords Forecasting · Neural networks · CUDA

1 Introduction

In large industrial enterprises in the current conditions of power consumption there is a problem, in the preparation of the application to the power company on the amount of power required at a given time. If the real readings of consumption will differ from the amounts claimed by more than 5 %, the company will be fined. The ultimate goal is to reduce the cost of energy saving on your electricity bill (see [1]).

R. Taranov (✉)
North-Caucasus Federal University, Stavropol, Russia
e-mail: roman_boxx@mail.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_3

Today, analysts do forecasting and calculation of the required volume of electricity on the vast majority of enterprises. The most common method of forecasting based on specific rates of electricity with expert adjustments (static method of regression analysis). The error calculation is usually more than 5 %, as a result there are significant financial losses. Also, as a model to approximate the load curve and the forecast can be successfully used polynomial expansion (see [2]). If a prediction is needed for a long period of time, such models are unsuitable due to computational complexity. In such cases, there may be used autoregressive methods (see [3]).

All these methods have significant drawbacks:

- their application requires direct dialog analysis participation, i.e. human resources;
- computational complexity;
- insufficient accuracy of the forecast;
- it applies only to a specific type of forecast (for example, only short-term forecasting);
- sensitivity to input data.

It is therefore necessary to use a more efficient method of forecasting, with a minimum participation of specialists and higher accuracy of the result. Now new approaches to forecasting electricity load are available, being developed and implemented. These new methods include artificial neural networks.

The aim of this work is to optimize the computational part of the learning algorithm of artificial neural network back propagation by parallelization of computing processes with the direct signal propagation across the network and error back-propagation.

2 Description of the Neural Network Model

2.1 Artificial Neural Network Model

ANN are forecasting methods that are based on simple mathematical models of the brain. They allow to complex nonlinear relationships between the response variable and its predictors (see [4]).

A neural network can be thought of as a network of neurons organized in layers. The predictors (or inputs) form the bottom layer, and the forecasts (or outputs) form the top layer. There may be intermediate layers containing hidden neurons (see [4]).

A simple example is show in Fig. 1.

This is known as a multilayer feed-forward network where each layer of nodes receives inputs from the previous layers. The outputs of nodes in one layer are inputs to the next layer. The inputs to each node are combined using a weighted

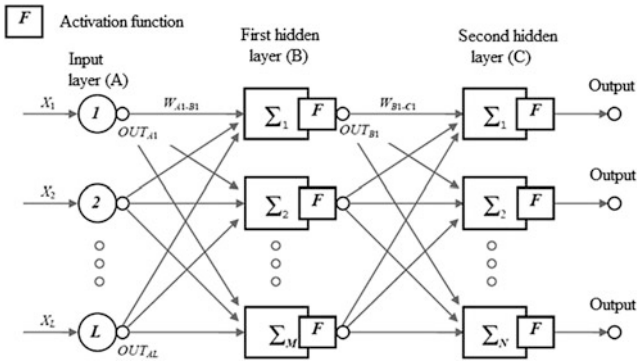


Fig. 1 Structure of neural network

linear combination. The result is then modified by a nonlinear function (1) before being output.

$$OUT = F(\xi) = \frac{1}{1 + \exp(-\xi)} \tag{1}$$

where, $\xi = \sum_{k-1}^{L} x_k \omega_k$; x_k—inputs; ω_k—weights; L—number of inputs.

The weights take random values to begin with, which are then updated using the observed data. Consequently, there is an element of randomness in the predictions produced by a neural network. Therefore, the network is usually trained several times using different random starting points, and the results are averaged (see [4]).

2.2 CUDA

CUDA is a parallel computing platform and programming model created by NVIDIA and implemented on graphics processing units (GPUs). CUDA gives developers access to the virtual instructions sets and memories of the parallel computational elements in CUDA GPUs. This approach of solving general-purpose problems on GPUs is known as general purpose graphics processing unit (GP-GPU) (see [5]).

CUDA is ideal for an embarrassingly parallel problem, where little or no interthread of interblock communication is request. It supports interthread communication with explicit primitives using on-chips resources. Interblock communication is, however, only supported by invoking multiple kernels in series, communicating between kernel runs using off-chip global memory. It can also be performed in a somewhat restricted way through atomic operations to or from global memory (see [6]).

CUDA splits problems into grids of blocks, each containing multiple threads. The blocks may run in any order. Only a subset of the blocks will ever execute at

any one point in time. A block must execute from start to completion and may be run on one of N SMs (symmetrical multiprocessors). Blocks are allocated the grid of blocks to any SM that has free slots. Initially this is done on a round-robin basis so each SM gets an equal distribution of blocks. For most kernels, the number of blocks needs to be in the order of eight or more times the number of physical SMs on the GPU (see [6]).

2.3 Learning Algorithm of Neural Network

In this work is used supervised learning of neural network. Supervised learning is the machine learning task which uses labeled training data as a set of examples. This labeled training data consist of inputs as vectors X for a trained system and a desired system output y for each input.

Supervised learning algorithm uses training data to produce an inferred function $y = f(x)$ which can be a classifier or a regression function. Classifier is produced if outputs are discrete and regression function is produced if outputs are continuous. The main idea of the supervised learning algorithm is to generalize similar inputs for a given output and thus inferred function can be produced. Further reading on the subject can be found in 'Intelligent Systems' (see [7]).

The parameters are estimated such that the cost function of neural network is minimized. Cost function is an overall accuracy criterion such as the following mean squared error:

$$E = \frac{1}{N}\sum_{n=1}^{N}(e_i)^2 \tag{2}$$

where, N is the number of error terms.

This minimization is done with some efficient nonlinear optimization algorithms other than the basic backpropagation training algorithm (see [8]).

Correction of weights output layer:

$$\omega_{p-k}(i+1) = \omega_{p-k}(i) + \eta\delta_k OUT_p \tag{3}$$

where, i—number of iteration; $\omega(p-k)$—weight, p—neuron of hidden layer, k—neuron of output layer; $\delta_k = OUT_k(1-OUT_k)(T_k - OUT_k)$; η—coefficient of speed training; OUT_p—output of hidden layer; T_k—desired system output; OUT_k—output of output layer.

Correction of weights hidden layer:

$$\omega_{p-q}(i+1) = \omega_{p-q}(i) + \eta\delta_q OUT_p \tag{4}$$

where, i—number of iteration; ω_{p-q}—weights, p—neuron from the previous layers, q—neuron from hidden layer; $\delta_q = OUT_q(1 - OUT_q)\sum_{k=1}^{N}\delta_k\omega_{q-k}$; η—coefficient of speed training; OUT_p—output of neuron from the previous layer; OUT_q—output of neuron from hidden layer; N—number of neurons in next layer.

2.4 Modification of a Model

Currently, there is no general algorithm for determining the structure of the ANN, suitable for each of the problem. Often, such a structure is selected by "trial and error", which often takes researchers a long time (see [9]). Since each structure is necessary to select the values of the weighting factors, the acceleration of the selection should speed up the learning process of ANN.

In this paper we propose a model based on the parallel passage of neurons both in the direct distribution of signals, and in the opposite direction to adjust the balance. The fact that the neurons in the same layer are not dependent on each other, allowing parallelized computational process (Figs. 2 and 3).

3 Implementation of ANN in the Programming Language *C++*

A neural network is implemented in the programming language *C++*. For computational experiments and compare the results of both algorithms were coded: for serial and parallel signals passing through the network.

The CUDA libraries were used for the implementation of the parallel signal path on the network. CUDA platform of parallel computing provides a set of extensions

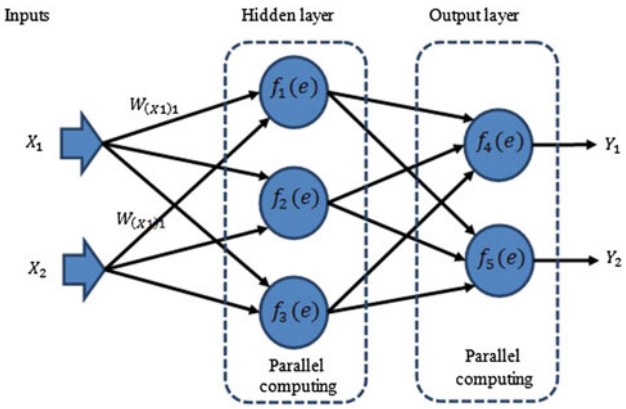


Fig. 2 Propagation network with parallel training

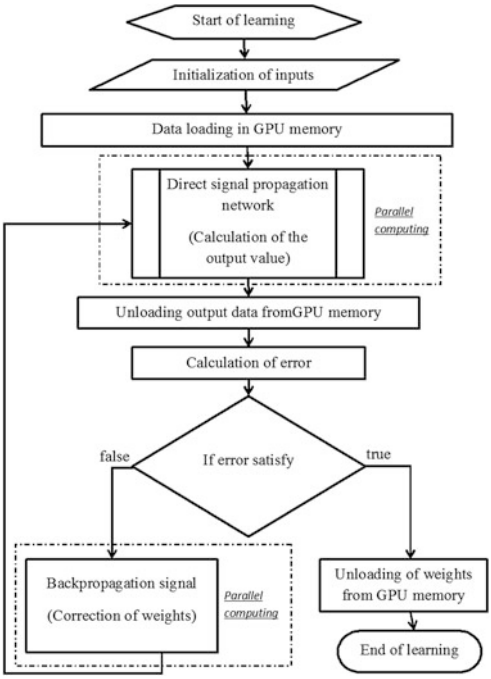


Fig. 3 Algorithm ANN training with parallel computing technology CUDA

for *C* or *C++*, allowing to express both data parallelism and task parallelism at the level of small and large structural units.

Function *ActiveSign(int iter)* implements the activation function by the formula number 1.

```
// Activation function
float CNeuroNet::ActiveSigm(int iter)
{
    float e = (float)E;

    float y = 1 / (1 + pow(e, -(m_Neuro_hidden[iter].sum)));

    return y;
}

void CNeuroNet::ProcNeuro(int iter)
{
    float Net = 0;

    for (int i = 0; i < INCOUNT; i++)
    {
        Net += this->m_Neuro_hidden[iter].x[i] *
                    m_Neuro_hidden[iter].w[i];
    }

    m_Neuro_hidden[iter].sum = Net;

    m_Neuro_hidden[iter].output = ActiveSigm(iter);
}
```

this *E*—is a constant, the base of the natural logarithm of the value of $\approx 2,71828182845904$; function *ProcNeuro (int iter)* performs works by sequential summation of the input parameters to the weights of neurons. The function is called cyclically for all the neurons of the layer. Parameter function *iter*—the number of iteration passage layer; *i*—number of neuron in the layer; cycle *for(){}* within the function implements a consistent flow of all the weighting factors relating to the current neuron; *INCOUT*—the number of the input parameters specified by the user.

Next fragment of code that implements the function of the activation algorithm for parallel signal flow through the network by means of an expanded library CUDA:

```
__global__ void MainKernel(float * input,
                           float * weights,
                           float *sum_vector,
                           float *output,
                           int Size,
                           int iStandartIteration)
{
   float e = (float)E;
   int xIndex = blockIdx.x * blockDim.x + threadIdx.x;
   sum_vector[xIndex] = 0;
   output[xIndex] = 0;

   for (int k = 0; k < Size; k++)
   {
      sum_vector[xIndex] += weights[xIndex * Size + k] *
                            input[iStandartIteration * Size + k];
   }

   output[xIndex] = 1. / (1 + pow(e, -(sum_vector[xIndex])));
}
```

Function *Main Kernel ⋘⋙()*—a function of the core is carried out on the graphic processor (GPU). Variable *blockIdx*—the index of the current block in the computation on the GPU. Variable *blockDim*—the dimension of the current block in the computation on the GPU. Variable *threadIdx*—the index of the current thread in the computation on the GPU. Cycle *for ()* does a sum of products of weight coefficients on the input parameters. This function is performed simultaneously (in parallel) for all neurons layer.

4 Experiments and Results

4.1 Formalization of the Problem

Consider the example of monthly forecasting electricity company CISC "ELECTROTECHNICAL FACTORIES "ENERGOMERA" with the use of neural networks.

For high reliability of the data used in the target company should be introduced high-precision multifunctional automated system of control and metering of electricity consumption (AMR).

For input parameters accepted—the average monthly temperature, which is characteristic for a given month, the number of holidays and days of the month, the average duration of the day, the planned production volume for the month. As well as the value of electricity consumption in the previous month, the maximum and minimum monthly consumption values for the previous year. Data on consumption of electricity imported from the AMR system cEnergo 4.5 (company Energomera). For the experiment were used real factory data CISC "Electrotechnical Factories "Energomera" for 2014. All input data is represented in a table format Microsoft Excel.

For the training of the National Assembly there must be a certain amount of training pairs, the more there are, the better the result is. As the task is to predict the monthly value, the minimum number of learning pairs 12—values for each month of the year. The training data set is a pair of ten inputs and one output value. The output value—is the electricity consumption for the last month of the period selected for training, and input parameters—the characteristics of the same month, which may affect the power consumption: the average air temperature, the number of working days and holidays, the duration of daylight, the volume of production. All input parameters must be normalized in the interval from 0 to 1. With the normalization of negative temperatures taken as zero value of 0,5.

The initial values of the weighting factors ANN are given random values belonging to a range from 0 to 0,1. The purpose of the learning process is to find the optimal values of the weights at which the trained neural network produces the resultant value within a given error for all training pairs.

4.2 Experiments

All calculations were performed on PC with the specifications:

- CPU: Intel(R) Core(TM) i5 760 2.8 GHz;
- RAM: 8,00 Gb;
- Operation system: Windows 7 Professional x64;

A GPU (graphics cards) NVIDIA requires to run the program with the parallel implementation of the algorithm ANN training. In the experiment, for this work used unit graphics card (GPU) NVIDIA GeForce GT 240.

In this case, the output layer, representing the value of power consumption per month plant comprises one neuron hidden layer initially (during calibration network can change)—128 neurons and the input—10 neurons.

One neuron is characterized by average monthly temperature characteristic of the projected month. While training in the forecast temperature is taken as a fact known temperature. Two neuron describe the number of working days and holidays. One

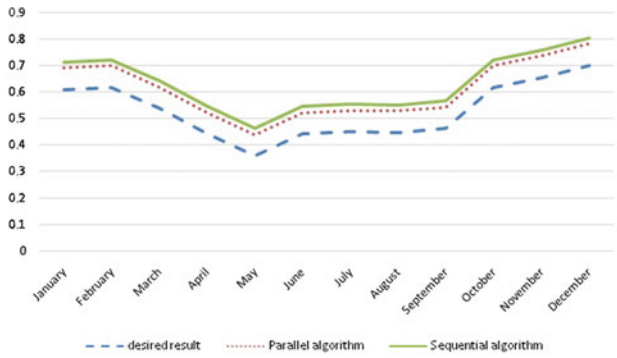


Fig. 4 Result of training networks for parallel and serial algorithms within the first ten seconds

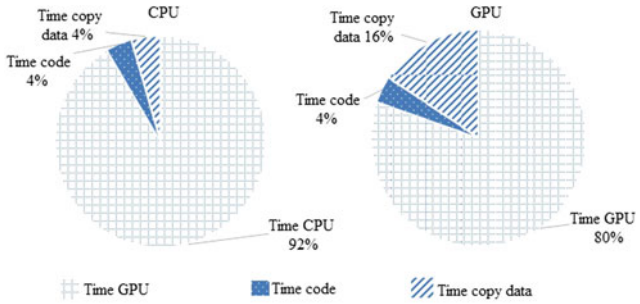


Fig. 5 The cost of computing power

neuron is responsible for the duration of daylight. Six neurons describe the planned production volume over the forecast period.

The efficacy of the parallel learning algorithm can be clearly shown in Fig. 4. The figure shows the result of learning network for 10 s for the parallel and serial algorithms. In both cases, we used the same data from one network and identical parameters.

With the implementation of the parallel algorithm a small percentage of the power is spent on the transfer of data to the GPU and back. This is shown schematically in Fig. 5. The time required to copy data in the implementation of parallel training will be significantly greater than for implementation of sequential algorithm, but the final time of training network for parallel algorithm—many times less than in the case of sequential algorithm. That is confirmed by the data in Table 1.

Table 1 presents the results of algorithms. These data show a significant advantage of the parallel algorithm. The program is executed under identical input values and with the same values of configurable network parameters (learning rate, number of neurons, the maximum number of iterations, the margin of error). The computational complexity of algorithms identical since all calculations are built on a set of formulas.

Table 1 Principles of a block learning algorithm

	Sequential algorithm	Parallel algorithm
Number of teaching periods	1 million	1 million
Time (sec.)	10365,5	1418,6
The absolute error (%)	2,9999	2,9999
Standard deviation (MSE)	0,01689	0,01689

5 Conclusion

This paper analyzes the popular methods of forecasting energy consumption. Presents arguments in favor of the choice of forecasting method using ANN. The theory of the INS is studied, the algorithm of neural network is trained. The resulting algorithm has been optimized through parallelization of individual blocks. In the *C++* programming language designed programs that implement both algorithms. Programs have been optimized in terms of performance, functionality and reliability. The process of parallelization implemented using parallel computing platform CUDA, which provides a set of extensions for *C* or *C++*. Inserted practical experiment using real indicators of factory CISC "Electrotechnical Factories "Energomera" and compared the results of the program.

During the study of the model ANN we achieved the mean relative error of the forecast for the month is less than 3 %. This model can easily be reconfigured for any company. Embedded in the model of parallelization of computational process CUDA makes it possible to gain an advantage even when processing small amounts of data. The desired accuracy is achieved in less time, in consequence of which is greatly reduced while the process of training the neural network. These results suggest that, with limited resources, developed a model of parallel learning neural network will provide a higher prediction accuracy with the same resources.

References

1. Federal Law of the Russian Federation: On Electricity. No. 35, dated 26.03.2003
2. Akhmet'yanov, R.R., Delehodyna, L.A., Kopylov, N.P.: Tasks predicting energy consumption in the integrated AMR Novosibirsk Scientific Center. Energy Conserv. J. **1** (2007)
3. Akhmet'yanov, R.R., Delehodyna, L.A., Kopylov, N.P.: The multiplicative model the seasonal energy enterprises. Avtometryya J. **3** (2008)
4. Hyndman, R.J., Athanasopoulos, G.: Forecasting: principles and practice. Publisher: OTexts (2013)
5. Cipolla-Ficarra, F.V.: Handbook of Research on Interactive information Quality in Latin Association of Human-Computer Interaction. Spain & International Association of Interactive Communication, Italy (2005)
6. Shane Cook: CUDA Programming. A Developers Guide to Parallel Computing with GPUs. Elsevier Inc. (2013)

7. Gerstner, W.: Supervised learning for neural network: a tutorial with java exercises. In: Mlynek, D., Teodorescu, H.-N. (eds.) Intelligent Systems, An EPFL Graduate Course (1999)
8. Rumelhart, D., McClelland, J.: Parallel Distributed Processing. MIT Press, Cambridge, MA (1986)
9. Arzamasians, A.A., Kryuchin, O.V., Zakharova, P.A., Zenkova, N.A: The universal software system for computer simulation based on artificial neural network with self-organizing structure. J. Tambov Univ. Rep. Ser. Nat. Tech. Sci. **5** (2006)

Part II
Probabilistic models, Algebraic Bayesian Networks and Information Protection

Models and Algorithms for the Information System's Users' Protection Level Probabilistic Estimation

Artur Azarov, Maxim Abramov, Alexander Tulupyev and Tatiana Tulupyeva

Abstract Questions of information security for today's business, somehow using information technologies are among the most important. Probabilistic relational model complex "information system—critical documents—the staff—the attacker" are being concidered, as well as an algorithm for estimating the probability of users of information systems security by social engineering attacks. Also a method for estimating the probability of achieving critical information stored in the information system by the malefactor was presented.

Keywords Social engineering attacks · User protection · Information security · Malefactor's competence profile

A. Azarov (✉) · M. Abramov · A. Tulupyev · T. Tulupyeva
St. Petersburg State University, Saint Petersburg, Russia
e-mail: artur-azarov@yandex.ru

M. Abramov
e-mail: mva16@list.ru

A. Tulupyev
e-mail: alexander.tulupyev@gmail.com

T. Tulupyeva
e-mail: tvt100a@mail.ru

A. Azarov · M. Abramov
Moscow Pedagogical State University, Moscow, Russia

T. Tulupyeva
North-West Institute of Management, Saint Petersburg, Russia

A. Azarov · M. Abramov · A. Tulupyev · T. Tulupyeva
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science, 39, 14-th Line V.O., Saint Petersburg 199178, Russia

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_4

1 Introduction

The issues of information protection for the modern business, which are using information technologies, are still the most important. These facts are confirmed by the increasing the frequency of the events associated with the information security incidents [1]. One of the most serious issues for the companies' private network is an illegal access that can be obtained through it users. This problem has a close relation to the recruitment of spies agents in enemy territory, and therefore has a long history, but in modern times it has close connection with modern technologies and, for this reasons, links it with the security of the entire network.

There are a number of researches devoted to the problem of manipulative influence on people [2–7]. In particular, the emphasis in this type of the research is on the manipulation of facts in the social media. Basically, considered manipulative effects are used in newspapers, television and radio. These types of impacts unites into the informational psychological impacts, which, in turn, are a part of malefactor's social engineering attacks. But it should be mentioned that social engineering attacks are not only the malefactor's impacts that leads to the information security incidents.

Currently, much of the research in the information security are dedicated to the improvement of technical base of information security [8–13]. These issues of information security are well understood, developed a large number of methods that minimize the probability of success of a malefactor's attacks. At the same time, there are still a lack of the methods that allows to protect informational systems users from social engineering attacks.

Main purposes of this article are presenting probabilistic relational models of "information system—critical documents—personnel—malefactor" complex and some algorithms of the evaluation of the probability of malefactor's attack action success, based on the given user's vulnerabilities and malefactor's resources.

2 Models' Description

Some limitation should be mentioned while informational systems' models are being described. The article focuses on the issues of informational systems' users' protection analysis, but not on the impacts on hardware and software components of information system. Relational probabilistic models of the "information system—critical documents—personnel—malefactor" complex are presented in this article. These complex is the development of the "informational system—personnel—critical documents" complex. The basic elements of this complex are:

(1) model of hosts of informational systems and computer network;
(2) model of users of informational systems, including the model of the user's vulnerabilities profile;

(3) model of critical documents, that are being stored on the hosts;
(4) model of malefactor, who implements social engineering attack actions on the users of informational system, with the limits on resources, time and types of social engineering attack actions.

Let consider these models. Assume that "information system—critical documents—personnel—malefactor" complex includes k hosts, l users, v user's vulnerabilities.

2.1 Model of the User of Information System and the Model of User's Vulnerabilities Profile

One of the possible representation of the model of informational systems' user may consist of:

- model of user's vulnerabilities profile;
- user's access to the different hosts of informational system;
- relationships between current user and the other users of informational system.

To sum up, the model may be presented as:

$$U_i = \left(\left\{ V_j, V_j^i(D) \right\}_{j=1}^{v}; \left\{ Ac_j^i \right\}_{j=1}^{k}; \left\{ L_j^i \right\}_{j=1}^{l} \right),$$

where $\left\{ V_j, V_j^i(D) \right\}$—user's vulnerabilities profile, where V_j—j user's vulnerability, and $V_j^i(D)$—the degree of j vulnerability i user, $\{Ac_j^i\}_{j=1}^{k}$—the access of i user to the j host, $\left\{ L_j^i \right\}_{j=1}^{l}$—type of the relationship between current user and the other users of informational system.

Conducted research revealed a relationship between level of manifestation of personality attributes and the user's vulnerability degree [14].

2.2 Model of Malefactor

It is concerned that the model of malefactor includes some information about malefactor's resources, which includes time available for the preparation and the implementation of social engineering attack actions, types of social engineering attack actions and basic knowledge of the malefactor about the users of informational system, which can be obtained by the processing open data in social networks. This model can be represented as: $A_i = \left(R^i; \left\{ At_j, At_j^i(D) \right\}_{j=1}^{v}; \{BK_j\}_{j=1}^{l} \right),$

where R^i—resources of the i malefactor, $\left\{At_j, At_j^i(D)\right\}_{j=1}^{v}$—types of social engineering attack actions and the opportunities to use these social engineering attack actions by the i malefactor, $\left\{BK_j\right\}_{j=1}^{l}$—basic knowledge of the malefactor about the users of informational system, which can be obtained by the processing open data in social networks. Malefactor may predict some user's vulnerabilities based on the information from user's personal pages in social networks.

Two types of malefactor's social engineering attack actions will be considered in this article.

The first type of these actions is manipulation attacks. In these types of attacks malefactor uses his knowledges, charisma and some other features. For example, in 2011 in Moscow personable man entered one of the bank. On the lapel of his jacket were located a sign of the Deputy of the Russian State Duma. He introduced himself as a famous Russian statesmen. He declared that he is a VIP client of this bank and has announced his personal bank account. Banks's cashier examined his personal bank account and was astonished due to the fact that she is talking with so rich and respectable man. The man informed that he wants to transfer 50 million rubles to his other bank account in the other bank. At the same time fraud did not show any personal documents. Cashier reported to the head of the office, she also has no doubt and gave the nod for the issue operations [15]. Thus, malefactor by forming employees reaction was able to successfully impersonate the account holder and implement some unlawful acts without hindrance from the bank staff, by manipulating them to act with no job descriptions.

The second type of the considered social engineering attacks is attacks which looks like services in exchange for private information. As an example of this attack is bribery of the informational system's user. It should be mentioned that in the considered malefactor's model resources of malefactor are used only for the implementation the second type of the social engineering attack actions.

2.3 Model of Hosts of Informational Systems and Computer Network

Model of host includes the set of the applications that are installed on this device and links between different hosts in the informational network. The model may be represented as: $CM_i = \left(\left\{Apps_j^i\right\}_{j=1}^{m}; \left\{L_j^i\right\}_{j=1}^{k}\right)$, where $\left\{Apps_j^i\right\}_{j=1}^{m}$—the set of the applications that are installed on this device, $\left\{L_j^i\right\}_{j=1}^{k}$—The set of the links between i host and other hosts of the network.

2.4 Model of Critical Documents

Model of critical document can be described with the financial (or other) damage, that can be caused to the organization in case of unauthorized access to this document by the malefactor; and with the hosts, on which this critical document is being stored. This model can be represented as: $CD_i = \left(Dm^i; \left\{ H_j^i \right\}_{j=1}^{k} \right)$, where Dm^i—financial (or other) damage, that can be caused to the organization in case of unauthorized access to this document by the malefactor, and $\left\{ H_j^i \right\}_{j=1}^{k}$—hosts, on which this critical document is being stored.

Representation of the elementary event "the success of the malefactor's social engineering attack impact".

It should be mentioned that the main target of the malefactor, while he is implementing social engineering attack actions, is the access to the private information that is being stored in the information system. Malefactor's attack actioned may be represented as the sequence of actions. The basic element of this sequence is the malefactor's social engineering attack action. Let consider the elementary event "the success of the malefactor's social engineering attack impact" One of the possible ways to represent the probability of this event for the i users may be as follows:

$$p = \varphi(R; At_i | V_{i1}, \ldots, V_{ik})$$

That means that the probability of the malefactor's social engineering attack action success depends on the variety of the parameters such as malefactor's resources, types of social engineering attack actions and the opportunities to use these social engineering attack actions, and the degrees of user's vulnerabilities.

The basic idea is that one malefactor's social engineering attack action effects only one, particular user's vulnerability. In this way, the probability of the malefactor's social engineering attack action success may be represented as follows

$$P_{\text{suc}} = \frac{At_j^i(D) V_j^i(D)}{At_{j\max}(D) V_{j\max}(D)} \tag{1}$$

At the same time, if one malefactor's social engineering attack action effects several user's vulnerabilities, the probability of social engineering attack action success may be represented as follows:

$$P_{\text{suc}} = 1 - \prod_{j=1}^{v} \left(1 - \frac{At(D) V_j^i(D)}{At_{j\max}(D) V_{j\max}(D)} \right) \tag{2}$$

The propagation of the malefactor's social engineering attack actions on the graph of user's social connection modeling.

Social connection of the users of informational systems may be represented with the graph of user's social connections. The nodes of this graph are the models of users of informational system and the ages—social connections between these users. This graph allows to model the distribution of the malefactor's social engineering attack actions among users of informational system. This type of modeling enables to determine the protection degree of critical information contained in the information system.

Let's consider user's social connection graph which is represented on the Fig. 1.

This graph represents the model of information system. This system includes only 5 users. It should be mentioned, that this graph displays only users if informational system and critical document, that are being stored in this system and the access of users to these documents.

The probability of the malefactor's social engineering attack action success was determined in the previous section. For each users which is presented on the graph above it can be mapped the probability $p_1, \ldots, p_5$, which can be figured out by (1) or (2). Each edge in this graph has it's own weight, that represents the type of the relationships between users of informational system. Let denote such weight as r_{ij}, where i and j are the numbers of users from who and where goes the link. One of the resources of malefactor is a time, that he has for social engineering attack actions implementation. Let consider each implementation as 1 time unit. The probability of accessing the private information by the malefactor may be represented as:

$$P(\text{Inf1}) = 1 - (1 - \bar{p}_1)(1 - \bar{p}_2)(1 - \bar{p}_4) \tag{3}$$

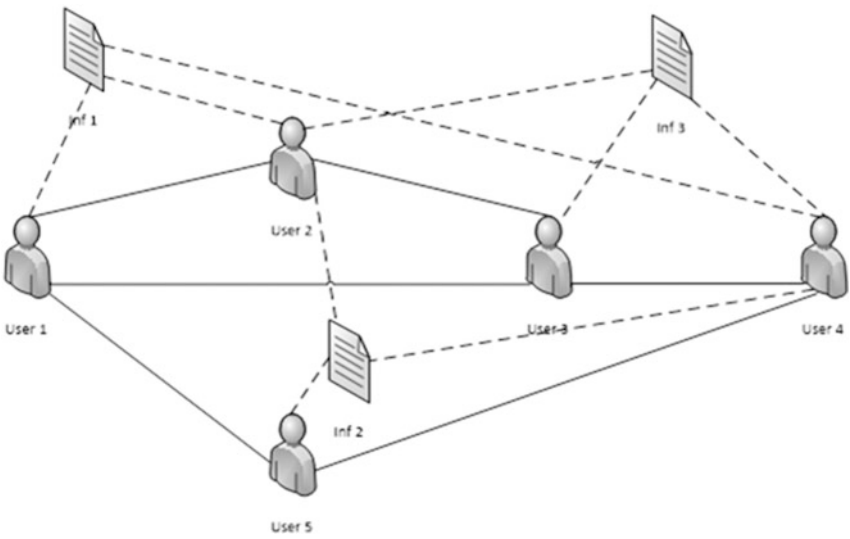


Fig. 1 User's social connection graph

It should be mentioned, that if time that is available for the malefactor is $T = 1$, the probabilities will look like $\bar{p}_1 = p_1, \bar{p}_2 = p_2, \bar{p}_4 = p_4$. If the malefactor has more resources, there are more available links for the malefactor to get the access to the critical documents. Such type of the probabilities will look like:

$$\bar{p}_1 = 1 - (1 - p_1)(1 - \hat{p}_2 r_{21})(1 - \hat{p}_3 r_{13})(1 - \hat{p}_5 r_{15}) \quad (4)$$

$$\bar{p}_2 = 1 - (1 - p_2)(1 - \hat{p}_1 r_{12})(1 - \hat{p}_3 r_{32}) \quad (5)$$

$$\bar{p}_4 = 1 - (1 - p_4)(1 - \hat{p}_3 r_{34})(1 - \hat{p}_5 r_{45}) \quad (6)$$

If $T = 2$, then $\hat{p}_1 = p_1, \hat{p}_2 = p_2, \hat{p}_3 = p_3, \hat{p}_5 = p_5$

In case $T = 3$, which means that the intensity of the malefactor's social engineering attack action distribution equals 3. Each $\bar{p}_1, \bar{p}_2, \bar{p}_3, \bar{p}_5$ will differ from the previous ones. Each (4), (5), (6), will be represented as follows:

For (4)

$$\hat{p}_2 = 1 - (1 - p_2)(1 - p_3 r_{32}) \quad (7)$$

$$\hat{p}_3 = 1 - (1 - p_3)(1 - p_2 r_{23})(1 - p_4 r_{43}) \quad (8)$$

$$\hat{p}_5 = 1 - (1 - p_5)(1 - p_4 r_{45}) \quad (9)$$

For (5)

$$\hat{p}_1 = 1 - (1 - p_1)(1 - p_3 r_{13})(1 - p_5 r_{15}) \quad (10)$$

$$\hat{p}_3 = 1 - (1 - p_3)(1 - p_1 r_{13})(1 - p_4 r_{14}) \quad (11)$$

For (6)

$$\hat{p}_3 = 1 - (1 - p_3)(1 - p_2 r_{23})(1 - p_1 r_{13}) \quad (12)$$

$$\hat{p}_5 = 1 - (1 - p_5)(1 - p_1 r_{15}) \quad (13)$$

Thus, the estimation of the probability of malefactor's social engineering attack actions success were presented.

3 Conclusion

The method of estimation the probability of malefactor's social engineering attack actions success was presented in this article. On the basis of these probabilities the level of critical document protection may be estimated. The further research will be devoted to the development of the algorithms of malefactor's social engineering attack actions imitations.

References

1. Information security business. Studies of current trends in information security business Kaspersky Lab. http://media.kaspersky.com/pdf/IT_risk_report_Russia_2014.pdf (date of the application: 30.04.2015)
2. Cialdini, Robert: Psychology of influence. Persuade, influence, protect.—SPb.: Peter, p. 336 (2010)
3. Demina, M.N.: Changes cognitive practices individual under the influence new information technologies. Sociol. Res. №6, 87–92 (2010)
4. Gavrilov, A.A.: Means the impact of media on public consciousness in the information society [Text]/A. Gavrilov. Young Sci. №8, 152–155 (2012)
5. Kanuhova, T.V.: The impact of media on mass consciousness in the information society. Basic Res. №3, 71–72 (2005)
6. Key Features informational message.: Modern standards of news text. Ifreestore. http://ifreestore.net/1153/24/ (date of the application: 14.01.2016)
7. Pugachev, V.P.: Information power and democracy. Soc. Sci. Present №4, 65–77 (1999)
8. Distefanoa, S., Puliafitob, A.: Information dependability in distributed systems: the dependable distributed storage system. Integr. Comput.-Aided Eng. **21**, 3–18 (2014)
9. Goo, J., M.-S., Yim, Dan, J. Kim, IEEE: 1. A path to successful management of employee security compliance: an empirical study of information security climate. Trans. Prof. Commun. **57**(4), pp. 286–308 (2014). 23p
10. James, C.: Information systems user security: a structured model of the knowing–doing gap. Comput. Hum. Behav. **28**(5), 1849–1858 (2012)
11. Kotenko, I.V., Stepashkin, M.V.: Systems-simulators: purpose, functions, architecture and implementation approach. Math. Univ. Instrum. **T. 49**(3), 3–8 (2006)
12. Kotenko, I.V., Yusupov, R.M.: Future research directions in the field of computer security. Data Prot. Inside № 2, 46 (2006)
13. Trčcek, D., Trobec, R., Pavešsić, N., Tasič, J.F.: Information systems security and human behaviour. Behav. Inf. Technol. **26**(2), 113–118 (2007)
14. Vanushecheva, O.U.: A Prototype Set of Programs to Build the Profile of Psychologically Caused by Vulnerabilities. Diplonaya work. SPb, SPbSU (2012)
15. Rudchenko, A.D., Yurchenko, A.V.: Management systems, business security. http://www.hse.ru/data/2015/10/05/1077104619/!-Тема № 6_(моя,2 лекции).pdf (date of the application: 11.01.2016)

An Adaptive Algorithm for the Steganographic Embedding Information into the Discrete Fourier Transform Phase Spectrum

Oleg Evsutin, Anna Kokurina, Roman Mescheryakov and Olga Shumskaya

Abstract On the example of digital images there are investigated properties of the discrete Fourier transform (DFT) used to embed information into the phase spectrum. The investigation helped to form a new steganografic algorithm that can be used in case of non- compressed images. Peculiarity of the algorithm is changeable size of information put into the blocks of stego-image. Characteristics of the suggested algorithm are comparable with the analog ones, they allow to get the correct information without mistakes.

Keywords Information security · Steganography · Data hiding · Digital images · Digital watermarking · Discrete Fourier transform

1 Introduction

Methods of digital steganography allows to solve wide circle of security problems by embedding different hidden information sequences into the digital objects.

According to the problems under consideration the methods could be divided into two big classes: methods meant to protect the embedding information and

O. Evsutin (✉) · A. Kokurina · R. Mescheryakov · O. Shumskaya
Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia
e-mail: eoo@keva.tusur.ru

A. Kokurina
e-mail: mrv@keva.tusur.ru

R. Mescheryakov
e-mail: anessa.kas@gmail.com

O. Shumskaya
e-mail: shumskaya.oo@gmail.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_5

methods to protect the digital object used to embed the information in. In the first case, confidentiality of the information embedded is provided be transaction within digital objects, in the second the goal of embedding is authentication of the digital objects.

Further we would call any digital object stego-container not depending on the concrete application.

Digital images usually play the role of stego-containers. Such data are widely used now in the nets of general access in parallel with video- and audio-data. Still the digital images are more simple compared to the video- and audio-data [1]. To make the description shorter we would call methods and algorithms steganographic if information is embedded into digital images.

Embedding information into a digital object is done by manipulating the data elements having some redundancy. If stego-containers are used for non-compressed images it could be the natural space redundancy showing up in the proximity of the neighboring pixels. If the compressed images are used for embedding, there are changed the coefficients of frequency transform having redundancy. Important to note, that the steganographic embedding could be done also without further compression.

So, steganographic methods, as many others, could be divided into the space and the frequency ones.

LSB is one of the classic methods used to hide the confidential information in the space of digital images. Publications [2–4] provide some algorithms worked out using this method and differ by the methods to make steganographic endurance and robustness higher.

For digital images authentication there are mainly used watermarks. Examples of watermarks embedding into the digital images space are given in [5, 6]. The main goal of the investigation is to make stego-containers more resistant to geometrical distortions.

Embedding the information into the frequency area is connected with utilization of the discrete cosine transform (DCT) coefficients. An example is the widely spread method of digital images with losses compression (JPEG). Investigation base on this method and became classic is given in [7–9].

Nowadays there are being published new works meant to improve the classic ways. Among them there are attempts to minimize stego-containers distortions by optimization [10–14].

In the papers there are described algorithms to embed information into the area of the discrete wavelet transform (DWT) coefficients. In some cases optimization methods are used to improve secrecy and robustness. In [15] there was suggested steganographic algorithm based on the joint utilization of DWT and DCT.

Still there not many publications where embedding utilizes discrete Fourier transform. It concerns mainly digital images because such methods for audio data are already rather numerous [16, 17].

2 The Suggested Algorithm of Data Embedding

2.1 Review of the Methods to Build in the Information into Digital Images Utilizing Discrete Fourier Transform

Let's have a look at the known steganographic algorithms for embedding information into the digital images based on the discrete Fourier transform. Many of them deal with elements of amplitude Fourier spectrum.

In case of the algorithm described in [18] the space of hiding is formed of the middle frequency elements that are within a ring zone of the given width on the complex plan. In order to embed one bit of a secret message a pare of symmetric elements are changed in such a way that their difference depends on the bit.

In [19] there is described an algorithm of digital watermarks embedding. A digital watermark is formed with the help of pseudo random key sequence and looks like amplitude Fourier spectrum which elements with values of the set {−1, 1} makes a ring in the space of middle frequencies. The authors says that the property of the ring symmetry of the digital watermark provide stability in case of geometry attack like "turn of image". Embedding operation is realized by two methods—additive and multiplicative.

The same algorithm is described in [20]. The difference—the watermark looks like a circle, not a ring and operation of embedding is additive. All elements of digital watermarks are from the set {0, 1}.

In [21] a binary digital watermark having circle symmetry is formed using log-polar mapping. In the process of embedding elements of the digital image amplitude Fourier spectrum corresponding to the digital watermark 1 are recalculated by averaging within the area 3 × 3 multiplied by the amplification coefficient.

Differently from the previous publications where embedding is done into a brightness component of the color space YCbCr, in [22] is used the color space Lab and the same digital watermark is embedded into Fourier images of both chromatic components. Embedding consists of summation of amplitude spectrum elements and elements of the digital watermarks.

In [23, 24] the image is divided into blocks 2 × 2 pixels and each of them is treated by DFT. The frequency coefficient in the top left corner of the amplitude spectrum is not changed, all others are embedded in the message bits according to LSB. In [23] in addition there is used genetic algorithm to improve quality.

It is worth pointing out the paper [25] where generalization of the Fourier transform is used—fractional Fourier transform. The hidden space is formed of complex digitals coefficients of this transform. The secret message (digital watermark) is also a sequence of complex values which real and imaginary parts are normally distributed and embedded additively.

One can see two ways suggested in these papers to come to the frequency field using DFT:

1. DFT is applied to the whole image;
2. DFT is applied to small 2 × 2 pixel blocks.

Both approaches have drawbacks. In the first case utilization of the small blocks does not allow to mark out middle frequencies in the Fourier spectrum, while this part is more suitable to embed the additional information. It makes frequency domain useless. In the second case variation of every Fourier image change all the stego-container image pixels. It amplifies the embedding artifacts.

In [26] the image is split into blocks 8 × 8, part of them are used for embedding. Unfortunately it is not taken into consideration that artifacts of different blocks could differ even if the volume of information is the same.

It is necessary to point out separately that other authors don't investigate the problem of an embedded information distortion which could appear during the back transfer from the frequency area into the space one. This problem appears because after application of the back Fourier transfer to a modified Fourier image there appear real values which are rounded when the pixel matrix is formed. As a result, the formed Fourier image got after DFT application differ from the one where information was embedded. So the information could be got incorrect.

In the paper there given are results of Fourier transform steganographic properties investigation in the context of the mentioned problem. There was formed an algorithm of embedding information into the phase spectrum. The choice of the phase image is due to the fact that phases belong to the interval $[-\pi, \pi]$ and don't depend on the concrete image. So there is no need of normalizing. The paper develops the previous investigation of the authors presented in [27].

2.2 *Investigation of the Discrete Fourier Transform Steganographic Properties*

As an elementary steganographic operation we suggest using in this paper the turn of the radius-vector at the complex area, where we call the radius-vector the geometric interpretation of a complex spectrum separate element.

In order to solve the problem mentioned in 2.1 and connected with the frequency spectrum distortion in the process of transfer from real values to digital ones it was necessary to investigate the properties of the introduced operation.

To do this we made some calculation experiments with standard test images. These calculations meant to show the interrelation between image characteristics in space and frequency areas and properties of steganographic operation. In the paper we don't give the full description due to the lack of the space.

The calculation experiments allowed to conclude:

1. The number of bits got correctly from the stego-image before the first mistake practically does not depend on the image size. Therefore in the process of building in the image it is reasonable to split the image into the blocks of small but trivial size: from 8 × 8 to 16 × 16 pixels.
2. Uniform blocks with similar number of pixels are generally less disturbed.
3. When two radius-vectors turn by the same angle, longer vector is more stable. Therefore when the same information volume is built in into two different pixel blocks more stable is the block with bigger sum of amplitude spectrum elements. These elements correspond to elements of the phase spectrum used.

Still up to now we have not managed to find the strong dependence of building in reliability on the image statistical characteristics. We can speak only about empiric conclusions true in many cases. In order to get more reliable result one should have redundancy while building in information into the frequency area of the digital image.

The corresponding algorithm with redundancy during calculations, which consists of iterative repetition during the process of building in, is shown below.

2.3 The Suggested Algorithm of Embedding Data into Digital Images

In order to realize the suggested steganographic operation parameters φ_0, φ_1, ε, defining the two not crossing intervals: the interval $(\varphi_0 - \varepsilon, \varphi_0 + \varepsilon)$ corresponding zero bit and the interval $(\varphi_1 - \varepsilon, \varphi_1 + \varepsilon)$ corresponding unity bit, should be given. We'll call these intervals "intervals of building in".

Input: stego-container—full-color or halftone digital image; secret information $L = l_1 l_2 \ldots l_n$, $l_i \in \{0, 1\}$; parameters of building in φ_0, φ_1, ε; maximal number of iterations with one block; threshold value of elements of frequency block spectrum A_{crit} average amplitude.

Output: stego-image.

Step 1: Go to the YCbCr color model and split matrix $\mathbf{Y}$ into not crossing blocks 8 × 8.

Step 2: Initialize counter of the built in bits $k \leftarrow 0$.

Step 3: While $k < n$ do the follow:

Step 3.1: Do DFT of the matrix $\mathbf{Y}$ next block 8 × 8 getting matrix of complex numbers $\mathbf{F} = \left(f_{ij}\right)_{i=1,j=1}^{8,8}$. Calculate matrix of amplitudes $\mathbf{A} = \left(a_{ij}\right)_{i=1,j=1}^{8,8}$ and phases $\mathbf{\Phi} = \left(\varphi_{ij}\right)_{i=1,j=1}^{8,8}$.

Step 3.2: Calculate average A_{mean} of elements a_{ij}, $i = \overline{2,7}$, $j = \overline{2,4}$, whose phases φ_{ij} are in one of building in intervals. If $A_{mean} < A_{crit}$ go to 3.3, otherwise to 3.1.

Step 3.3: Emphasize 18 bit long substring L' in the line L.

Step 3.4: Perform building in of secret bits. If the number of bits built in the block $q = 0$ go to 3.5, otherwise to 3.6.

Step 3.5: Treat the empty block.

Step 3.6: Calculate $k \leftarrow k + q$ and make $L \leftarrow l_{k+1} l_{k+2} \ldots l_n$.

Step 3.7: Make IDFT of the transformed block and modernize the values of the corresponding block in matrix $\mathbf{Y}$.

Step 4: Go to the RGB color model and finalize the algorithm.

Algorithms of matrix $\mathbf{Y}$ separate blocks processing used at steps 3.4 and 3.5 are investigated separately.

Algorithm of secret blocks building in.

Input: matrix of DFT complex coefficients is $\mathbf{F} = \left(f_{ij}\right)_{i=1,j=1}^{8,8}$; the corresponding matrix of phases is $\mathbf{\Phi} = \left(\varphi_{ij}\right)_{i=1,j=1}^{8,8}$; bit line is $L' = l_1 l_2 \ldots l_{18}$; maximal number of iterations is τ .

Output: matrix $\mathbf{F} = \left(f_{ij}\right)_{i=1,j=1}^{8,8}$ is the matrix with built in bits or the same as the initial matrix; the number of built in bits is q.

Step 1: Make matrixes $\mathbf{G} = \left(g_{ij}\right)_{i=1,j=1}^{8,8}$, $\mathbf{P} = (p_{uv})_{u=1,v=1}^{6,3}$, $p_{uv} = -1$, $u = \overline{1,6}$, $v = \overline{1,3}$. Write matrix $\mathbf{F}$ into matrix $\mathbf{G}$. Initialize counter of built in bits $q \leftarrow 0$.

Step 2: For all $i = \overline{2,7}$, $j = \overline{2,4}$ do as follows:

Step 2.1: If $\varphi_{ij} \notin (\varphi_0 - \varepsilon, \varphi_0 + \varepsilon)$ and $\varphi_{ij} \notin (\varphi_1 - \varepsilon, \varphi_1 + \varepsilon)$, take the next element of matrix $\mathbf{\Phi}$.

Step 2.2: $p_{i-1,j-1} \leftarrow l_{q+1}$, $q \leftarrow q + 1$.

Step 2.3: If φ_{ij} is in an interval corresponding the built in bit, go to 2.1.

Step 2.4: If $l_q = 0$, $\varphi_{ij} \leftarrow \varphi_0$, otherwise $\varphi_{ij} \leftarrow \varphi_1$.

Step 2.5: Turn radius-vectors corresponding elements f_{ij} and $f_{9-i,9-j}$ of matrix $\mathbf{F}$ taking DFT properties into consideration.

Step 3: Initialize counter of iterations $t \leftarrow 0$.

Step 4: Make matrix $\mathbf{F}$ IDFT getting the changed block of luminosity components $\mathbf{Y}$.

Step 5: Go to the RGB color model forming the block of pixels.

Step 6: Go to the YCbCr color model and perform DFT getting matrix $\mathbf{F} = \left(f_{ij}\right)_{i=1,j=1}^{8,8}$. Calculate matrix of phases $\mathbf{\Phi} = \left(\varphi_{ij}\right)_{i=1,j=1}^{8,8}$.

Step 7: Initialize the counter of errors $s \leftarrow 0$.

Step 8: For all $i = \overline{2,7}$, $j = \overline{2,4}$ do as follows:

Step 8.1: If $p_{i-1,j-1} = -1$ and $\varphi_{ij} \in (\varphi_0 - \varepsilon, \varphi_0 + \varepsilon)$ or $\varphi_{ij} \in (\varphi_1 - \varepsilon, \varphi_1 + \varepsilon)$, $f_{ij} \leftarrow g_{ij}$, $f_{9-i,9-j} \leftarrow g_{9-i,9-j}$, $s \leftarrow s + 1$ and take the next element of the matrix $\mathbf{\Phi}$.

Step 8.2: If $p_{i-1,j-1} = 0$ and $\varphi_{ij} \notin (\varphi_0 - \varepsilon, \varphi_0 + \varepsilon)$, $\varphi_{ij} \leftarrow \varphi_0$.

Step 8.3: If $p_{i-1,j-1} = 1$ and $\varphi_{ij} \notin (\varphi_1 - \varepsilon, \varphi_1 + \varepsilon)$, $\varphi_{ij} \leftarrow \varphi_1$.

Step 8.4: Turn radius-vectors corresponding elements f_{ij} and $f_{9-i,9-j}$ of the matrix $\mathbf{F}$ taking DFT properties into consideration and make $s \leftarrow s + 1$.

Step 9: If $s = 0$ go to 13, otherwise $t \leftarrow t + 1$.

Step 10: If $t < \tau$ go to 4.

Step 11: Write matrix $\mathbf{G}$ into matrix $\mathbf{F}$ and make $q \leftarrow q - 1$.

Step 12: If $q > 0, L' \leftarrow l_1 l_2 \ldots l_q$ and go to 1.

Step 13: Return matrix $\mathbf{F}$ and q back and finish the algorithm.

The case when L or L', during the operation, are shorter than 18 bits we don't consider to make the consideration easier.

Algorithm of an empty block processing

Input: matrix of the DFT complex coefficients $\mathbf{F} = \left(f_{ij}\right)_{i=1,j=1}^{8,8}$.

Output: changed matrix $\mathbf{F} = \left(f_{ij}\right)_{i=1,j=1}^{8,8}$.

Step 1: Make matrix $\mathbf{F}$ IDFT getting the changed block of luminosity components $\mathbf{Y}$.

Step 2: Go to the to the RGB color model forming the block of pixels.

Step 3: Go to the YCbCr color model and perform DFT getting matrix $\mathbf{F} = \left(f_{ij}\right)_{i=1,j=1}^{8,8}$. Calculate matrix of phases $\mathbf{\Phi} = \left(\varphi_{ij}\right)_{i=1,j=1}^{8,8}$.

Step 4: Initialize the counter of the "false" bits $s \leftarrow 0$.

Step 5: For all $i = \overline{2,7}$, $j = \overline{2,4}$ do as follows:

Step 5.1: If $\varphi_{ij} \in (\varphi_0 - \varepsilon, \varphi_0 + \varepsilon)$ or $\varphi_{ij} \in (\varphi_0 - \varepsilon, \varphi_0 + \varepsilon)$ generate a random value r out of the mentioned interval and make $\varphi_{ij} \leftarrow r$.

Step 5.2: Turn radius-vectors corresponding elements f_{ij} and $f_{9-i,9-j}$ of the matrix $\mathbf{F}$ taking DFT properties into consideration and make $s \leftarrow s + 1$.

Step 6: If $s = 0$ go to 7, otherwise to 1.

Step 7: Return matrix $\mathbf{F}$ and finalize the algorithm.

Algorithm of extraction is obvious and is not described here.

2.4 Experiments

In this part there are given results of some calculation using the suggested steganographic algorithm. All calculations we are done with 10 full-color images with 256 × 256 pixels (classic test images Airplane, Baboon, Goldhill, Lenna, Peppers and so on). The main parameters of building in were $\varphi_0 = -\frac{\pi}{2}$, $\varphi_1 = \frac{\pi}{2}$, $\varepsilon = 1$.

Figure 1a show the averaged dependence of the building in quality (PSNR) on the length of message for $\tau = 20$ and $A_{crit} = 0,8$. One can see that even if the volume of building in is big there are no distortions of stego-container.

Graphics in Fig. 1b show that PSNR decrease only slightly if the number of iterations grows.

The graphic in Fig. 1c got for $\tau = 20$ shows that the threshold of the block frequency elements average amplitude suitable for building in influence the quality of stego-image. Therefore $A_{crit} > 1.5$ are not recommended.

So, we can conclude that the suggested algorithm allows not only escape distortions but also provide high quality of building in.

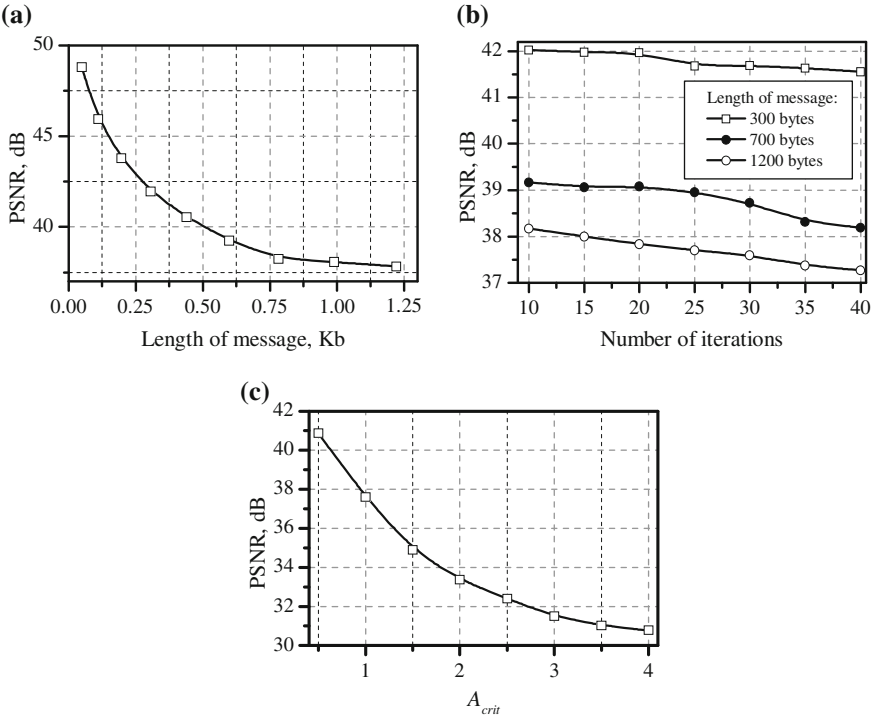


Fig. 1 Influence of the suggested algorithm parameters on the stego-image quality: length of message (**a**); number of iterations τ (**b**); threshold A_{crit} (**c**)

3 Conclusion

There were investigated steganographic properties of the discrete Fourier transform in case of building in information into the phase spectrum. As the basic steganographic operation there is used the turn of the radius-vector at the complex plane.

The main peculiarity of the investigation is solving the problem of information distortion while information is built in into the digital area of the digital image. Distortions appear in the process of the transform from the frequency area into the space and is due to rounding the real values to digital.

The problem mentioned is typical not only for the introduced steganographic operation but also for other operations connected with building in information into the frequency area of digital images. Unfortunately there are no publications on the subject.

We received algorithm that allows to get information without mistakes using unequal building in parts of information in image blocks and iterative procedure. Such characteristic as capacity and quality of building in the algorithm is comparable to the analogs described in 2.1.

Further investigation can make the process of building in more effective on the bases of optimization.

Acknowledgments This work was supported by the Ministry of Education and Science of the Russian Federation within 1.3 federal program « Research and development in priority areas of scientific-technological complex of Russia for 2014-2020 » (grant agreement № 14.577.21.0153 on November 28, 2014; identifier RFMEFI57714X0153).

References

1. Salomon, D.: Data Compression: The Complete Reference, 4th edn. Springer, London (2007)
2. Gorbachev, V.N., Kainarova, E.M., Denisov, L.A.: Embedding of Binary Image in the Gray Planes. Comput. Opt. **37**, 385–390 (2013)
3. Wang, Z.H., Chang, C.C., Li, M.C.: Optimizing least-significant-bit substitution using cat swarm optimization strategy. Inf. Sci. **192**, 98–108 (2012)
4. Kanan, H.R., Nazeri, B.: A novel image steganography scheme with high embedding capacity and tunable visual image quality based on a genetic algorithm. Expert Syst. Appl. **41**, 6123–6130 (2014)
5. Glumov, N.I., Mitekin, V.S.: A new semi-fragile watermarking algorithm for image authentication and information hiding. Comput. Opt. **35**, 262–267 (2011)
6. Verichev, A.V., Fedoseev, V.A.: Digital image watermarking on triange grid of feature points. Comput. Opt. **38**, 555–563 (2014)
7. Zhao, J., Koch, E.: Embedding robust labels into images for copyright protection. In: International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies, pp. 242–251. R. Oldenbourg Verlag, Munich (1995)
8. Holliman, M., Memon, N., Yeo, B.L., Yeung, M.: Adaptive public watermarking of DCT-based compressed images. Int. Soc. Opt. Eng. **3312**, 284–295 (1998)
9. Westfeld, A.: High capacity despite better steganalysis (F5-A Steganographic Algorithm). In: 4th International Workshop on Information Hiding, pp. 289–302. Springer, London (2001)
10. Yu, L., Zhao, Y., Ni, R., Zhu, Z.: PM1 steganography in JPEG images using genetic algorithm. Soft. Comput. **13**, 393–400 (2009)
11. Huang, F., Huang, J., Shi, Y.Q.: New channel selection rule for JPEG steganography. IEEE Trans. Inf. Forensics Secur. **7**, 1181–1191 (2012)
12. Ejaz, N., Anwar, J., Ishtiaq, M., Baik, S.W.: Adaptive image data hiding using transformation and error replacement. Multimedia Tools Appl. **73**, 825–840 (2013)
13. Li, F., Zhang, X., Yu, J., Shen, W.: Adaptive JPEG steganography with new distortion function. Ann. Telecommun. **69**, 431–440 (2014)
14. Evsutin, O.O., Kokurina, A.S., Shelupanov, A.A., Shepelev, I.I.: An improved algorithm for data hiding in compressed digital images based on PM1 method. Comput. Opt. **39**, 572–581 (2015)
15. Soliman, M.M., Hassanien, A.E., Onsi, H.M.: An adaptive watermarking approach based on weighted quantum particle swarm optimization. Neural Comput. Appl. **27**, 469–481 (2016)
16. Rekik, S., Guerchi, D., Selouani, S.A., Hamam, H.: Speech steganography using wavelet and fourier transforms. EURASIP J. Audio Speech Music Process. **20**, 14 (2012)
17. Natgunanathan, I., Xiang, Y., Rong, Y., Peng, D.: Robust patchwork-based watermarking method for stereo audio signals. Multimedia Tools Appl. **72**, 1387–1410 (2013)
18. Cedillo-Hernandez, M., Garcia-Ugalde, F., Nakano-Miyatake, M., Perez-Meana, H.: Robust watermarking method in DFT domain for effective management of medical imaging. SIVIP **9**, 1163–1178 (2015)

19. Solachidis, V., Pitas, I.: Circularly symmetric watermark embedding in 2-D DFT domain. IEEE Trans. Image Process. **10**, 1741–1753 (2001)
20. Poljicak, A., Mandic, L., Agic, D.: Discrete fourier transform-based watermarking method with an optimal implementation radius. J. Electron. Imaging **20**, 033008-1–033008-8 (2011)
21. Ridzon, R., Levicky, D.: Content protection in grayscale and color images based on robust digital watermarking. Telecommun. Syst. **52**, 1617–1631 (2013)
22. Rabie, T.: Digital Image Steganography: An FFT Approach. In: Benlamri, R. (ed.) NDT 2012. Communications in computer and information science, vol. 294, pp. 217–230. Springer, Heidelberg (2012)
23. Mandal, J.K., Khamrui, A.: A Genetic Algorithm Based Steganography in Frequency Domain (GASFD). In: International Conference on Communication and Industrial Application, pp. 1–4, IEEE (2011)
24. Bhattacharyya, D., Kim, T.: Image Data Hiding Technique Using Discrete Fourier Transformation. In: Kim, T., Adeli, H., Robles, R.J., Balitanas, M. (eds.) UCMA 2011. Communications in computer and information science, vol. 151, pp. 315–323. Springer, Heidelberg (2011)
25. Savelonas, M.A., Chountasis, S.: Noise-resistant watermarking in the fractional fourier domain utilizing moment-based image representation. Sig. Process. **90**, 2521–2528 (2010)
26. Chen, W.Y.: Color image steganography scheme using DFT, SPIHT codec, and modified differential phase-shift keying techniques. Appl. Math. Comput. **196**, 40–54 (2008)
27. Evsutin, O.O., Kozlova, A.S., Meshcheryakov, R.V., Kokurina, A.S.: Information embedding into the phase spectrum of the discrete fourier transform. In: 25th International Crimean Conference on Microwave & Telecommunication Technology, pp. 371–372, Sevastopol (2015)

Decremental and Incremental Reshaping of Algebraic Bayesian Networks Global Structures

Daniel G. Levenets, Mikhail A. Zotov, Artem V. Romanov, Alexander L. Tulupyev, Andrey A. Zolotin and Andrey A. Filchenkov

Abstract The paper considers algorithms for global structures generation in algebraic Bayesian networks. A decremental algorithm for constructing a secondary structure after deleting vertex from the adjacency graph is proposed supplemented by a listing of the algorithm code and by the proof of its correctness. The results of the statistical tests for decremental algorithm are proposed graphically together with a comparative analysis of the results. Moreover, the description of incremental algorithm for adding vertex in tertiary structure is provided supplemented by a listing of the algorithm code and proof of its correctness.

Keywords Joint graph · Probabilistic graphical model · Incremental algorithm · Performance statistical estimate · Structure learning · Machine learning

D.G. Levenets (✉) · M.A. Zotov · A.V. Romanov · A.L. Tulupyev · A.A. Zolotin
St. Petersburg State University, Saint Petersburg, Russia
e-mail: daniellevenets@gmail.com

M.A. Zotov
e-mail: zotov1994@mail.ru

A.V. Romanov
e-mail: exxemka.spb@gmail.com

A.L. Tulupyev
e-mail: alexander.tulupyev@gmail.com

A.A. Zolotin
e-mail: andrey.zolotin@gmail.com

D.G. Levenets · M.A. Zotov · A.V. Romanov · A.L. Tulupyev · A.A. Zolotin
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 39, 14-th line Vasilyevsky Ostrov, Saint Petersburg 199178, Russia

A.A. Filchenkov
ITMO University, 49, Kronverksky Pr, Saint Petersburg 197101, Russia

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_6

1 Introduction

Probabilistic graphical models (PGM) represent the graph in which nodes correspond to random elements, and edges represent dependencies between them. Algebraic Bayesian networks (ABN) related to Bayesian belief networks, being one of the representatives of the PGM class and having a primary, secondary, tertiary and quaternary structures, inherit the characteristics given above. One of the advantages of PGM is the possibility to decompose [3, 12] large amounts of data into smaller ones, that enables to speed up the processing of the information received. Being a field of artificial intelligence, the theory of PGM undoubtedly requires machine learning algorithms, one of which (the synthesis of global structures) is discussed in details in this paper. A global structure may be represented by a join tree or a join graph [1]. However, the problem of synthesizing such a structure is complex, which significantly slows down the work with dynamically changing data. Decremental and using similar techniques incremental algorithms discussed in this article are used to reconstruct the global structure [2, 4, 11] allow us to dynamically change the structure of the join graph without rebuilding the graph completely.

The major goal of the paper is to describe the decremental algorithm. Furthermore, the algorithm is compared to the greedy algorithm to outline its advantages. Details related to incremental algorithms are not given as they were published in [14], but the paper is extended by description of the incremental algorithm for synthesis of tertiary structure. We demonstrate that incrementation is not only applicable to the synthesis of a separate minimal joint graph (MJG), but also to the synthesis of a complete set of MJG, which is also a step forward.

2 Background

2.1 Problem Statement

PGMs are generative models representing domain knowledge. They usually work with priori and posteriori (given evidence) information. This is a very different situation to the most of other machine learning models. The typical scenario of supervised learning model consists of the two steps: learning step and application step. The learning step requires a set of training samples with known target labels: $D_{train} = \{(o_1, y_1), \ldots, (o_n, y_n)\}$, where o_i is an object and y_i is its label. The model is learnt on the training set and is used to predict label y_{new} of new object o_{new}, which is described with the same feature set.

In contrast, PGMs do not usually focus on a target variable. Despite the fact that the learning dataset may be the same, namely D_{train}, no particular difference is made between y and variables which describe o. If o is a vector $(x_1, \ldots, x_m)$, than y may

be simply incorporated as x_0. Moreover, the main distinction is in application step: as an input the learnt generative model receives an evidence, which contains information on a subset of variables. Thus, it can be understood as a partial description of a new object. The problem being usually solved is to predict information on other variables of the object. More precisely, we receive information $I(X^e)$ of $X^e \subset X$ and want to estimate $I(X|X^e)$, where I is a function that represents information with uncertainty on a subset of variables. ABNs use a family of probability distributions on the variable set, while Bayesian belief networks use a single probability distribution, and fuzzy Bayesian networks can work with fuzzy distributions.

Usually, BN learning problem is stated as selection of model from model space $\mathcal{M}$, maximizing training data likelihood penalized for model complexity:

$$M^* = \arg\max_{M \in \mathcal{M}} (\Pr(D_{train}|M) - \lambda \cdot \text{Complexity}(M)),$$

which reflects only training data representation quality, not evidence processing. To avoid this, the problem should be stated in the following way:

$$M^* = \arg\max_{M \in \mathcal{M}} (\Pr(D_{train}|M) - \lambda \cdot \text{Complexity}(M) + \kappa \cdot \mathbf{E}Q(\varepsilon; M)),$$

where $Q(e; M)$ is a quality measure for processing evidence e by model M and ε is a random element defined on evidence space.

ABNs, as well as other PGMs, consist of several so-called structures. Global structures reflect relations on variables and variable sets, while local structure relates to parametrization of probability distributions family. For this problem statement an interesting problem arises: if our considerations on probability distribution for ε change, another model will achieve maximum, even when the training set does not change. In general, this would require to retrain the model, which is extremely computationally. Fortunately, when only $\mathbf{E}Q(\varepsilon; M)$ changes, we do not need to rebuild all the structures: the only structure which is required to be rebuilt is the secondary structure [10].

Several algorithms for secondary structure synthesis exist, however, most of them require the secondary structure to be completely rebuilt. To suggest an algorithm which can rebuild an existing secondary structure, we need to introduce formal definitions for secondary structure, which is represented by a join graph.

2.2 Definitions and Notations

We will use terms and notations formed in the [6–10].

Loaded graph on alphabet A is a triple $\langle V, E, W \rangle$, where V is vertex set, E is edge set and W is **load function**, which assigns to each vertex a **load**, a non-empty subset of A. Let W_u denote load of vertex u The context of the paper allows to

equate vertices and their loads for ease as the mutually correlated ones, so we will use the notation u and W_u interchangeably.

Vertices u and v are **separable** if their separator is not empty. Two separable vertices are called **backbone connected** in the graph G if such a path P exists, that for each vertex s in P condition $W_u \cap W_v \subseteq W_s$ is satisfied.

Undirected loaded graph is a **join graph** if each its edge connects separable vertices, and any two separable vertices are backbone connected. Join graph with the minimum and maximum number of edges are called **minimal join graph** and **maximal join graph** respectively.

Separator $s_{u,v}$ is non-empty intersection of loads of two vertices $W_u \cap W_v$. Separator can be understood as a load of possible edge between these vertices. Let **Sep** denote set of separators for the graph G.

Tertiary structure is a directed loaded graph $\langle \mathrm{Sep}, E_{\mathrm{III}}, W \rangle$ over the separator set represented with a Hasse diagram, inverted upside down. Edge between two vertices a and b in tertiary structure exists, iff $W_a \cap W_b = W_a$ and $|W_a| < |W_b|$, where $\nexists c \in \mathrm{S}\, W_c \cap W_a = W_c$ and $|W_a| < |W_b| < |W_c|$. **Parent** of a separator in the tertiary structure is a separator preceding to it. **Son** of a separator in the tertiary structure is a separator next to it.

We also require additional notation for the algorithms. $\mathrm{PathExists}(G, e, u, v)$ is function that determines whether vertices u and v are backbone connected in G after removal of its edge e [5, 6, 15]. For edge e in G e.First, e.Second denotes two vertices of the edge and e.RemoveAllowed is a tag that indicates whether removal of e will persist keep the graph to be a join graph. Its initial value is true.

3 Decremental Algorithm for Removing the Vertex

As the input, the algorithm receives a minimal join graph $G = \langle V, E, W \rangle$ and one of its vertices $v \in V$, to be removed. $G' = \langle V \backslash \{v\}, E'.W \rangle$, where E' is a set of edges of the minimal join graph G'. Decremental algorithm takes a set of edges E as a basis for the minimum join graph G'. We assumed that the decremental algorithm will work faster than a greedy (direct) one since half or even most of the edges will remain unchanged.

3.1 Algorithm Description

In the pseudo-code of listing 1 a decremental algorithm for removing vertex from a minimum join graph is provided. The algorithm consists of two parts. In the first part (lines 2–19) a join graph is formed together with a set of vertices $V \backslash v$ and a set of edges, consisting of a set of edges of the initial graph without edges incident to the vertex v as well as of all possible edges between the vertices of such edges (without v). In the second part the edges the lack of which will not lead to

disruption of the main links are removed from the secondary structure of the above-described ABN (lines 20–31).

Let us consider the algorithm in more detail. In lines 4–12 set E'' consisting of the edges incident to the vertex v is formed (lines 9, 12), as well as a set of vertices V'', consisting of the vertices of the edges E'', but not including a vertex v (line 8, 11). Then all possible edges between the vertices from the set V'' (line 16), load intersections of which are not empty (line 15) are added in the set E'. After this the graph $G' = \langle V', E' \rangle$ (line 19) is formed and the edges absence of which will not lead to disruption of the backbone paths must be removed therefrom (lines 19–30). As a result of this transformation, graph G' becomes a minimal join graph. Let us note that the edges are removed by the greedy algorithm described in papers [5, 6, 13].

3.2 Algorithm Correctness

When adding an allowable edge to a minimum join graph, a new join graph is obtained, while removal of a vertex and edges incidence to it followed by addition of a acceptable clique on its place is a transition from one join graph to another with a smaller number of vertices. Therefore, after the first part of the algorithm, the graph G' turns into a join graph, while the admissibility of the edges is confirmed by the test carried out in line 15.

Algorithm 1 Decremental algorithm for removing the vertex from a minimal join graph.

input: $\mathrm{G} = \langle \mathrm{V}, \mathrm{E}, \mathrm{W} \rangle$, $\mathrm{v} \in \mathrm{V}$
output: $\mathrm{G}' = \langle \mathrm{V}', \mathrm{E}', \mathrm{W} \rangle$

```
 1: function DeleteIncremental
 2:     E' = E
 3:     V' = V
 4:     E'' = ∅
 5:     V'' = ∅

 6:     foreach (e in E)
 7:         if (e.First == v) then
 8:             V'' = V'' ∪ {e.Second}
 9:             E'' = E'' ∪ {e}
10:         if (e.Second == v) then
11:             V'' = V'' ∪ {e.First}
12:             E'' = E'' ∪ {e}

13:     foreach (x in V'')
14:         foreach (y in V'')
15:             if ((W_x ∩ W_y) ≠ ∅) then
16:                 E' = E' ∪ {x, y}
```

```
17:     E′ = E′ \ E″
18:     V′ = V′ \ {v}
19:     G′ = ⟨V′, E′⟩

20:     while (true)
21:         e* = NaN

22:         foreach(e in E′)
23:             if (e.RemoveAllowed) then
24:                 e* = e
25:                 break foreach //Exit from foreach

26:         if (e* == NaN) then
27:             break while //Exit from while

28:         if (PathExists(G′, e*, e*.First, e*.Second)) then
29:             E′ = E′ \ {e*}
30:         else
31:             e*.RemoveAllowed = false

32:     return G′ = ⟨V′, E′⟩
```

Removal of the edges maintaining the graph being backbone is nothing else than the transition from one join graph to another with a decrease in the number of edges. In papers [5, 6] it was proved that the set of join graphs forms the matroid. According to the matroid properties, the sequence of permissible deletions of edges in the join graph is convergent, thus transforming initial graph into the minimum join graph. As shown previously, after executing the first part of the algorithm, the graph G' turns into a join graph, while a valid removal of the edges takes place in the second part of the algorithm. Thus, according to the properties of a matroid, the algorithm shown in listing 1 generates minimal join graph that does not include the vertex v, Q.E.D.

4 Statistical Experiments

In order to ensure that the incremental algorithm is faster than the greedy one, we conduct computational experiments, where the runtime of minimal join graph synthesis by decremental and greedy algorithms respectively is measured. The detailed methodology of computational experiments was described in [15], therefore let us confine ourselves to a brief description of the calculated statistics.

On Fig. 1 one can find the following statistics: $R-, R+$—the lower and upper limits of the confidence interval, $Q3$, $Q1$—the third and the first quartile and finally Rg—the average. The alphabet with a 64 characters capacity was chosen for load generation of graph vertices. The horizontal axis displays the number of nodes in a

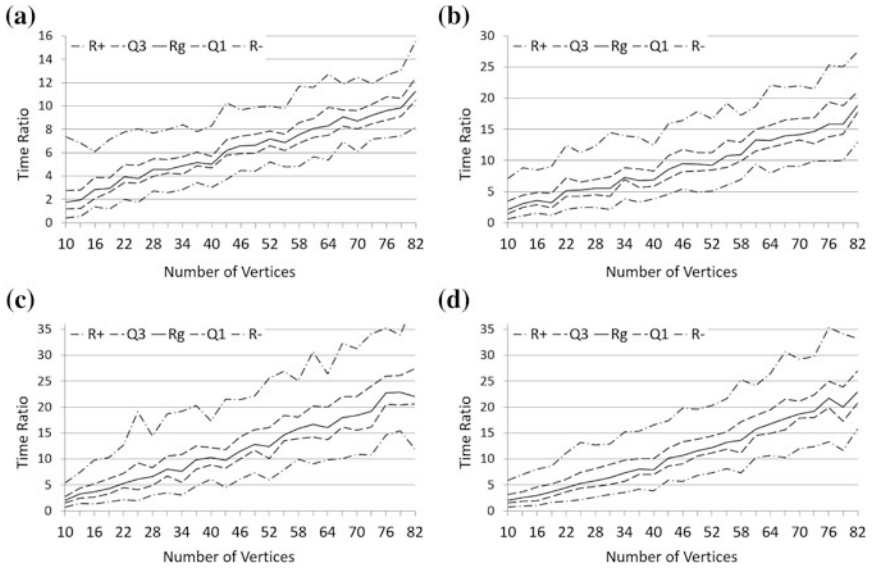


Fig. 1 The results statistical experiments. **a** vertices with 2–4 symbols loads—80 %, 5–7—15 %, 8–10—5 %, 11–13—0 %; **b** vertices with 2–4 symbols loads—70 %, 5–7—15 %, 8–10—10 %, 11–13—5 %; **c** vertices with 2–4 symbols loads—50 %, 5–7—25 %, 8–10—15 %, 11–13—10 %; **d** vertices with 2–4 symbols loads—30 %, 5–7—55 %, 8–10—15 %, 11–13—10 %

graph and a vertical—the statistical value (the ratio of a runtime of greedy algorithm to decremental).

Analysis of the computational experiments results leads us to the conclusion that the decremental algorithm is much faster than the greedy algorithm in the subrange of 22–82 vertices, while in the subrange 10–19 a transitional state takes place (as evidenced by the lower limit of the confidence interval), when the decremental algorithm may be slower than the greedy one. However, let us note that the time of minimal join graph synthesis for 10–19 vertices is negligible.

5 Incremental Tertiary Structure Processing Algorithm

Let us consider the case of the incremental adding of the vertex in the ABN tertiary structure. Given a set of vertices S and tertiary structure of the ABN constructed on a set of non-empty separators Sep.

Adding a new vertex u we have three options:

1. $\forall v \in S\, W_u \cap W_v = \emptyset$
2. $\forall v \in S\, W_u \cap W_v \in$ Sep or $W_u \cap W_v = \emptyset$
3. $\exists v \in S\, W_u \cap W_v \notin$ Sep and $W_u \cap W_v \neq \emptyset$

In the first case, the added vertex does not generate separators, hence tertiary structure remains unchanged. In the second case, the separators obtained by adding new vertices are already in the set of separators Sep the tertiary structure is based on, and thus the tertiary structure contains them as well.

Let us consider the third case where new separators that do not belong to the set Sep are produced by intersections of loads of newly added vertex with the loads of vertices of set V.

5.1 Algorithm Description

The pseudo-change for reconstruction of the tertiary structure when adding a new piece of knowledge in ABS is provided in listing 5.2. The algorithm input is a constructed tertiary structure, a set of primary structure vertices and a new vertex. To change the tertiary structure first it is necessary to allocate the new separators, generated by added vertex (this is carried out in lines 3–5) and then to get a set Sep_u that contains all the new separators.

Next, for each element s of the set Sep_u (line 6–31) we are to build a set of parents (line 7) and a set of sons (line 8), containing elements of set S, which should be linked by an edge with s in a tertiary structure. A set of parents for the vertex s is constructed as follows: we consider only vertices c from the set S, loads of which are contained in the load of s (line 13), then we remove from the set of parents all the elements, which loads are contained in the W_c (lines 14–16), and if there is no such element in the set of parents whose load contains W_c (line 17–21) then we add the element c to the set of parents.

Similarly the set of sons is constructed, the difference is that it will contain elements loads of which contain W_s (line 23–31). So we obtained the set of parents, containing tertiary structure vertices of the highest power, loads of which are contained in the W_s, and the set of sons, containing tertiary structure vertices of the lowest power, loads of which contain load W_s.

The next step is to remove the edges existing in tertiary structure between elements of the set parents and sons, since they are connected to each other through a separator s (line 32–35). Next, we add a new separator (line 36) and the new edges (lines 37–41) in a set of vertices and edges of the tertiary structure respectively.

5.2 Proof of Algorithm Correctness

A graph $H = (S, E_H, W)$ is a tertiary structure, iff the following conditions are satisfied: (1) sets Sep and S coincide elementwise; (2) there is an edge between two vertices a and b iff $W_a \cap W_b = W_a$ and $\nexists c \in S : \{W_c \cap W_a = W_c \text{ and } |W_c| > |W_a|\}$.

As the input, the algorithm receives a graph satisfying the requirements given above. Let us show that the structure, obtained at the output of the algorithm, will

be a tertiary structure. The first condition is satisfied, because all separators which are not contained in the set of S (the set Sep_u) are added in line 36. The second condition is also fulfilled, since the sets of sons and parents are constructed according to the rule described in it and after that all the edges that are not already in the set E_H are added and those edges which are linked through the added separators are removed. Thus, the algorithm is valid and enables one to obtain a tertiary structure when adding a new knowledge pattern in the ABN primary structure.

Algorithm 2 Incremental algorithm for the tertiary structure rebuilding when adding a new knowledge pattern.

input: $\mathrm{H} = \langle \mathrm{S}, \mathrm{E_H} \rangle$, $\mathrm{V''}$, u
output: $\mathrm{H} = \langle \mathrm{S'}, \mathrm{E'} \rangle$

```
 1: function AddSeparators
 2:     Sep_u = ∅
 3:     foreach (v in V'')
 4:         if ((W_v ∩ W_u ≠ ∅) and (W_v ∩ W_u not in V)) then
 5:             Sep_u = Sep_u ∪ {(W_v ∩ W_u)}

 6:     foreach (s in Sep_u)
 7:         parents = ∅
 8:         sons = ∅
 9:         foreach (c in S)
10:             current = W_s ∩ W_c
11:             if (current ≠ ∅) then
12:                 f = true
13:                 if (current == W_c) then
14:                     foreach (p in parents)
15:                         if (W_p ∩ W_c == W_p) then
16:                             parents = parents \ {p}
17:                         else
18:                             if (W_p ∩ W_c == W_c) then
19:                                 f = false
20:                     if (f) then
21:                         parents = parents ∪ {W_c}
22:                 else
23:                     if (current == W_s) then
24:                         foreach (son in sons)
25:                             if (W_son ∩ W_c == W_c) then
26:                                 sons = sons \ {son}
27:                             else
28:                                 if (W_son ∩ W_c == W_son) then
29:                                     f = false
30:                         if (f) then
31:                             sons = sons ∪ {W_v}
```

```
32:         foreach (p in parents)
33:             foreach (son in sons)
34:                 if ((p, son) in E_H) then
35:                     E_H = E_H \ {(p, son)}
36:         S = S ∪ {s}
37:         foreach (p in parents)
38:             E_H = E_H ∪ {(p, s)}
39:         foreach (son in sons)
40:             E_H = E_H ∪ {(s, son)}
41:     return H
```

6 Conclusion

Paper demonstrates the application of decrementalisation and incrementalisation of algorithms necessary for the synthesis of the ABN secondary structure as a stage of these the network learning. First, the incremental algorithm for removing the vertex from the minimal join graph is considered as well as its code listing and the result justification. Moreover, the collected statistics shows that decrementalisation has a positive effect—time of MJG construction is significantly smaller in the range of 22–82 vertices and that is confirmed by charts. On the other hand, the approach to the implementation of one step of the reverse operation is also considered –the incremental algorithm for tertiary structure construction in case of adding a new knowledge pattern to the primary structure extended with an algorithm code listing and with a proof. This result is new, it is expected to accelerate the synthesis of the entire set of MJG. The proposed algorithms provide the basis for development and improvement of software systems for representation, processing and visualization of ABN.The paper presents results of the project partially supported with RFBR grant 15-01-09,001-a “Combined Probabilistic-Logic Graphical Approach to Representation and Processing of Unsertain Knowledge Systems: Algebraical Bayesian Networks and Related Models”, and partially supported by the Government of the Russian Federation, Grant 074-U01.

References

1. Cano, A., Gomez-Olmedo, M., Moral, S.: Binary probability trees for bayesian networks inference. Symbo. Quant. Approaches Reasoning uncertainty Proc. **5590**, 180–191 (2009)
2. Flores, M., Gamez, J., Olesen, K.: Incremental compilation of bayesian networks based on maximal prime subgraphs. Int. J. Uncertainty Fuzziness knowl.-Based Syst. **19**(2), 155–191 (2011)
3. Leonelli, M., Smith, J.: Bayesian decision support for complex systems with many distributed experts. Ann. Oper. Res. **235**(1) (2015)

4. Nepomniaschaya, A.: Efficient parallel implementation of the ramalingam decremental algorithm for updating the shortest paths subgraph. Comput. Inform. **32**(2), 331–354 (2013)
5. Oparin, V.V., Filchenkov, A.A., Sirotkin, A.V., Tulupyev, A.L.: Matroidal representation for the adjacency graphs family built on a set of knowledge patterns. Sci. Tech. J. Inf. Technol. Mech. Opt. (4), 73–76 (2009). http://ntv.ifmo.ru/file/article/697.pdf (In Russian)
6. Oparin, V.V., Tulupyev, A.L.: Synthesis of a joint graph with minimal number of edges: an algorithm formalization and a proof of correctness. Tr. SPIIRAN (11), 142–157 (2009).http://mi.mathnet.ru/trspy52 (In Russian)
7. Tulupyev, A.L.: Join tree with conjunctions ideals as an acyclic algebraic bayesian network. Tr. SPIIRAN (3), 198–227 (2006). http://mi.mathnet.ru/trspy222 (In Russian)
8. Tulupyev, A.L.: Algebraic bayesian networks: global probabilistic-logic inference in adjacent tree (2007) (In Russian)
9. Tulupyev, A.L., Nikolenko, S.I., Sirotkin, A.V.: Bayesian Networks: a Probabilistic-logic Approach. SPb: Nauka (2006) (In Russian)
10. Tulupyev, A.L., Sirotkin, A.V., Nikolenko, S.I.: Bayesian Belief Networks. SPb: SPbSU Press (2009) (In Russian)
11. Yue, K., Fang, Q., Wang, X., Li, J., Liu, W.: A parallel and incremental approach for data-intensive learning of bayesian networks. IEEE Trans. Cybern. **45**(12) (2015)
12. Zhu, M., Liu, S.: A decomposition algorithm for learning bayesian networks based on scoring function. J. Appl. Math.
13. Zotov, M.A., Tulupyev, A.L., Sirotkin, A.V.: Complexity statistical estimates of straightforward and greedy algorithms for algebraic Bayesian networks syntesis. Nechetkie Sistemy i Myagkie Vychisleniya [Fuzzy Systems and Soft Computing] **10**(1), 75–91 (2015) (In Russian)
14. Zotov, M., Levenets, D., Tulupyev, A., Zolotin, A.: Synthesis of the secondary structure of algebraic bayesian networks: an incremental algorithm and statistical estimation of its complexity. Tr. SPIIRAN **16**(1), 122–132 (2016) (In Russian)
15. Zotov, M., Tulupyev, A.: Secondary structure synthesis in algebraic bayesian networks: a statistical technique for complexity estimates and comparative analysis of greedy and straightforward algorithms. Comput. Tools Educ. **1**, 3–18 (2015) (In Russian)

Algebraic Bayesian Networks: Local Probabilistic-Logic Inference Machine Architecture and Set of Minimal Joint Graphs

Ekaterina A. Mal'chevskaya, Alexey I. Berezin, Andrey A. Zolotin and Alexander L. Tulupyev

Abstract The paper considers a C# software package that implements algorithms of the local probabilistic-logical inference in algebraic Bayesian networks and the synthesis of their secondary structure. The package supports local network consistency verification and maintenance, priori and posteriori inference operations. In addition, two assembly algorithms for generating the set of minimal joint graphs and a usage example for one of them are presented. These algorithms provide algebraic Bayesian networks visualisation and are intended for structured entry data representation in GUI as well as for global evidence propagation and other probabilistic-logic inference operations implementation based on their local homologues.

Keywords Algebraic Bayesian networks · Probabilistic-logic inference · Secondary structure synthesis · Assembly algorithms · Minimal joint graph

E.A. Mal'chevskaya (✉) · A.I. Berezin · A.A. Zolotin · A.L. Tulupyev
St. Petersburg State University, Saint Petersburg, Russia
e-mail: katerina.malch@gmail.com

A.I. Berezin
e-mail: beraliv.spb@gmail.com

A.A. Zolotin
e-mail: andrey.zolotin@gmail.com

A.L. Tulupyev
e-mail: alexander.tulupyev@gmail.com

E.A. Mal'chevskaya · A.I. Berezin · A.A. Zolotin · A.L. Tulupyev
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 39, 14-th line Vasilyevsky Ostrov, Saint Petersburg 199178, Russia

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_7

1 Introduction

Increasing amounts of information and its uncertainty spur the development of mathematical models and methods to process the corresponding data volumes. Algebraic Bayesian networks (ABN) [9] are a new example of probabilistic graphical models (PGM). These models are better known with another well-known and extensively used example that is Bayesian belief networks; they serve as a representation of knowledge bases with uncertainty [2, 5, 6]. In contrast to other PGM, ABN are intended for systematic processing of both scalar and interval estimates of truth value probabilities. A probabilistic-logic inference machine is proposed over the ABN structure described previously [10], allowing both local priori and a posteriori inference in a concerned knowledge pattern [7], and similar types (or homologues) of global inference over the secondary structure [1]. In the case of global inference, the secondary structure construction time plays an important role. The purpose of this article is to consider previously obtained theoretical results on the local probabilistic-logic inference and to present the architecture of the new software complex implementing these results as well as minimum joint graph synthesis algorithms.

2 Probabilistic-Logic Inference

2.1 The Tasks of Probabilistic-Logic Inference in ABN

An ABN can be represented as an undirected graph with ideals of conjuncts associated with the graph vertices. Conjuncts like other propositional formulae are defined over a fixed alphabet, at the same time scalar or interval probability estimate is attributed to every conjunct. Let us call an ideal of conjuncts with probability estimates a knowledge pattern (KP). Note that a knowledge pattern can be constructed not only over the ideal of conjuncts, but also over the ideal of disjuncts or quanta propositions [8].

There are three tasks of local probabilistic-logic inference (P-LI) in ABN: the local priori inference, the first and the second tasks of local posteriori inference. Next, let us consider the tasks in detail, as well as introduce the basic definitions of KP consistency and reconcilability [7].

The conjunct is the conjunction of some number of atomic variables of the form $x_{i_1} \wedge x_{i_2} \wedge \cdots \wedge x_{i_k}$.

Quantum over the alphabet $A = \{x_i\}_{i=0}^{n-1}$ is a conjunction which contains either the atoms or its negation for every atom of the alphabet.

Suppose given a KP with scalar probability estimates (C, p). We say that it is *consistent* and denote it by Consistent$[(C, p)]$ iff there is a probability p_F, defined over a set of propositional formulae $F(A)$ (where A is the alphabet, which the KP is built over) such that

$$\forall c \in C : p_F = p(c)$$

Suppose given a KP with interval probability estimates $(C, \mathbf{p})$. We say that it is *consistent*, and denote it as Consistent$[(C, \mathbf{p})]$ iff for every conjunct $c \in C$ and every $\varepsilon \in \mathbf{p}(c)$ there is a function $p_{c,\varepsilon} : C \to [0, 1]$ such that $p_{c,\varepsilon} = \varepsilon$ and $(C, p_{c,\varepsilon})$ is consistent in terms of previous definition.

Suppose given a KP with interval probability estimates $(C, \mathbf{p})$. We say that it is *reconcilable* iff there is a consistent KP with interval probability estimates $(C, \mathbf{p}')$ such that

$$\forall c \in C : \mathbf{p}'(c) \subseteq \mathbf{p}(c)$$

The local priori inference task is to build the truth values of propositional formulae defined over the same alphabet as a KP based on a consistent KP.

Before giving a definition of evidence propagation tasks let us represent the concept of evidence used in these tasks.

Evidence is new "conditional" data that arrived in KP, and taking that into account we need to reconsider all (or only some) probability estimates. The evidence is represented as $\langle \cdots \rangle$.

The first task of the posterior inference is to estimate a probability of the evidence, given KP elements probability estimates.

The second task of the posterior inference is to evaluate the conditional probabilities of KP elements assuming that an evidence has arrived.

The theory of ABN considers 3 types of evidences: deterministic, stochastic and interval.

Deterministic evidence is the assumption that a conjunction of atoms or negotiation of atoms appears to be truth. This evidence is denoted by $\langle c_i, c_j \rangle$. Here c_i, c_j are the conjuncts which composed of atoms(without negation) and the negation of atoms respectively.

Stochastic evidence is the assumption that consistent KP with scalar probability estimates that determine the corresponding subideal elements probabilities is set over C' which is the subideal of C. This evidence is denoted by $\langle (C', \mathbf{P_c}) \rangle$.

Interval evidence is the assumption that consistent KP with interval probability estimates which determine the corresponding subideal elements probabilities is set over C' which is the subideal of C. This evidence is denoted by $\langle (C', \mathbf{P_c^-}, \mathbf{P_c^+}) \rangle$.

All the tasks above form a local probabilistic-logic inference in ABN, which was implemented in the software package further described.

2.2 Software Package Architecture

Designing the architecture of the software package, we paid a particular attention to the structure of the classes and interfaces that are responsible for the creation, storage and processing of KP. Three types of KP are considered in software

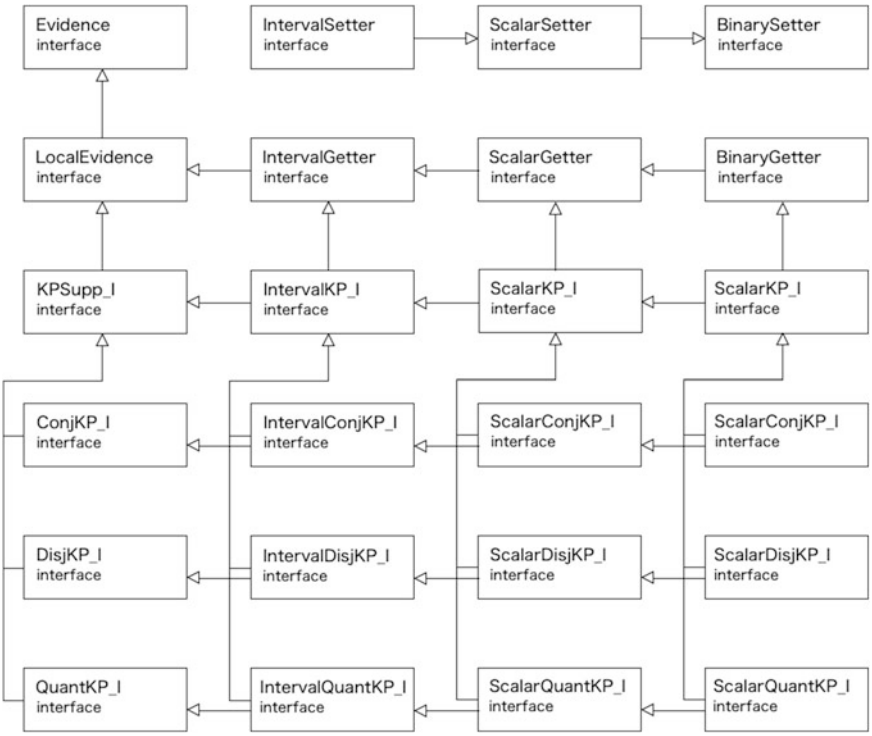


Fig. 1 Interfaces diagram

package: interval, scalar and binary. It has been observed that a binary KP is a particular case of a scalar KP and a scalar KP is a particular case of the interval KP. The observation gave rise to two inheritance chains. The first chain is the inheritance of getters, methods that read KP probability estimates, where a binary KP might be read as a scalar KP while scalar KP might be read as an interval KP. The second chain is the inheritance of setters, methods that set the KP probability estimates. The structure of interfaces interconnections which aggregates observations mentioned above, shown in Fig. 1.

The software package has a separate inferrers structure, responsible for the KP consistency checking and maintainance. It also lets convert a KP over an ideal of conjuncts into a KP over quanta propositions. The Fig. 2 shows the inferrers structure classes and interfaces dependencies.

The propagators structure in the software package is responsible for the local posteriori P-LI, that is to solve the first and second tasks of the evidence propagation. Methods described in the inferrers structure are used to maintain the consistency of the KP after evidence is propagated, corresponded to the DRY (Do not Repeat Yourself) principle. The Fig. 3 shows the dependence of classes and interfaces, responsible for the deterministic evidences propagation from each other.

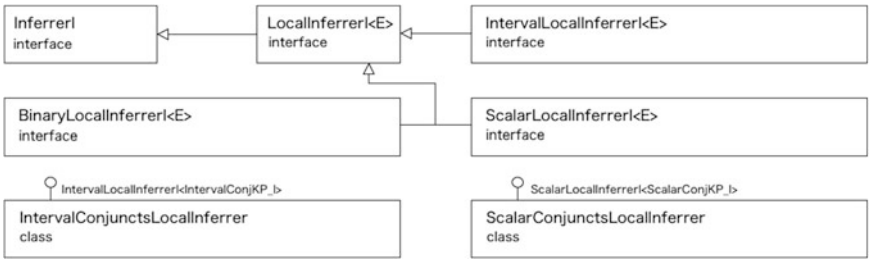


Fig. 2 Inferrers structure

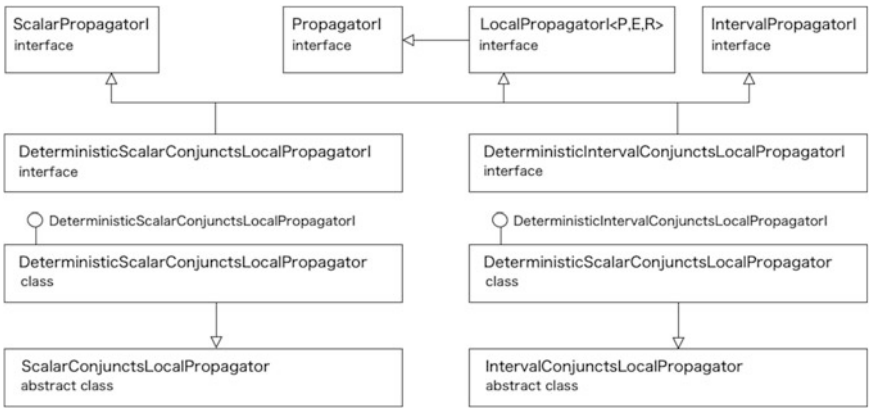


Fig. 3 Propagators structure

2.3 Probabilistic-Logic Inference Implementation

The developed software package allows to construct knowledge patterns, which can be built over the ideals of conjuncts, disjuncts or quanta propositions, with different types of probability estimates: binary, scalar or interval and conduct a P-LI over the derived KP.

As an example of the P-LI let us consider maintenance of consistency in the knowledge pattern, which is built over the ideal of conjuncts with interval probability estimates.

Let us consider the KP over the alphabet consisting of two symbols:

$A = \{x_1, x_2\}$. Let's take probability estimates of conjuncts:

$$\left(\mathbf{P}_{\mathbf{c}}^{-}\right)^{\mathrm{T}} = (0 \quad 0 \quad 0.2 \quad 0.2) \text{ and } \left(\mathbf{P}_{\mathbf{c}}^{+}\right)^{\mathrm{T}} = (1 \quad 1 \quad 0.7 \quad 0.7).$$

As a result of their adjustment with the methods of the class Interval- Conjuncts-Local- Inferrer we receive the detailed probability estimates:

$$\left(\mathbf{P}_{\mathbf{c}}^{-}\right)^{\mathrm{T}} = (0.2 \quad 0.2 \quad 0.2 \quad 0.2) \text{ and } \left(\mathbf{P}_{\mathbf{c}}^{+}\right)^{\mathrm{T}} = (1 \quad 1 \quad 0.7 \quad 0.7).$$

It is obvious that the estimates are narrowed, and verification confirms the consistency of the obtained KP.

3 SMJG

The set of minimal joint graphs (MJG) is considered in the section, assembly algorithms for set synthesis are presented, which then allow to select the optimal one for probabilistic-logic inference.

3.1 Necessity of Use Secondary Global Structure

Having considered ABN KP representations and local P-LI algorithms implementations [11], it is impossible not to mention the usage context of the upcoming C# library.

At first, note should be taken that researchers mostly cannot limit themselves with processing a single KP: the tasks they fulfil in the main include set of KP representations and processing, i.e. ABN. This requires to bring following issues to a close: adding information about set of KP and having a convenient secondary global structure visualisation, to represent links among KPs.

Besides, built library includes global P-LI algorithms in ABN. Therefore, it requires to settle a problem of ABN secondary global structure synthesis in order to provide dissemination of impact on local changes across the network, propagation of evidence, verification of a variety of ABN consistency degrees.

The current approach in the ABN theory is that the secondary global structure is MJG, though these several graphs can be corresponded to same set of KP. For this reason, we consider structural tasks in details paying more attention to MJG and assembly models [12].

3.2 Definitions

A primary structure is a set which should consist of inclusion maximal subalphabets: $\forall w_i, w_j \in$ Workloads, $i \neq j \Rightarrow w_i \subsetneq w_j$. A collection of these sets is denoted by PrimaryStructures.

Compression $\sigma_U : G \to G$ is a map $\sigma_U(G) = K_U$, where $K_U = \langle F_U, E_U, d(e)\rangle$, such that

1. $F_U = \{f_i | f_i = \sigma_U(P_U^i), P_U^i \subseteq V(G \downarrow U)\};;$
2. $E_U = \{e_i | e_i = \sigma_U(E_{i,j}), E_{i,j} \subset (G \downarrow U) \setminus (G \Downarrow U)\};;$
3. $d : E_U \to N : d(\sigma_U(E_{ij})) = |E_{ij}|$, where $|E_{ij}|$ is the multiplicity.

A feud of a weight U, or f_i, is a vertex obtained by compressing some possession. *A curia* of the weight U, or K_U, is a multigraph which came out from compressing the narrowing $G \downarrow U$. *A homage* H_U is a curia K_U being a tree in which all the edges have multiplicity 1.

A sinew $s_{i,j}^U$ is a graph, built on vertices of the restriction $M \downarrow U$ and the corresponding homage $H_i^U : \sigma_U(s_{i,j}^U) = H_i^U$. The set of sinews of the weight U is denoted by Sinews_U.

A sheaf is a graph, constructed by unification of sinews, selected by one for every separator:

$$\bigcup_{U_k \in Sep} s_{i_k,j_k}^{U_k}, \ where \ s_{i_k,j_k}^{U_k} \in Sinews_{U_k}.$$

3.3 MJG and Assembly Algorithms

Next, the matter will be subsets of MJG synthesis algorithms, where subsets can be either the whole set of MJG (SMJG) or a singleton.

Denoted ModelKit as the following quadruple ⟨Loads, StereoSeparators, StereoHoldings, NecessaryEdges⟩, where Loads is a primary structure, StereoHoldings is a set of strong narrowing components, where narrowing influences every stereoseparator:

$$\text{StereoHoldings} = \{\text{Components}(\downarrow U) | \ U \in \text{StereoSeparators}\},$$

where Components returns set of graph connected components.

NecessaryEdges is a set of necessary edges:

$$\text{NecessaryEdges} = \bigcup_{U \in \text{BSep}} E(\downarrow U),$$

where BSep is a set of biseparators.

MJG subsets synthesis algorithms given ModelKit are named *assembly algorithms*.

Next, we will see several SMJG synthesis algorithms. The efficiency of the suggested algorithms is mentioned in [3].

3.4 MK1 Algorithm Description

In listing 1 there is a pseudocode of SMJG synthesis algorithm MK1.

Algorithm 1 Simple algorithm generating SMJG MK1

input: Loads
output: ModelKit

```
 1: StereoSeparators = ∅
 2: StereoHoldings = ∅
 3: NecessaryEdges = ∅
 4: ⟨ Separators, SeparatorTable ⟩ = SeparatorSet(Loads)
 5: for each s in Separators
 6:     Vertices = NarrowedVertices(s, Loads)
 7:     if |Vertices| = 2 then
 8:         NecessaryEdges = NecessaryEdges ∪ {Vertices[0], Vertices[1]}
 9:     else
10:         Holdings = Components(StrongNarrowing(Vertices, s, SeparatorTable))
11:         if |Holdings| > 1 then
12:             StereoSeperators = StereoSeperators ∪ {s}
13:             StereoHoldings[s] = Holdings
14: return ⟨ Loads, StereoSeparators, StereoHoldings, NecessaryEdges ⟩
```

The idea of MK1 is following: first, the set of non-empty separators is constructed (line 4), second, the set of vertices which are from respective narrowing is constructed for every separator going over all loads (line 6). Third, it is checked for the constructed set of vertices if the separator is a biseparator (i.e. the number of vertices equals 2) or a stereoseparator (i.e. it is not a biseparator and the number of possessions are greater than 1). In first case (lines 7–8) the edge is added to the set of necessary edges. In second case the possessions are constructed as the connected components of strong narrowing (lines 10–13) [3] (Fig. 4).

3.5 MK4 Algorithm Description

In listing 2 there cites a pseudocode of SMJG synthesis algorithms MK4.

Algorithm 2 Advanced SMJG synthesis algorithm MK4

input: Loads
output: ModelKit

```
 1: StereoSeparators = ∅
 2: StereoHoldings = ∅
 3: NecessaryEdges = ∅
 4: ⟨ Separators, MainVertices, SeparatorTable ⟩ =
 5:       SeparatorSetAndMainVertices(Loads)
 6: Sons = Sons(Separators)
 7: SortedSeparators = SortSeparators(Separators)
 8: for each s in SortedSeparators
 9:       Vertices = MainVertices[s]
10:       if |Vertices| = 2 then
11:             if Sons[s] = ∅ then
12:                   NecessaryEdges = NecessaryEdges ∪ {Vertices[0], Vertices[1]}
13:             else
14:                   for each c in Sons[s]
15:                         Vertices = Vertices ∪ Vertices(c)
16:       else
17:             Holdings = HoldingsBy4thStructure(s,Sons,Vertices)
18:             if |Holdings| > 1 then
19:                   StereoSeperators = StereoSeperators ∪ {s}
20:                   StereoHoldings[s] = Holdings
21:             else
22:                   for each c in Sons[s]
23:                         Vertices = Vertices ∪ Vertices(c)
24: return ⟨ Loads, StereoSeparators, StereoHoldings, NecessaryEdges ⟩
```

The idea of MK4 is following: at first a set of non-empty separators is constructed simultaneously with the set of main vertices of the separators (lines 4–5), then the tertiary structure is constructed (line 6). Separators are considered ascending (line 7), if the separator *s* has two main vertices and has no children, it is called a biseparator and processed accordingly (lines 10–12). If it has children, it is called a monoseparator of type 1, and then the set of vertices is constructed joining the sets of his sons vertices with its own main vertices (lines 14–15). If the separator has more than two main vertices, the set of his possessions is synthesised constructing a half-sibling graph and finding the connected components in it. If the number of possessions is greater than 1, then the separator is processed as the stereoseparator (lines 18–20). If the number of possessions equals 1, the separator is a monoseparator but not of type 1, and a set of his vertices is constructed same way as for a monoseparator of type 1 (lines 22–23). That both monoseparators' types processing is necessary that while constructing the monoseparators' parent possessions all the monoseparators' vertices are taken into account but not only those which the main one are [4].

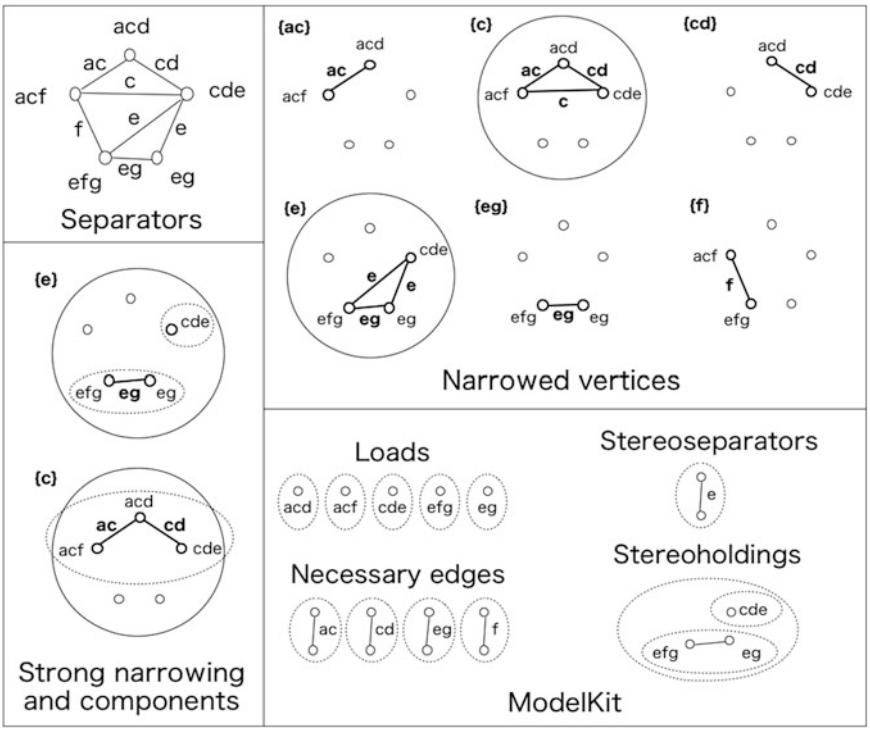


Fig. 4 The work example of MK1

3.6 The Work Instance of Assembly Algorithms

Similarly, the example of the algorithm MK4 could be referred to.

4 Conclusion

Thus, we considered the primary and the secondary structure of the ABN. For the primary structure we introduced the basic terms and concepts relating to the PLI and described the developed C# library implementing all PLI algorithms using side library lp_solve for linear programming problems solving. Designed library corresponds to the basic principles of object-oriented programming, which makes it expandable and has API, making it reusable in future global PLI algorithms implementation. Also we described 2 algorithms for minimal joint graph generation, supplemented by a program code listings and examples of their use. The obtained secondary structure gives an opportunity to extend application of local inference machine so that it can conduct a variety of global probability-logic inference types.

Acknowledgments The paper presents results of the project partially supported with RFBR grant 15-01-09001-a "Combined Probabilistic-Logic Graphical Approach to Representation and Processing of Unsertain Knowledge Systems: Algebraical Bayesian Networks and Related Models".

References

1. Butz, C.J., Oliveira, J.S., Madsen, A.L.: Bayesian network inference using marginal trees. Int. J. Approx. Reason. **68**, 127–152 (2015)
2. Farasat, A., Nikolaev, A., Srihari, S.N., Blair, R.H.: Probabilistic graphical models in modern social network analysis. Soc. Netw. Anal. Min. **5**(1) (2015)
3. Filchenkov, A.A., Frolenkov, K.V., Sirotkin, A.V., Tulupyev, A.L.: Minimal joint graph subsets synthesis system. Tr. SPIIRAN **27**, 200–244 (2013). http://mi.mathnet.ru/trspy656 (in Russian)
4. Fil'chenkov, A.A., Tulupyev, A.L.: Minimal joint graph structure synthesis. Tr. SPIIRAN **11,** 104–129 (2009). http://mi.mathnet.ru/trspy50 (in Russian)
5. Kappel, D., Habenschuss, S., Legenstein, R., Maass, W.: Network plasticity as Bayesian inference. PLoS Comput. Biol. **11**(11) (2015)
6. Nelson, A.W., Malik, A.S., Wendel, J.C., Zipf, M.E.: Probabilistic force prediction in cold sheet rolling by Bayesian inference. J. Manuf. Sci. Eng.-Trans. ASME **136** (2014)
7. Sirotkin, A.V.: Algebraic Bayesian networks: computational complexity of algorithms for probabilistic-logical inference under uncertainty (2011) (in Russian)
8. Tulupyev, A.L., Nikolenko, S.I., Sirotkin, A.V.: Bayesian networks: a probabilistic-logic approach. SPb.: Nauka (2006) (in Russian)
9. Tulupyev, A.L., Sirotkin, A.V., Nikolenko, S.I.: Bayesian belief networks. SPb.: SPbSU Press (2009) (in Russian)
10. Tulupyev, A.L., Sirotkin, A.V., Zolotin, A.A.: Matrix equations in a posteriori inference of truth estimates in algebraic Bayesian networks. Vestnik St. Petersburg University. Mathematics **48**(3), 168–174 (2015)
11. Tulupyev, A.L., Stolyarov, D.M., Mentyukov, M.V.: A representation for local and global structures of an algebraical Bayesian network in java applications. Tr. SPIIRAN **5**, 71–99 (2007). http://mi.mathnet.ru/trspy301 (in Russian)
12. Zotov, M.A., Tulupyev, A.L., Sirotkin, A.V.: Complexity statistical estimates of straightforward and greedy algorithms for algebraic Bayesian networks syntesis. Nechetkie Sistemy i Myagkie Vychisleniya [Fuzzy Systems and Soft Computing] **10**(1), 75–91 (2015) (in Russian)

Modeling of Injured Position During Transportation Based on Bayesian Belief Networks

A.I. Motienko, A.L. Ronzhin, O.O. Basov and M. Zelezny

Abstract Development of rescue robotics for injured transportation connected with problem of selection of injured position based on trauma type. The paper presents a model of a position of the injured during transportation based on Bayesian belief networks. The developed Bayesian belief network structure is represented by the signs of trauma, trauma itself, positions for transportation of the sufferer corresponding to trauma, and the relationships between them. Conditional probabilities tables is determined based on the expert information; available medical research focused on the identification of the similar relationships between the elements of the diagnostic process; historical statistics. The simulation results show that the developed Bayesian belief network enables one to solve probabilistic forecasting tasks based on subjective and incomplete data. The former are obtained during questioning the sufferer; the latter are based on the computer vision systems (examination) and sensors for various purposes (manipulation) that are installed on specialized robots. The developed tools are focused on the rescue robotics based on intelligent hardware and software for human-robot interaction.

Keywords Robotics · Search-and-rescue · Human-machine interaction · Transportation of the injured · First aid · Rescue work · Emergency · Bayesian belief network

1 Introduction

According to world statistics (starting from the second half of the 20th century), injuries come first in the list of causes of death of able-bodied population under the age of 45 years. In their turn, natural disasters and fires are in second place

A.I. Motienko · A.L. Ronzhin (✉) · O.O. Basov
SPIIRAS, 39, 14th Line, Saint Petersburg 199178, Russia
e-mail: ronzhin@iias.spb.su

M. Zelezny
University of West Bohemia, Pilsen, Czech Republic

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_8

(after road accidents) among the causes of such deaths. In Russia, the largest number of human losses is due to natural disasters such as floods and earthquakes. Around 13 to 15 thousand people die in fires each year. Conducting rescue and other urgent work aimed at eliminating the effects of emergency situations is one of the main tasks of the Russian Unified System for Emergency Situations and Civil Defense [1].

The purpose of any rescue and other emergency operations is to save people and assist to sufferers, localize accidents and eliminate the damage hindering rescue operations, and create conditions for the subsequent recovery works. According to the World Health Organization, about 85 % of people die from injuries simply because they were not provided with first aid, or first aid was rendered untimely or improperly during the emergency situation (and only 15 % die from fatal injury) [2].

Rescue works are characterized by the presence of factors that threaten life and health of those who conduct them (rescue workers, firefighters, etc.), so special training, clothing and equipment are required. It is possible to minimize the risk to the rescuers by applying the so-called unmanned technologies—automated robotic systems and tools [3], intended for ensuring the implementation of first aid measures [4] and for transporting the sufferer. However, to determine signs of life (consciousness) and to examine the sufferer, robotic tools need to possess relevant sensors (temperature, pressure, humidity ones) and systems (for computer vision, analysis and synthesis of speech for questioning the victim, etc.), nomenclature and configuration of which should be identified based on weight, size and power characteristics of the robot. It is necessary to create an appropriate scientific and methodological apparatus that will allow one to determine the optimal position of the body of the sufferer, even in terms of lacking data about the signs of injury, sufferer's condition and symptoms of illness.

2 Bayesian Belief Networks Structure Formation and Training for Modeling Casualty Movement Process

The developed Bayesian belief network (BBN) structure is represented by the signs of trauma, trauma itself, positions for transportation of the sufferer corresponding to trauma, and the relationships between them (Fig. 1) [4]. The nodes of the BBN graph are introduced based on injury patterns that can be detected visually, by questioning, or by means of simple manipulations. The proposed model consists of the following subgraphs (for clarity, duplicate graph nodes are indicated by the dotted line):

- subgraph G_1 "Position for transportation" that defines the relationship between positions for transportation and injuries corresponding to them;

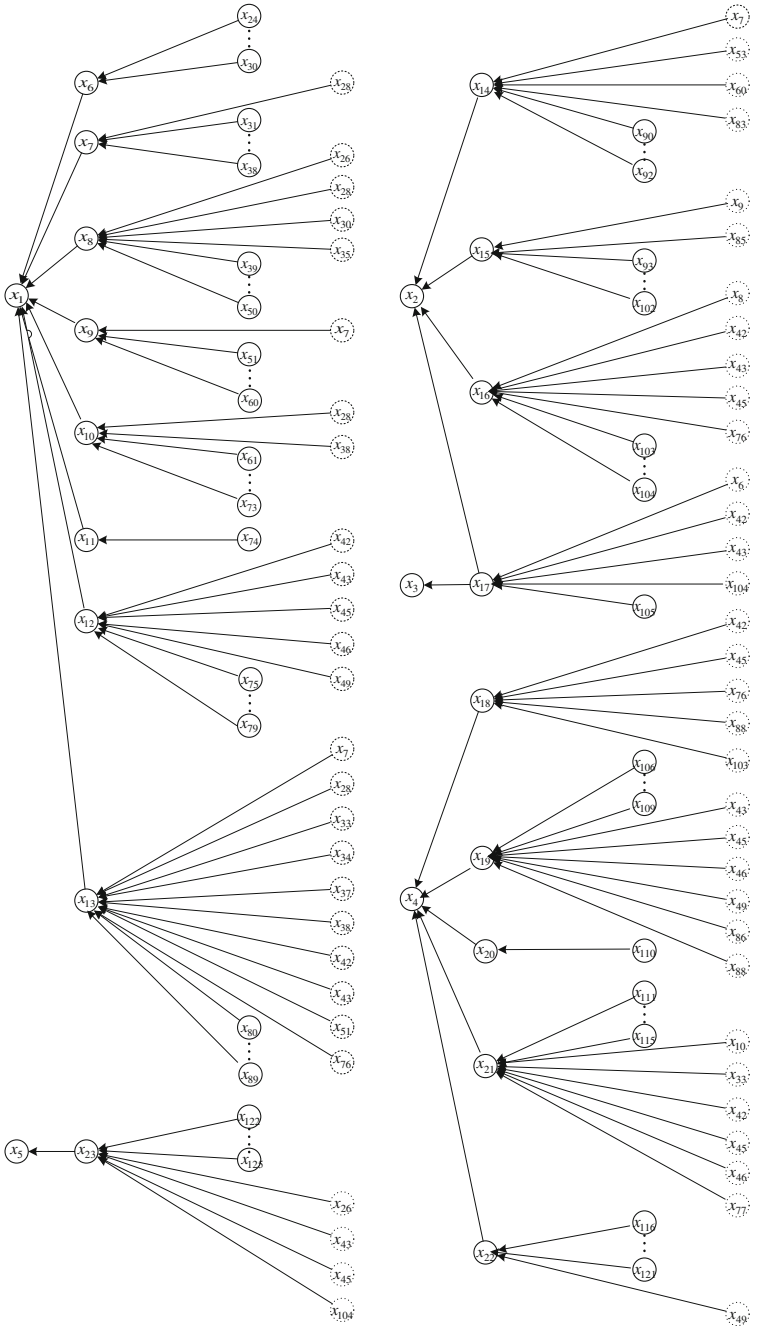


Fig. 1 A model of a position for transportation of the sufferer based on BBN

- subgraph G_2 "Spinal injury" with a table of conditional probabilities (CP) $p(x_6|\tilde{x}_{24}\tilde{x}_{25}\tilde{x}_{26}\tilde{x}_{27}\tilde{x}_{28}\tilde{x}_{29}\tilde{x}_{30})$;
- subgraph G_3 "Depressed case" with a table of CP $p(x_7|\tilde{x}_{28}\tilde{x}_{31}\tilde{x}_{32}\tilde{x}_{33}\tilde{x}_{34}\tilde{x}_{35}\tilde{x}_{36}\tilde{x}_{37}\tilde{x}_{38})$;
- subgraph G_4 "Fractures of the pelvis and lower limb" with a table of CP $p(x_8|\tilde{x}_{26}\tilde{x}_{28}\tilde{x}_{35}\tilde{x}_{39}\tilde{x}_{40}\tilde{x}_{41}\tilde{x}_{42}\tilde{x}_{43}\tilde{x}_{44}\tilde{x}_{45}\tilde{x}_{46}\tilde{x}_{47}\tilde{x}_{48}\tilde{x}_{49}\tilde{x}_{50})$;
- subgraph G_5 "Brain concussion" with a table of CP $p(x_9|\tilde{x}_7\tilde{x}_{51}\tilde{x}_{52}\tilde{x}_{53}\tilde{x}_{54}\tilde{x}_{55}\tilde{x}_{56}\tilde{x}_{57}\tilde{x}_{58}\tilde{x}_{59}\tilde{x}_{60})$;
- subgraph G_6 "Chest injury" with a table of CP $p(x_{10}|\tilde{x}_{28}\tilde{x}_{61}\tilde{x}_{62}\tilde{x}_{63}\tilde{x}_{64}\tilde{x}_{65}\tilde{x}_{66}\tilde{x}_{67}\tilde{x}_{68}\tilde{x}_{69}\tilde{x}_{70}\tilde{x}_{71}\tilde{x}_{72}\tilde{x}_{73}\tilde{x}_{74})$;
- subgraph G_7 "Amputated lower limb" with a table of CP $p(x_{11}|\tilde{x}_{75})$;
- subgraph G_8 "Facial trauma" with a table of CP $p(x_{12}|\tilde{x}_{42}\tilde{x}_{43}\tilde{x}_{45}\tilde{x}_{46}\tilde{x}_{76}\tilde{x}_{77}\tilde{x}_{78}\tilde{x}_{79}\tilde{x}_{80}\tilde{x}_{81})$;
- subgraph G_9 "Abdominal trauma" with a table of CP $p(x_{13}|\tilde{x}_7\tilde{x}_{28}\tilde{x}_{33}\tilde{x}_{34}\tilde{x}_{37}\tilde{x}_{38}\tilde{x}_{42}\tilde{x}_{43}\tilde{x}_{78}\tilde{x}_{83}\tilde{x}_{84}\tilde{x}_{85}\tilde{x}_{86}\tilde{x}_{87}\tilde{x}_{89}\tilde{x}_{90}\tilde{x}_{91}\tilde{x}_{92})$;
- subgraph G_{10} "Blood loss" with a table of CP $p(x_{14}|\tilde{x}_7\tilde{x}_{53}\tilde{x}_{60}\tilde{x}_{85}\tilde{x}_{93}\tilde{x}_{94}\tilde{x}_{95})$;
- subgraph G_{11} "Injury to the occipital lobe" with a table of CP $p(x_{15}|\tilde{x}_9\tilde{x}_{87}\tilde{x}_{96}\tilde{x}_{97}\tilde{x}_{98}\tilde{x}_{99}\tilde{x}_{100}\tilde{x}_{101}\tilde{x}_{102}\tilde{x}_{103}\tilde{x}_{104}\tilde{x}_{105})$;
- subgraph G_{12} "Injury to buttocks and the dorsum of the legs" with a table of CP $p(x_{16}|\tilde{x}_8\tilde{x}_{42}\tilde{x}_{43}\tilde{x}_{45}\tilde{x}_{78}\tilde{x}_{106}\tilde{x}_{107})$;
- subgraph G_{13} "Back injuries" with a table of CP $p(x_{17}|\tilde{x}_6\tilde{x}_{42}\tilde{x}_{43}\tilde{x}_{107})$;
- subgraph G_{14} "Bruise, cuts, abrasions" with a table of CP $p(x_{18}|\tilde{x}_{42}\tilde{x}_{45}\tilde{x}_{78}\tilde{x}_{106}\tilde{x}_{109})$;
- subgraph G_{15} "Shoulder girdle injuries" with a table of CP $p(x_{19}|\tilde{x}_{43}\tilde{x}_{45}\tilde{x}_{46}\tilde{x}_{76}\tilde{x}_{89}\tilde{x}_{110}\tilde{x}_{111}\tilde{x}_{112}\tilde{x}_{113})$;
- subgraph G_{16} "Amputated upper limb" with a table of CP $p(x_{20}|\tilde{x}_{114})$;
- subgraph G_{17} "Trauma to eyes, chest and respiratory tract" with a table of CP $p(x_{21}|\tilde{x}_{10}\tilde{x}_{33}\tilde{x}_{42}\tilde{x}_{45}\tilde{x}_{46}\tilde{x}_{115}\tilde{x}_{116}\tilde{x}_{117}\tilde{x}_{118}\tilde{x}_{119}\tilde{x}_{120})$;
- subgraph G_{18} "Upper limbs trauma" with a table of CP $p(x_{22}|\tilde{x}_{76}\tilde{x}_{121}\tilde{x}_{122}\tilde{x}_{123}\tilde{x}_{124}\tilde{x}_{125}\tilde{x}_{126})$;
- subgraph G_{19} "Neck injury" with a table of CP $p(x_{23}|\tilde{x}_{43}\tilde{x}_{45}\tilde{x}_{54}\tilde{x}_{107}\tilde{x}_{127}\tilde{x}_{128}\tilde{x}_{129}\tilde{x}_{130})$.

The optimal positions for transportation of the sufferers $x1 \ldots x5$, list of the most common injuries $x6 \ldots x23$ and their signs that are used to preset the nodes of the developed BBN, as well as methods to determine these injures (by examination, questioning and manipulation) are presented in [4].

The notation $\tilde{x}$ indicates that here in the formula both a propositional formula x and its negation $\bar{x}$ can be used.

For the subgraph G_1, a table of CP will be

$$p(x_1|\tilde{x}_6\tilde{x}_7\tilde{x}_8\tilde{x}_9\tilde{x}_{10}\tilde{x}_{11}\tilde{x}_{12}\tilde{x}_{13}) = 1,$$

if there is at least one propositional formula $x_i, i = 6 \ldots 13$;

$$p(x_2|\tilde{x}_{14}\tilde{x}_{15}\tilde{x}_{16}\bar{x}_{17}) = 1,$$

if there is at least one propositional formula $x_i, i = 14 \ldots 16$;

$$p(x_2|x_{14}x_{15}x_{16}x_{17}) = 1; \quad p(x_2|\bar{x}_{14}\bar{x}_{15}\bar{x}_{16}x_{17}) = 0.5; \quad p(x_2|\tilde{x}_{14}\tilde{x}_{15}\tilde{x}_{16}x_{17}) = 0.75,$$

with any other combination of propositional formulas x_i and their negations $\bar{x}_i$ $(i = 14 \ldots 16)$;

$$p(x_3|x_{17}) = 1; p(x_3|\bar{x}_{17}) = 0; p(x_4|\tilde{x}_{18}\tilde{x}_{19}\tilde{x}_{20}\tilde{x}_{21}\tilde{x}_{22}) = 1,$$

if there is at least one propositional formula $x_i, i = 18 \ldots 22$;

$$p(x_5|x_{23}) = 1; p(x_5|\bar{x}_{23}) = 0$$

For the subgraph G_5, the independence of the events $(x_7, x_{51} \ldots x_{60})$ and their equal significance for determining the posterior probability of an event x_9 "Brain concussion" is considered acceptable at the initial stage of creating the table of CP. Then:

$$p(x_9|\tilde{x}_7\tilde{x}_{51}\tilde{x}_{53}\tilde{x}_{54}\tilde{x}_{55}\tilde{x}_{56}\tilde{x}_{57}\tilde{x}_{58}\tilde{x}_{59}\tilde{x}_{60}) = 0.1, \tag{1}$$

if in (1) there is one propositional formula $x_i, i = 7, 51, 53, 54 \ldots 60$;

$$p(x_9|\tilde{x}_7\tilde{x}_{51}\tilde{x}_{53}\tilde{x}_{54}\tilde{x}_{55}\tilde{x}_{56}\tilde{x}_{57}\tilde{x}_{58}\tilde{x}_{59}\tilde{x}_{60}) = 0.2, \tag{2}$$

if in (2) there are two propositional formulas $x_i, i = 7, 51, 53, 54 \ldots 60$; and so on, up to the event:

$$p(x_9|x_7x_{51}x_{53}x_{54}x_{55}x_{56}x_{57}x_{58}x_{59}x_{60}) = 1.$$

A similar approach is true for all the subgraphs $G_2, \ldots, G_{19}$ of the injuries $(x_6 \ldots x_{23})$. Specification of CP tables is carried out based on the expert information; available medical research focused on the identification of the similar relationships between the elements of the diagnostic process; historical statistics.

3 Analysis of Casualty Movement Process

The procedure for calculating the probabilities of the outcomes of one or more BBN variables based on the CP tables and known data (evidence) about the meaning of signs of injuries is called sampling. This procedure is performed during both training and using the network, so the effectiveness of the BBN largely depends on

the sampling algorithm and its implementation. All the known methods of generating a sample can be divided into two categories: algorithms representing BBN in the most convenient form for sampling (also called clustering algorithms) and approximation algorithms operating on stochastic computing techniques [5–7].

In the study, we applied a sampling algorithm for BBN that uses representation (clustering) of the original network as the so-called junction tree. This approach allows us to turn from sampling a network of a general form to a tree graph, which significantly reduces the computation time, eliminating the need for many intermediate calculations. This algorithm has several advantages:

1. It is possible to use a junction tree for networks of any topological complexity, which makes the algorithm generic and applicable to a very wide range of tasks;
2. Unlike stochastic algorithms, a junction tree algorithm allows obtaining the exact (not approximate) values of the required probabilities; at the same time the running speed of the algorithm is high enough.

In all popular programs for working with BBN, the junction tree algorithm (in a particular implementation) is the main algorithm for sampling. In addition, for a number of tasks involving working with BBN and requiring accurate results with complex network topology, network representation in the form of a junction tree is the only acceptable method of sampling [8]. As this algorithm is widely spread and holds much promise, in this study we have modelled a process of choosing a position for transportation of the injured using Netica [9].

Figure 2 shows an example of using Netica to calculate the probability of x_1 (supine position for transportation). The most likely position for transportation (Fig. 3) was determined as the value of a multitude of permissible positions that shows the maximum probability of the presence of injury (injuries), provided a specific amount of evidence $(x_{24} \ldots x_{130})$, including signs of injuries, symptoms and other parameters.

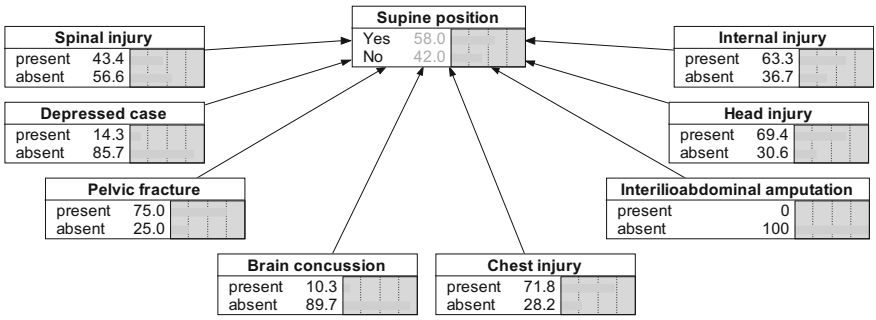


Fig. 2 Posteriori inference on the subgraph G_1 "Position for transportation" in Netica

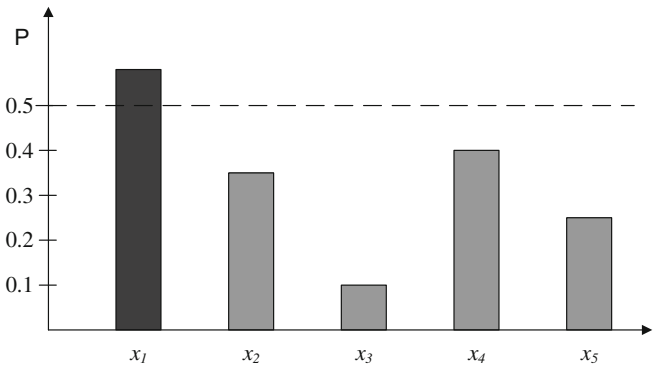


Fig. 3 Determining the most probable position for transportation of the injured

4 Conclusion

The simulation results show that the developed Bayesian belief network enables one to solve probabilistic forecasting tasks based on subjective and incomplete data. The former are obtained during questioning the sufferer; the latter are based on the computer vision systems (examination) and sensors for various purposes (manipulation) that are installed on specialized robots.

Probabilistic estimates of the observed properties (evidence) provide actual values that make it possible to consider credibility of the evaluations of unobserved properties (injuries and positions for transportation). A model of choosing a position for transportation of the injured offers the possibility to:

- reveal evidence that is the most important in terms of its impact on the decision, to list it and rank in order of importance;
- identify variables x_i $(i = 24 \ldots 125)$ that do not afford basis for conclusions;
- create an ordered list of steps that most effectively lead to a clear decision on the position for transportation (for example, a list of questions to be asked to the sufferer during situation analysis; the list and sequence of sensors application, etc.).

A study of the stated possibilities will allow one to modify the developed model of choosing a position for transportation of the sufferer; to synthesize algorithms for the estimation of the signs of trauma, physical and emotional state of a person; and to optimize the algorithm for probabilistic inference. Tools under consideration are focused on the development of rescue robotics based on intelligent hardware and software for human-machine interaction [10–15].

Acknowledgment This work is partially supported by the Russian Foundation for Basic Research (grant № 16-08-00696-a).

References

1. Motienko, A.I., Basov, O.O.: Application of automated robotic means of transportation for first aid to sufferer. In: International Youth Scientific and Technical Conference on Progressive Technologies and Processes, vol. 2, pp. 216–220. Kursk. (2015) (in Russ.)
2. Konnova, L.A., Balabanov, V.A., Artamonova, G.K.: Basics of first aid: a textbook for students and students of higher educational institutions enrolled in the bachelor degree program. In: Latyshev, O.M. (ed.) Technosphere Safety and the Specialty Fire Safety. Saint-Petersburg University of State Fire Service of Emercom of Russia, Saint-Petersburg (2015) (in Russ.)
3. Motienko, A.I., Ronzhin, A.L., Pavljuk, N.A.: The modern development of rescue robots, opportunities and principles of their application. Sci. Bull. NSTU, **3**(60), 147–165 (2015) (in Russ.)
4. Motienko, A.I., Makeev, S.M., Basov, O.O.: Analysis and modeling of position choice process for transportation of the sufferer on the basis of bayesian belief networks. SPIIRAS Proc. **43**, 135–155 (2015) (in Russ.)
5. Henrion, M.: Propagation of uncertainty by probabilistic logic sampling in bayes networks. In: Lemmer, J.F., Kanal, L.N. (eds.) Uncertainty in Artificial Intelligence, vol. 2, pp. 149–163. Elsevier Science Publishing Company, Inc., New York, NY, USA (1988)
6. Fung, R., Chang, K.-C.: Weighting and integrating evidence for stochastic simulation in bayesian networks. In: Fifth Conference on Uncertainty in Artificial Intelligence (UAI-89), pp. 475–482. Amsterdam (1989)
7. Shachter, R., Peot, M.: Simulation approaches to general probabilistic inference on belief networks. In: Fifth Conference on Uncertainty in Artificial Intelligence (UAI-89), pp. 311–318. Amsterdam (1989)
8. Maslennikov, E.D., Sulimov, V.B.: Predictions based on bayesian belief networks: algorithm and software implementation. In: Numerica methods and programming, vol. 11, no. 2, pp. 222–235 (2010) (in Russ.)
9. Netica. http://www.norsys.com/
10. Ronzhin, A.L., Yusupov, R.M.: Multimodal interfaces for autonomous robotic systems. Izvestiya SFedU. Eng. Sci. **162**(1), 195–206 (2015) (in Russ.)
11. Basov, O.O.: Principles of construction of polymodal info-communication systems based on multimodal architectures of subscriber's terminals. SPIIRAS Proc. **39**, 109–122 (2015). (in Russ.)
12. Kozyrenko, N.K., Meshcheryakov, R.V., Hodashinsky, I.H., Anufrieva, N.A.: Mathematical model and algorithms of people health evaluation. SPIIRAS Proc. **33**, 117–146 (2015). (in Russ.)
13. Yusupov, R.M., Ronzhin, A.L.: From smart devices to smart space. Her. Russ. Acad. Sci. **80** (1), 45–51 (2010)
14. Karpov, A.A., Ronzhin, A.L.: Information enquiry kiosk with multimodal user interface. Patt. Recogn. Image Anal. **19**(3), 546–558 (2009)
15. Ronzhin, A.L., Budkov, V.Y.: Multimodal interaction with intelligent meeting room facilities from inside and outside. In: Balandin, S., et al. (eds.) NEW2AN/ruSMART 2009. LNCS vol. 5764, pp. 77–88. Springer, Heidelberg (2009)

Passive Steganalysis Evaluation: Reliabilities of Modern Quantitative Steganalysis Algorithms

N. Prokhozhev, O. Mikhailichenko, A. Sivachev, D. Bashmakov and A. Korobeynikov

Abstract This paper presents initial results from experiments which perform statistically accurate evaluation of the reliability of modern quantitative statistical steganalysis algorithms. The focus here is on the algorithms for detection of simple LSB steganography in grayscale images. The following algorithms were evaluated: RS- analysis, Sample pair analysis, Difference image histogram, Triples analysis, Weighted stego-image. ROC curves were obtained as a result the reliability of image steganalysis algorithms. We also describe some of the questions for a general methodology of evaluation of steganalysis algorithms and potential pitfalls caused by the differences of image resolution.

Keywords Statistical steganalysis · LSB steganography · Regular-singular steganalysis · Difference histogram steganalysis · Triplet steganalysis

N. Prokhozhev · O. Mikhailichenko · A. Sivachev (✉) · D. Bashmakov · A. Korobeynikov
ITMO Univercity, Kronverksky Pr. 49, 197101 Saint Petersburg, Russia
e-mail: avsivachev@corp.ifmo.ru; sivachev239@mail.ru
URL: http://en.ifmo.ru

N. Prokhozhev
e-mail: 144339@corp.ifmo.ru

O. Mikhailichenko
e-mail: 116198@corp.ifmo.ru

D. Bashmakov
e-mail: dabashmakov@corp.ifmo.ru

A. Korobeynikov
e-mail: 105929@niuitmo.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_9

1 Introduction

Nowadays steganography is widely used for hide information transferring and hidden messaging. Therefore, while solving problems of the informational security, steganography is being increasingly used by secret agencies, criminal groups and terroristic organizations.

One of the most widespread types of steganography containers is a digital image. Digital images make up a large percentage of the Internet data traffic allows using them as a channel for hidden information transfer with significant throughput and great secrecy.

There is a large number of free software for LSB image steganography. There are two relevant steganography resistance approaches: passive and active.

In the simplest case, the procedure of passive steganography resistance reduces to the problem of binary classification. A tested container should be classified as either original (clear) image or steganogram. It's a main goal quantitative steganalytic algorithms are widely used for. For digital images and LSB steganography, the result of algorithm execution is the estimate of the number of changed pixels in the tested image.

2 Purpose

The main purpose of this paper is the estimation of efficiency of modern steganographic algorithms when resisting LSB steganography with small payload rates. Results illustrate the lower bounds of payload, when the classification accuracy falls dramatically. In the case of such payloads the effective resistance is practically impossible. This lower bound can be used to determine the value of maximal throughput of steganography channel, considering existing abilities of passive resistance.

3 Method of Experiment

The tested image is being selected from a test set. Then LSB bit is being inverted for a fixed number of pixels. In this way, the apply steganography effect is simulated. Payload is measured as a percentage of maximum value for the tested image. Maximum payload of the image is the total amount of pixels in the image. Payload varies in the range of 1−6 %, basing on the stated error of evaluated steganalytic algorithms, roughly estimated at 2 %. Modified image is being tested with steganalytic algorithm. The result of the estimation is a number of pixels affected with steganographic embedding.

The following steganalytic algorithms have been evaluated:

- RS-analysis (RS) [1];
- Sample pair analysis (SP) [2];

- Difference image histogram (DIH) [3];
- Triples analysis (TR) [4];
- Weighted stego-image (WSI) [5].

Test set consists of three grayscale image collections:

- Collection 1–3126 images with resolution from 392 × 550 to 5184 × 3456 [9];
- Collection 2–5214 images with resolution from 1339 × 1357 to 5100 × 4026 [10];
- Collection 3–30682 images with resolution from 700 × 500 to 1300 × 734 [11].

Collections are public and include a great number of images with wide range of characteristics and features.

The pixels' coordinates are pseudo-randomly selected, with uniform distribution.

4 Evaluation Procedure

As noted above, the procedure of passive steganography resistance reduces to the problem of binary classification. Then the classification sets are represented as

$$Y = \{-1, +1\}, \tag{1}$$

and classifier is represented as follows:

$$a(x) = sign[f(x, w) - w_0], \tag{2}$$

where x is a random object, $f(x, w)$ is a classifying function (steganalytic algorithm), w is a parameters vector (the amount of flipped pixels), w_0 is a threshold.

In this case, the steganalytic algorithm acts as a classifier, and the classification procedure is done by comparing result of algorithm execution and some threshold.

The distinction surface is defined as follow:

$$f(x, w) = w_0. \tag{3}$$

The choice of a threshold value w_0 is a difficult task, since it directly affects the probabilities of both true positive and false positive detection. Too small value may result in high probability of both false classifications. In practice, it will result in clear images classified as stego images. Too great value will enlarge the probability of false negative classification.

True positive classification rate (TPR) is defined as follows:

$$TPR(a, X^m) = \frac{\sum_{i=1}^{m} [a(x_i) = +1][y_i = +1]}{\sum_{i=1}^{m} [y_i = +1]}, \tag{4}$$

where

$$X^m = (x_1, \ldots x_m) \tag{5}$$

is a sample, $y_1, \ldots, y_m$ are classification sets of the sample.

False positive classification rate (FPR) is defined as follows:

$$FPR(a, X^m) = \frac{\sum_{i=1}^{m} [a(x_i) = +1][y_i = -1]}{\sum_{i=1}^{m} [y_i = -1]}. \tag{6}$$

The ratio of these two values affects classification quality, and, consequently, the practical efficiency of passive resistance. The ideal classifier has $TPR = 1$ and $FPR = 0$.

ROC curves illustrate the ratio, for greater clarity [6].

The set of ROC curves for different evaluated algorithms for the same experiment conditions sets makes possible the comparative evaluation of their efficiency.

5 Results

The graph shows ROC curves (Fig. 1) for all evaluated algorithms and the same ROC curves for different payload rates. The same experimental results for distinct image collections are stated in Table 1.

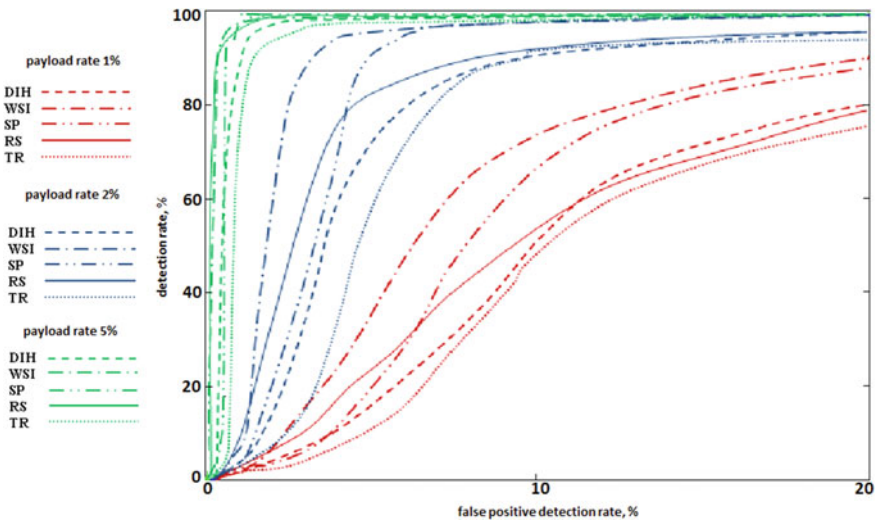


Fig. 1 ROC curves for the evaluated algorithms

Table 1 False positive rate for TPR at 97.5 %

Evaluated algorithm	Collection#	Payload, %				
		1	2	3	4	5
RS-analysis	1	48.1	30.0	7.1	2.1	0.7
	2	3.6	1.5	1.0	0.6	0.3
	3	49.7	32.7	8.3	3.6	1.9
Difference image histogram	1	42.4	33.4	8.0	2.9	1.5
	2	5.8	2.2	1.2	0.8	0.5
	3	44.9	23.5	18.2	7.8	4.6
Sample pair analysis	1	43.9	9.6	2.1	1.1	0.5
	2	5.2	1.8	1.1	0.7	0.3
	3	44.6	12.8	4.8	2.2	1.3
Triples analysis	1	62.7	37.5	8.3	2.7	1.7
	2	4.4	1.9	1.4	1.0	0.7
	3	64.4	36.7	9.4	3.9	1.9
Weighted stego-image	1	36.7	7.1	1.8	0.7	0.5
	2	2.6	0.9	0.5	0.4	0.1
	3	38.3	8.8	2.4	1.1	0.7

6 Findings

Modern quantitative statistical steganalytic algorithms show close accuracy of estimation. Passive steganorgaphy resistance based of them provides comparable classification quality and makes their practical efficiency very similar.

The classification quality falls significantly for steganogram with payload rates, greater than 5 %.

The tested image resolution has significant impact on the classification quality. So, for the images of Collection 2, with resolution as minimum as 1339 × 1357, the false classification rate is almost 10 times less than for Collections 1 and 3, with minimal image resolutions of 392 × 550 and 700 × 500, respectively (see Table 1).

7 Conclusion

Steganalysis based on modern quantitative statistical steganalytic algorithms have an ability to provide effective passive LSB steganography resistance against steganograms with payload value greater than 5 %. Using payload value less than that threshold reduces passive resistance efficiency dramatically. The passive resistance is ineffective against LSB steganography with payload rate less than 1 % for the image with resolutions not exceeding the resolution of modern displays.

The small image with resolution 600 × 400 pixels, used as a stego image with payload rate of 1−2 % is practically undetectable by modern quantitative steganalytic algorithms. In the same time, such payload rate makes possible to embed 5−10 KB of data. Taking into account the possibility of preliminary compression of embedded data and the usage of matrix embedding [7], the evaluated algorithms need further improvements.

References

1. Fridrich, J., Goljan, M., Du, R.: Reliable detection of LSB Steganography in color and grayscale images, State University of New York, Binghamton, NY, USA
2. Lu, P., Luo, X., et. al.: An improved sample pairs method for detection of LSB embedding. In: Proceedings of the 6th Information Hiding Workshop. Springer LNCS, vol. 3200, pp. 116–128 (2004)
3. Zhang, T., Ping, X.: Reliable detection of LSB steganography based on the difference image histogram. In: Proceedings of the IEEE ICSAAP 2003, Part III, pp. 545–548 (2003)
4. Ker, A.D.: A general framework for structural steganalysis of LSB Replacement. In: Proceedings of the Information Hiding, pp. 296–311 (2005)
5. Ye, M., Liu, F., Yang, C., He, X.: Steganalysis based on weighted stego-image for lsb replacement steganography. In: Intelligent Information Hiding and Multimedia Signal Processing IIH-MSP '09, pp. 945–948 (2009)
6. Schaathun, H.G.: Machine learning in image steganalysis. Alesund University College, pp. 164–167 (2012)
7. Kim, Y., Duric, Z., Richards, D.: Modified matrix encoding technique for minimal distortion steganography. In: Proceedings of 8th International Workshop Information Hiding, vol. 4437, pp. 314−327, 10–12 Jul 2006

Bayesian Belief Networks in Risky Behavior Modelling

Alena Suvorova and Tatiana Tulupyeva

Abstract The area of risky behavior modelling has unsolved issues in current practice: there is a need for numerical estimates of risky behavior rate. We propose the approach for risky behavior modelling in terms of Bayesian Belief Networks on the base of the data about behavior episodes. The paper includes the description of the model, results of model testing on automatically generated dataset and discussion of possible further development.

Keywords Bayesian belief network · Risky behavior · Rate estimates · Poisson random process

1 Introduction

Bayesian Belief Network (BBN) is a type of probabilistic graphical models that represents a set of random variables and their conditional dependencies [1]. Formally, BBN is a directed acyclic graph that should satisfy several requirements [1]. In general BBNs represent complex influencing factor relationships, they are applied in wide range of areas: finance, medicine, information technology [2–6]. BBNs are known as good representation of knowledge and decision support under

A. Suvorova (✉) · T. Tulupyeva
SPIIRAS, Saint Petersburg, Russia
e-mail: suvalv@gmail.com

T. Tulupyeva
e-mail: tvt100a@mail.ru

A. Suvorova
National Research University Higher School of Economics, Saint Petersburg, Russia

T. Tulupyeva
Saint Petersburg State University, Saint Petersburg, Russia

T. Tulupyeva
North-West Institute of Management of the Russian Presidential Academy of National Economy and Public Administration, Saint Petersburg, Russia

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_10

uncertainty [7] and one of their important advantage is the ability to combine different source of information [8].

These features of BBN play an important role in research that involves both empirical data and expert knowledge [9]. One of such research areas is individual's risky behavior modelling. On the one hand, researchers collect data about respondents' risky behavior in different field studies; on the other hand, there are associations, causalities and assumptions based on results of previous studies, facts and theories from other research areas. To construct better models of respondents' behavior and, furthermore, to compute more accurate characteristics of that behavior, it is important to combine both data-based and expert-based information.

Knowledge of risky behavior characteristics supports decision making in many practical issues. Unusual behavior of the user of information system can be a marker that defines insecure events. Risky sexual behavior (e.g. having multiple sexual partners, unprotected sex) is widely known to be associated with high risk of sexually transmitted infections, including HIV-infection [10]. The "gold standard" for behavior rate measuring is a diary method, in other words, simple recording of episodes [11]. However this method is extremely time-consuming, resource-consuming and even hardly possible for many kinds of behavior [12], for example, due to private nature of the behavior [13] or social desirability bias. In [14] authors provided the method based on data about several behavior episodes.

Risky behavior studies were mostly focused on exploring factors associated with risky behavior or factors influenced by that behavior [15–17]. The results were mostly data-based with expert knowledge included in the form of variable selection. For those studies regression models were primary and dominating method especially in medicine and public health [16, 17]. Also the results did not provide any numerical characteristic of behavior, e.g. rate, frequency or risk that can be further implemented into automated system for decision making.

Human reliability analysis that studies causes, consequences and contributions of human failures in socio-technical systems [9] is close to risky behavior modelling in many parts: it deals with both expert information and empirical data, it faces with uncertainty of initial data, and it is related to human behavior. Human reliability analysis is based on many methods and now BBNs received an increasing attention [9].

However, this practice can not be applied directly to risky behavior modelling. The closest outcome in human reliability analysis that can be considered as behavior characteristic is probability of human error [9] but we need more detailed outcome for risky behavior rate estimate, not just "it happens with probability p". The other issue is an availability of data sources: there is no rather simply collected data from technical systems, the information about behavior comes from self-reports about risky behavior episodes.

The purpose of the paper is to describe the approach for risky behavior modelling in terms of Bayesian Belief Networks.

2 Risky Behavior Modelling

2.1 Initial Data

The model [18] is based on data about the length of intervals between three last episodes of risky behavior and the length of minimum and maximum intervals between episodes during period of interest T. The data about episodes in most applications is obtained from respondents' self-reports [14]. We assume that for each respondent occurrence of episodes follows Poisson random process: the occurrence of the next episode is independent from the previous ones, length of interval between concurrent episodes follows exponential distribution. This assumption corresponds to the features of risky behavior and, at the same time, allows less complicated calculations.

Adding data about minimum and maximum intervals decreases the influence of recent behavior represented by the last episodes. However, combining all the data about episodes leads to very complicated joint distribution [19] even in case of Poisson random process and requires much more calculation for behavior rate estimate. Any change or revision of the model, again, will require re-calculation of joint distribution (if it will be possible in elementary functions).

On the contrary, as it was mentioned earlier, BBN allows determining complex relationships in terms of simpler dependencies between small parts. Modelling risky behavior as BBN gives a way to add all available data into the model as well as include expert assumptions about relationships between them and their distributions. Revising the assumptions or adding new variables to the model requires re-arranging small part of the BBN only. Moreover, the existence of software tools (including freeware) for dealing with BBNs, for example [20] or [21], allows researchers focus on description of the model while calculations are performed automatically. Considering these advantages of BBNs we proposed a BBN for risky behavior modelling.

2.2 Model Description

The structure of BBN model is a graph $G(V, L)$ with vertices $V = \{t_{01}, t_{12}, t_{23}, t_{\min}, t_{\max}, \lambda, n\}$ and edges (or links) $L = \{(u, v) : u, v \in V\}$ (Fig. 1), where λ is random variable for behavior rate; t_{ij} is random variable for the length of the interval between ith and jth episodes from the end (0 corresponds to interview moment); $t_{\min}$ and $t_{\max}$ are random variables for the length of minimum and maximum intervals; n is random variable for the number of episodes during period of interest.

All variables were discretized. Conditional probabilities for variables were as follows [18] (where $l_s = 1, \ldots, k_s$, k_s was the number of disjunctive intervals for

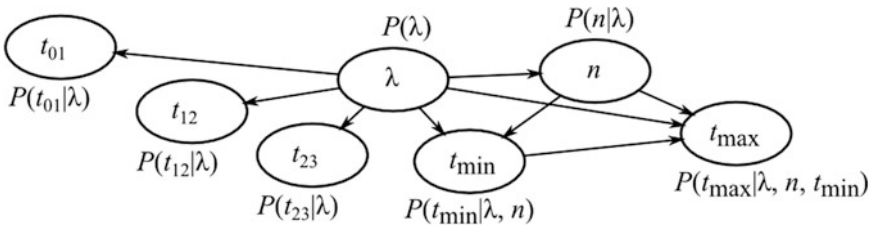


Fig. 1 Bayesian Belief network for risky behavior modelling

discretized $t_{j,j+1}$ variable; $s = 0, \ldots, 4$; $j = 0, 1, 2$; $i = 1, \ldots, m$, m was the number of disjunctive intervals for discretized λ variable):

$$p\left(t_{j,j+1}^{(l_j)} \middle| \lambda^{(i)}\right) = e^{-a\lambda^{(i)}} - e^{-b\lambda^{(i)}},\ t_{j,j+1}^{(l_j)} = [a; b);$$

$$p\left(t_{\min}^{(l_3)} \middle| n,\ \lambda^{(i)}\right) = e^{-an\lambda^{(i)}} - e^{-bn\lambda^{(i)}},\ t_{\min}^{(l_3)} = [a; b);$$

$$p\left(n \middle| \lambda^{(i)}\right) = \frac{\left(\lambda^{(i)}T\right)^n}{n!} e^{-\lambda^{(i)}T};$$

$$p\left(t_{\max}^{(l_4)} \middle| n,\ \lambda^{(i)},\ t_{\min}^{(l_3)}\right) = e^{(n-1)\lambda^{(i)}t_{\min}^{(l_3)}} \times \left(\left(e^{-\lambda^{(i)}t_{\min}^{(l_3)}} - e^{-\lambda^{(i)}b}\right)^{n-1} - \left(e^{-\lambda^{(i)}t_{\min}^{(l_3)}} - e^{-\lambda^{(i)}a}\right)^{n-1}\right),\ t_{\max}^{(l_4)} = [a; b).$$

For all further examples we used the following discretization: for the rate variable λ $\lambda^{(1)} = [0, 0.01)$, $\lambda^{(2)} = [0.01, 0.03)$, $\lambda^{(3)} = [0.03, 0.05)$, $\lambda^{(4)} = [0.05, 0.1)$, $\lambda^{(5)} = [0.1, 0.2)$, $\lambda^{(6)} = [0.2, 0.5)$, $\lambda^{(7)} = [0.5, 1)$, $\lambda^{(8)} = [1, \infty)$; for the variables $t_{j,j+1}$, $t_{\min}$, $t_{\max}$ $t^{(1)} = [0, 0.1)$, $t^{(2)} = [0.1, 1)$, $t^{(3)} = [1, 7)$, $t^{(4)} = [7, 30)$, $t^{(5)} = [30, 180)$, $t^{(6)} = [180, \infty)$.

The model was represented in GeNIe&Smile [20]. The calculations and statistical analysis were performed using Smile library [20] and R [22].

2.3 Testing Results

To test the model we used automatically generated dataset. First, we generate 200 values for behavior rate that followed Gamma distribution with the shape $k = 1.1$ and the scale $\theta = 0.1$. The parameters were chosen to produce more "real-behavior"-like dataset with behavior rate in most cases less than 1 and concentrated around 0.1 episodes per day. Next, for each rate value according to assumptions of the model we generated 20 "respondents" or 20 sequences of

behavior episodes each of those for period of 180 days in total. In other words, we had 4000 sequences of episodes. Than we calculated lengths of minimum, maximum intervals between episodes and lengths of intervals between the last three episodes. After deletion of incomplete cases (e.g. cases with only one episode during 180 days) the final dataset consisted of 3672 cases (or "respondents").

All variables were discretized as it was described in Sect. 2.2. The prior distribution of behavior rate for BBN model presented on Fig. 2: the probability of the last episode was set to 0.125, the remaining part was distributed uniformly according to the length of the interval.

We considered both individual predictions and prediction for the whole group. The predicted rate value for each case or "respondent" was the interval with maximum posterior probability. Table 1 shows the confusion matrix for the individual predictions. The average accuracy [23] for this eight-class classification was 87.7 %. Moreover, the false positive predictions in most cases belonged to adjacent intervals.

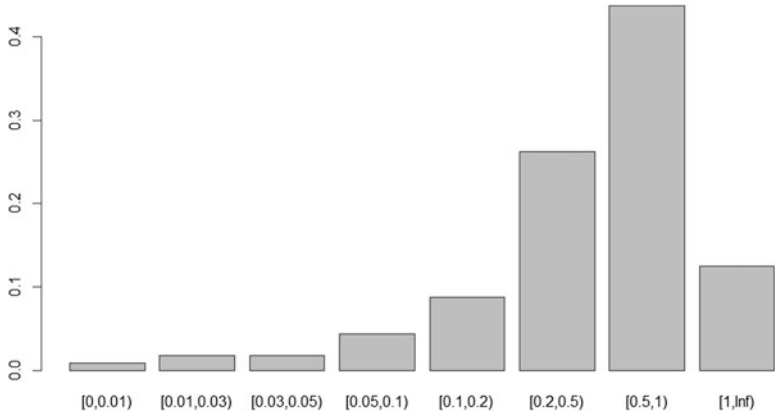


Fig. 2 Prior rate distribution

Table 1 Confusion matrix for the individual predictions

Initial value	Predicted value							
	$\lambda^{(1)}$	$\lambda^{(2)}$	$\lambda^{(3)}$	$\lambda^{(4)}$	$\lambda^{(5)}$	$\lambda^{(6)}$	$\lambda^{(7)}$	$\lambda^{(8)}$
$\lambda^{(1)}$	6	12	3	1	0	0	0	0
$\lambda^{(2)}$	83	310	85	36	1	0	0	0
$\lambda^{(3)}$	11	205	145	123	6	0	0	0
$\lambda^{(4)}$	0	121	249	470	153	7	0	0
$\lambda^{(5)}$	0	3	18	252	568	119	0	0
$\lambda^{(6)}$	0	0	0	6	285	369	0	0
$\lambda^{(7)}$	0	0	0	0	0	20	0	0
$\lambda^{(8)}$	0	0	0	0	0	0	0	0

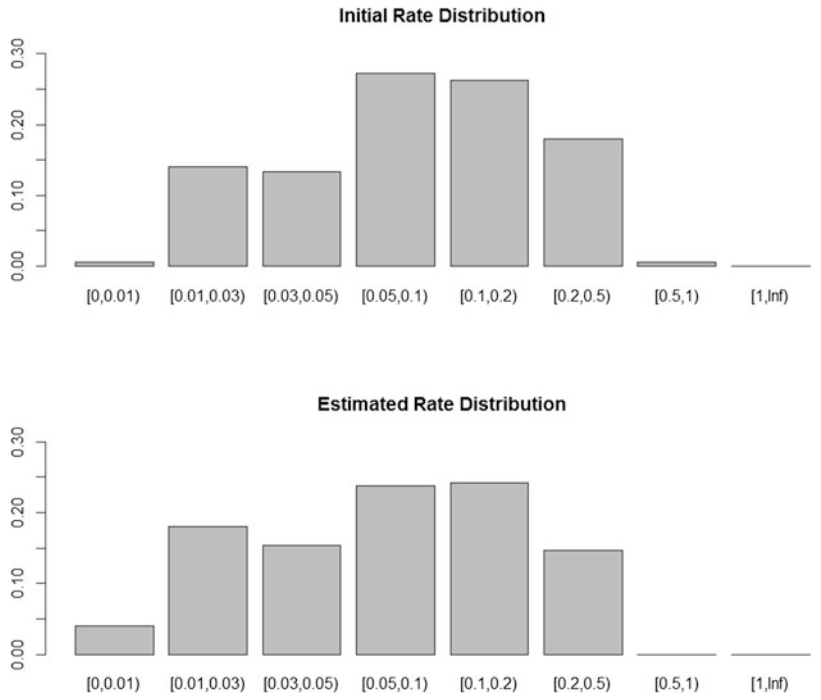


Fig. 3 Comparison of the initial and the estimated rate distributions

The initial rate distribution for generated dataset and comparison with the rate distribution estimated on the base of data about episodes is presented on Fig. 3. The estimated rate distribution was the result of inference on the BBN combined for the group of "respondents" according to Dirichlet distribution [24]. The χ^2 test for comparison the distributions showed that we did not reject the hypothesis about similarity of distributions ($\chi^2 = 48$, $df = 42$, $p-\text{value} = 0.24$). In other words, the estimated rate distribution looked similar to the initial distribution with major part of cases concentrated between 0.05 and 0.2 cases per day.

2.4 Real Data Example

To illustrate how the model works for real data we considered data about the last episodes of alcohol consumption and data about minimum and maximum intervals between the episodes during the last 6 months. The data about 380 respondents was collected in 2011 among patient of one of STD clinics in St.Petersburg, Russia. We used the same prior rate distribution that was described in Sect. 2.3 (Fig. 2). The estimated distribution presented on Fig. 4.

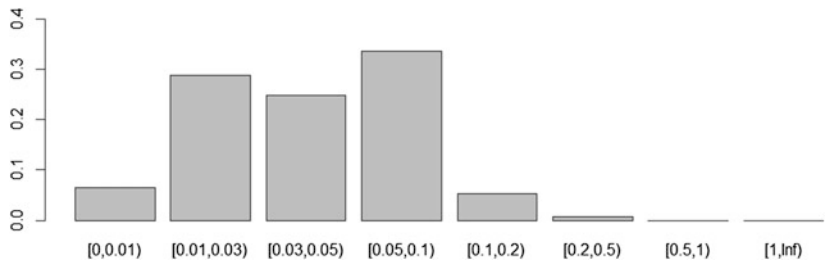


Fig. 4 Estimated rate distributions for alcohol consumption

According to the estimates we can conclude that with probability greater than 0.9 respondents in the sample drank alcohol rarely than one tome per 10 days. At the same time, the probability of really rare consumption was relatively low too that allowed to conclude that in general there was a need for behavior-changing events.

3 Conclusion

The area of risky behavior modelling has unsolved issues in current practice. To provide a tool for estimating behavior characteristics we proposed the model in terms of Bayesian Belief Networks that combined both empirical data and expert knowledge about behavior. Testing the model on automatically generated dataset showed good results. The rate estimates for real data illustrated the possible conclusions in practical issues. However, for practical usage the model requires more detailed testing with different initial rate distributions and, on the other hand, testing on data about real behavior with available data about episodes and real rate.

Acknowledgments This work was partially supported by the by RFBR according to the research projects No. 14-01-00580 and No. 16-31-60063.

References

1. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge (2000)
2. Du, Y., Guo, Y.: Evidence reasoning method for constructing conditional probability tables in a Bayesian network of multimorbidity. Technol. Health Care, 23(s1) (2015)
3. Jitwasinkul, B., Hadikusumo, B.H., Memon, A.Q.: A Bayesian belief network model of organizational factors for improving safe work behaviors in Thai construction industry. Saf. Sci. **82**, 264–273 (2016)
4. Trucco, P., Cagno, E., Ruggeri, F., Grande, O.: A Bayesian belief network modelling of organisational factors in risk analysis: a case study in maritime transportation. Reliab. Eng. Sys. Saf. **93**(6), 845–856 (2008)

5. Semakula, H.M., Song, G., Achuu, S.P., Zhang, S.: A Bayesian belief network modelling of household factors influencing the risk of malaria: a study of parasitaemia in children under five years of age in sub-Saharan Africa. Environ. Model Softw. **75**, 59–67 (2016)
6. Sun, L., Erath, A.: A Bayesian network approach for population synthesis. Transp. Res. Part C: Emerg. Technol. **61**, 49–62 (2015)
7. Neapolitan, R.E.: Learning Bayesian Networks. Pearson Prentice Hall (2003)
8. Tulupyev, A., Nikolenko, S., Sirotkin, A.: Bayesian Networks: A Probabilistic Logic Approach. Nauka, SPb (2006)
9. Mkrtchyan, L., Podofillini, L., Dang, V.N.: Bayesian belief networks for human reliability analysis: a review of applications and gaps. Reliab. Eng. Syst. Saf. **139**, 1–16 (2015)
10. Leigh, B.C., Stall, R.: Substance use and risky sexual behavior for exposure to HIV: issues in methodology, interpretation, and prevention. Am. Psychol. **48**(10), 1035 (1993)
11. Bolger, N., Davis, A., Rafaeli, E.: Diary methods: capturing life as it is lived. Annu. Rev. Psychol. **54**(1), 579–616 (2003)
12. Schroder, K.E., Carey, M.P., Vanable, P.A.: Methodological challenges in research on sexual risk behavior: II. Accuracy of self-reports. Ann. Behav. Med. **26**(2), 104–123 (2003)
13. Graham, C.A., Catania, J.A., Brand, R., Duong, T., Canchola, J.A.: Recalling sexual behavior: a methodological analysis of memory recall bias via interview using the diary as the gold standard. J. Sex Res. **40**(4), 325–332 (2003)
14. Tulupyeva, T., Paschenko, A., Tulupyev, A., Krasnoselskikh, T., Kazakova, O.: HIV Risky Behavior Models in The Context of Psychological Defense and Other Adaptive Styles. Nauka, SPb (2008)
15. Ramrakha, S., Caspi, A., Dickson, N., Moffitt, T.E., Paul, C.: Psychiatric disorders and risky sexual behaviour in young adulthood: cross sectional study in birth cohort. BMJ **321**(7256), 263–266 (2000)
16. Lemelin, C., Lussier, Y., Sabourin, S., Brassard, A., Naud, C.: Risky sexual behaviours: the role of substance use, psychopathic traits, and attachment insecurity among adolescents and young adults in Quebec. Can. J. Hum. Sex. **23**(3), 189–199 (2014)
17. Vacirca, M.F., Ortega, E., Rabaglietti, E., Ciairano, S.: Sex as a developmental transition: the direct and indirect roles of peers. Psychol. Sex. **3**(2), 108–122 (2012)
18. Suvorova, A.: Socially significant behavior modeling on the base of super-short incomplete set of observations. Inf. Measur. Control Syst. **9**(11), 34–38 (2013)
19. Stepanov, D.V., Musina, V.F., Suvorova, A.V., Tulupyev, A.L., Sirotkin, A.V., Tulupyeva, T. V.: Risky behavior Poisson model identification: heterogeneous arguments in likelihood. Trudy SPIIRAN **23**, 157–184 (2012)
20. GeNIe& SMILE: Decisions Systems Laboratory. School of Information Sciences. University of Pittsburg. http://genie.sis.pitt.edu/
21. AgenaRisk Bayesian Network Tool. http://www.agenarisk.com
22. Core Team. *R:* A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2015). http://www.R-project.org/
23. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Inf. Process. Manage. **45**(4), 427–437 (2009)
24. Suvorova, A.V., Tulupyev, A.L., Sirotkin, A.V.: Bayesian belief networks for risky behavior rate estimates. Nechetkie sistemy i myagkie vychisleniya (Fuzzy Syst. Soft Comput.) **9**(2), 115–129 (2014)

Back Up Data Transmission in Real-Time Duplicated Computer Systems

S.A. Arustamov, V.A. Bogatyrev and V.I. Polyakov

Abstract The authors consider several options for the organization of inter computers exchange between duplicated computer systems through backed up switching facilities. They also analyze the impact of the selected option of inter computer exchange through backed up switches in a view of the probability of timeliness and error-free transmission of packages in real-time systems.

Keywords Fault tolerance · Duplicated computer system · Reliability · Redundancy · Real-time

1 Introduction

Fault tolerance and reliability of IT and commutations systems is reached as a result of redundancy and rejoining of resources, including the presence of data transmission channels back up [1].

For a real-time communication systems providing reliability of data transmission function includes the structural reliability and timely information delivery under conditions of possible communication faults.

For information technologies (IT) and communication systems with duplicated data channels potential improvement of the transmission reliability is feasible when a redundant transmission of copies of packages through multiple channels simultaneously has been implemented. For multi-channel queuing systems with a general queue in [2–4] the authors propose the simultaneous maintenance of each query copy by several devices. For service option called «broadcasting service with a

S.A. Arustamov · V.A. Bogatyrev · V.I. Polyakov (✉)
ITMO University, 49 Kronverksky Pr., Saint Petersburg 197101, Russia
e-mail: v_i_polyakov@mail.ru

S.A. Arustamov
e-mail: sergey.arustamov@gmail.com

V.A. Bogatyrev
e-mail: Vladimir.bogatyrev@gmail.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_11

copy request» [4], a request is routed to the non-busy equipment at the time of appearance, the service in each node of a copy (back up) of the request is being made independently.

In [9] another modification of a broadcast service is considered where, if the number of devices employed at the time of the request is less than a certain threshold, the request is backed up (copied) to all available nodes, otherwise it is served by a single node. As the main key productivity indicator (KPI) of efficiency of maintenance in [2–4] the authors consider the average residence time of requests in the system. However, for real-time systems as the main performance indicator it's recommended to deploy the probability there is no a timeout request exceeds predefined maximum accepted time.

Models of reliable and timely service in clusters with real-time execution of redundant copies of requests have been considered in [5].

In this paper we are developing models described in [5] applicable for systems composed out of computer systems with duplicated data channels, each of them being interpreted as a single-channel queuing system with unlimited queue.

Timeliness of data transmission in the IT and communication real-time systems may be assessed of transmission waiting time the probability for the time less than the limit t_0 [5].

2 The Structure of the System Under Study

As the object of study, consider the system composed from duplicated computer systems (DCS), presented in Fig. 1.

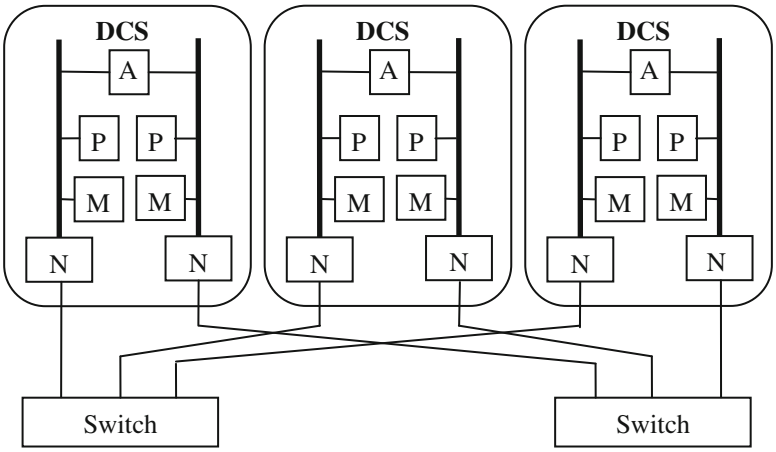


Fig. 1 Duplicated computer system structure

Each part of DSC includes a processor (P), the memory module (M), network adapters (NA). In addition an adapter for interconnection between parts of dual systems has also been included into the structure. Performing of critical queries is duplicated in both parts of the dual systems. When combining duplicate complexes in the fault-tolerant system we allocate m production and M-m backup systems. For doubling computing data transmitted between production and backup systems must be recorded into memory M modules of both parts of a dual system of addressable node. Once operation is performed without duplication of calculations every query is executed in one part of DCS. In this case the distribution of the flow of requests that do not require duplication of computing is supported by balancing of both parts of dual system [6–11].

In the system including M duplicated parts the ability to operate performing duplicated calculations is required at least in m duplicated parts. At that, between each pair of m operational complexes an exchange with the delivery of information transmitted to the memory unit at two duplicated receiver information (receiver may be any of m operation parts) must be guaranteed.

In modern cluster systems, consolidating the resources of several servers, the peculiarity of duplicated query execution is provided by organization independent queues for each server, to represent a single-channel queuing system. In such systems, unlike multi-channel systems with a shared queue [7–9], service time for execution query scattered drastically that increases queries in different lines, including requests for different duplication systems.

This case may potentially decrease the probability of timeliness transmission of packages between interactive nodes.

On the one hand a redundant transmission copy of packets through different switching nodes, each of which is a single-channel queuing system with infinite queue increases the stability of the computing process to failures and bugs in the nodes. On the other hand, the reservation request causes an increase in workload communication nodes and average waiting time for requests and, as a consequence, to an increased probability of delay queued requests over the maximum allowable time t0. At the same time, redundant transmission of packages through multiple switching nodes, based on stochastic service could potentially lead to an increase in the probability of timely transmission of packets provided they are timely transmitted of at least through one of the backup paths. Thus, there is a structural contradiction, requiring research targeting to formulate recommendations for a solution for cluster real-time systems whose key performance indicator of efficiency is the probability of timely service requests seems quite topical.

The research aims to study the organization of an exchange between the parts of dual computing, taking into consideration the impact on structural reliability and probability of timely exchange in redundant computer systems.

To achieve this goal the authors is conducting:

- Development of options of data exchange between duplicate computer systems via a redundant (duplicated) communication environment, including options with redundant copies of multipath transmission packages;

- Assessment of the impact of options of duplicated exchange system on reliability of their performance and timeliness exchange under redundant data transmission;
- Choice of rational option between duplicate complexes based on achieved operational reliability and the probability of not exceeding the maximum permissible transmission delay time value expectations.

3 Options for the Organization of Exchange Between DCS

Let's define options for organizing the exchange between the DCS with duplicate calculations when entering the transmitted data is required to the two memory modules M of both parts of DCS addressable computing node.

For systems with duplicated computing separate the following options of transmission from the ith to the jth DKS:

Option B1: The first (left) and second (right) parts of ith DCS transmit packages correspondently through the first Sw1 and second switch Sw2.

As a result of exchange SA1i → Sw1 → SA1j M1; SA2i → Sw2 → SA2j M2, ith data set recorded in the memory of both part of memory modules M of jth DCS.

Necessary conditions required for option B1 applicability are: operation ability of both switches and both DCS network adapters while packages are being transmitted from the i-th to j-th complex (node).

Consider the two version of option B1:

- Implementation of the first modification B11 presumes organization of the independent queues for transmission through each line (multiple accesses to each line is performed regardless of access to the second line);
- Unlike B11 the second modification B12 organize common request queue for transmission through the first and second lines with an implementation of unique procedure of multiple access: as a result both lines are available simultaneously and both copies of requests are transmitted through them.

Option B2: The first (or second) part of *i*-th complex forwards the packet through the first (second) switch and the appropriate network adapter in the memory of the first (second) part of *j*th DKS. From memory module that received the data, they are transmitted to the memory module of another part of DCS through internal communication adapter A.

Thus, the following connection is implemented $SA_{1i} \rightarrow Sw_1 \rightarrow SA_{1j}\ M_1$; $M_{1j} \rightarrow A_j\ M_{2j}$. Mapping parts of DCS to transfer switches is conducted to balance their workload.

Conditions for implementation of variant B2 during the transmission from the *i*th to the *j*th complex is the availability of at least of the one switch and its connection via serviceable ports and network adapters simultaneously to the first (or second) part of DCS interacting nodes provided serviceability adapter for internal adapters in unit receiver. When duplex communication is considered the operation ability of internal adapter is required for each of interactive part of DCS.

Note that the transmission though adapter can be implemented in parallel as DCSs are located closely (for instance, in one form factor).

We can use the also hybrid option, which allows to adapt to the intensity of DCS exchange and system component failures. For example, the option may be used presuming the usage of mode B1 under low intensity with switching to B2 in case the increase of workload. It is also reasonable to envisage combined option when originally B2 is applied without accumulation of exchange failures with switching to mode B1 after failures of internal communication switches that potentially can improve fault tolerance ability of the system.

To adaptively increase robustness of the system we can recommend to use mode B3, wherein we initially implement option B12. In case of non-delivery within the specified time packages through two channels we may use by the results of a successful error-free transmission over a single channel with the further implementation of the exchange between part of DCS. The last option of inter DCS exchange is applicable with operation ability of internal exchange adapters, with one of the copies of the two channels transmitted package must be transmitted through the inter DCS exchange during $t_0 - t_1$, where t_1 time of internal exchange.

4 Evaluation of the Probability of Timeliness Inter DCS Exchange

In assessing the residence time of requests for different inter DCS exchange each switching node represents the queuing system type M/M/1.

In the initial state of the system (without failure), the probability of not exceeding the maximum accepted latency transmission delay t0 using options B11, B12, B2 and B3 in the error-free transmission of packets to be calculated as follows:

$$
\begin{aligned}
d_{11} &= 1-\Lambda v \exp(-t_0(v^{-1}-\Lambda)),\\
d_{12} &= [1-\Lambda v \exp(-t_0(v^{-1}-\Lambda))]^2,\\
d_2 &= 1-(\Lambda/2) v \exp(-(t_0-t_1)(v^{-1}-\Lambda/2)),\\
d_3 &= 1-[\Lambda v \exp(-(t_0-t_1)(v^{-1}-\Lambda))]^2
\end{aligned}
$$

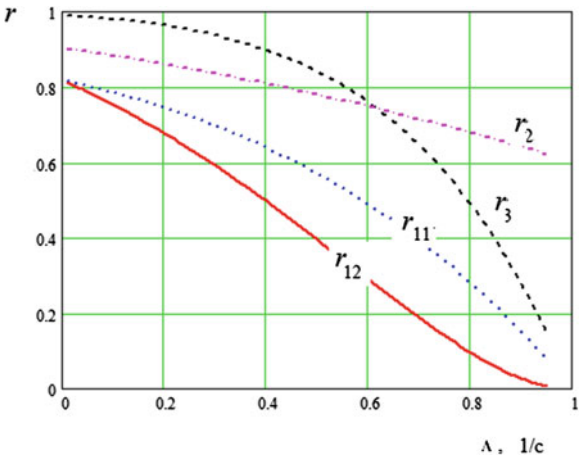


Fig. 2 The probability of exceeding waiting queue maximum allowable packet delay t_0

and with consideration of possible faulty transmissions

$$r_{11} = [1-\Lambda v \exp(-t_0(v^{-1}-\Lambda))](1-\lambda)^{2L},$$
$$r_{12} = \{[1-\Lambda v \exp(-t_0(v^{-1}-\Lambda))](1-\lambda)^{L}\}^2,$$
$$r_2 = [1-(\Lambda/2)v \exp(-(t_0-t_1)(v^{-1}-\Lambda/2))](1-\lambda)^{L},$$
$$r_3 = 1-(1-g)^2, g = \{1-[\Lambda v \exp(-(t_0-t_1)(v^{-1}-\Lambda))]\}(1-\lambda)^{L},$$

where $v = L/b$ average time of transmission of a average package volume L, bit rate b, λ—bit error probability, Λ—the intensity of the requests for package transmission.

Dependence of the probability of not exceeding expectations packet queue maximum accepted delay $t_0 = 1$ s is presented in Fig. 2 by the curves r_{11}, $-r_{12}$, r_2, r_3.

The calculation is performed under $t_1 = 0.2$ s, $L = 10^4$ bit, $b = 10^4$ bit/s и $\lambda = 10^{-4}$ 1/bit.

The calculations show an evident advantage of option B3, under low intensive queries and increase the efficiency of the exchange option B2 in their growth. Option B3 also increases the marginal rate of requests and robustness of the system under steady-on servicing transfer requests packets.

5 Conclusion

The authors conducted a research of options for the exchange in DCS through redundant switches in redundant cluster computer systems.

The research demonstrated the influence of the exchange organization in DCS on the probability of timely servicing of requests for the transfer of packets through redundant data transmission medium.

Additional features provide a high probability of error-free data transmission that timely provides adaptive change options exchange in the accumulation of failures or change in intensity requests in DCS exchange.

The results can be used to justify the choice of the organization of exchange in fault-tolerant systems, redundant computer systems, including those working in real time.

References

1. Koren, I.: Fault Tolerant Systems. Morgan Kaufmann Publications, visit our San Francisco, 378 pp. (2009)
2. Lee, M.H., Dudin, A.N., Klimenok, V.I.: The SM/V/N queueing system with broadcasting service. Math. Probl. Eng. **2006**, Article ID 98171, 18 (2006)
3. Dudin, A.N., Sun, B.: A multiserver MAP/PH/N system with controlled broadcasting by unreliable servers. Autom. Control Comput. Sci. **5**, 32–44 (2009)
4. Dudin A.N., Sun, B.: Unreliable multi-server system with controllable broadcasting service. Autom. Remote Control **70**(12), 2073–2084
5. Bogatyrev, V.A., Bogatyrev, A.V.: Functional reliability of a real-time redundant computational process in cluster architecture systems. Autom. Control Comput. Sci. **49**(1), 46–56 (2015)
6. Bogatyrev, V.A.: Exchange of duplicated computing complexes in fault tolerant systems. Autom. Control Comput. Sci. **46**(5), 268–276 (2011)
7. Bogatyrev, V.A., Bogatyrev, S.V., Golubev I.Y.: Optimization and the process of task distribution between computer system clusters. Autom. Control Comput. Sci **2012** (3), 103–111 (2011)
8. Bogatyrev, V.A.: Fault tolerance of clusters configurations with direct connection of storage devices. Autom. Control Comput. Sci. **45**(6), 330–337 (2011)
9. Bogatyrev, V.A.: An interval signal method of dynamic interrupt handling with load balancing. Autom. Control Comput. Sci. **34**(6), 51–57 (2000)
10. Bogatyrev, V.A.: Protocols for dynamic distribution of requests through a bus with variable logic ring for reception authority transfer. Autom. Control Comput. Sci. **33**(1), 57–63 (1999)
11. Lee, M.H., Dudin, A.N., Klimenok, V.I.: The SM/V/N queueing system with broadcasting service. Math. Probl. Eng. **2006**, Article ID 98171, 18 (2006)

Data Coherence Diagnosis in BBN Risky Behavior Model

Aleksandra V. Toropova

Abstract A problem of coherence diagnosis for risky behavior model based on the data about behavior episodes retrieved from an interview with a respondent is considered. The extension of the model is described and the examples of data coherency diagnostics are provided. For more convenient work with suggested method the software is provided.

Keywords Bayesian belief network · BBN · Data coherence diagnosis

1 Introduction

Many studies on Artificial Intelligence face with the problem of data coherence diagnosis [1–4]. Many models and decision-making systems include expert knowledge and expert estimates. Expert methods are used in situations where selection, justification and impact assessment cannot be made using exact calculations [5]. Researchers can use the information received from the experts only if they have a possibility to present it in a form suitable for further research. Therefore, it is necessary to either formalize the experts' knowledge, or evaluate its coherence and reliability [6, 7]. Such problems arise in studies focused on troubleshooting power system [4, 8], the diagnosis of distributed systems security problems anomalies [1, 9], as well as in other areas. In all these cases, the coherence of the data is extremely important.

In many fields of sociological, psychological and marketing research, we face the problem of risky behavior rate or frequency estimate on the basis of respondents' self-reports about their behavior. We need to estimate behavior rate using the responses to the questionnaire or the results of the interview [8, 10].

A.V. Toropova (✉)
SPIIRAS, 39, 14 Line, Saint Petersburg 199178, Russia
e-mail: alexandra.toropova@gmail.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_12

An approach to the risky behavior rate estimate based on Bayesian belief networks and data obtained from interviews about last episodes of respondent's behavior is proposed in [10–12].

The initial model [11] was based on the data about the three latest episodes of respondents' risky behavior and minimum and maximum intervals between the episodes. These data were usually obtained from questionnaires or interviews [13]. Respondents could give false (not corresponding to actual behavior) answers to make a positive impression or due to memory-related issues: episodes of risky behavior could happen a long time ago and, hence, be hard to remember. For example, the risky sexual behavior data (information needed for decision-making in various areas, including education, medicine and public health) was under-reported very often due to its very private nature, and such data often became a subject of significant social desirability bias. Sometimes respondents answering question could make mistakes or be confused [14, 15].

As an example, consider the following scenario. During an interview conducted on Monday, the respondent replied that the last behavior episode was on the last Monday, the previous one was on the last Wednesday and the last but two episode was a month ago. At the same time, the respondent defined the minimum interval between episodes as a "week". Hence, the provided data were incoherent because the interval between the last episode and the previous one was less than the minimum.

Note, that this is a very simplified example of the problem; obviously such inconsistencies can be easily identified. However, there are possible more complex situations because of the sampling variables included in the model.

Thus, applications that used data obtained from respondents often faced with the problem of incoherent data. Therefore it is important to have tools to diagnose such situations.

In the paper we describe modified model that solves this problem. For more convenient work with the model software is provided. Also we discuss an extended example of the model usage.

2 Model Description

Figure 1 shows a generalized risky behavior model $M = (G(V, L), \mathbf{P})$ as a Bayesian belief network [16, 17]. The model structure is represented by the directed graph $(G(V, L))$, where $V = \{t_{01}, t_{12}, t_{23}, t_{\min}, t_{\max}, \lambda, n\}$ is corresponded to the set of nodes, $L = \{(u, v) : u, v \in V\}$ is corresponded to the set of directed links between nodes. In other words, Fig. 1 shows random elements included in the model and relations between them. We used GeNIe 2.0 [18] to create Bayesian belief network and to implement the probabilistic reasoning algorithms. All figures were also constructed in GeNIe 2.0.

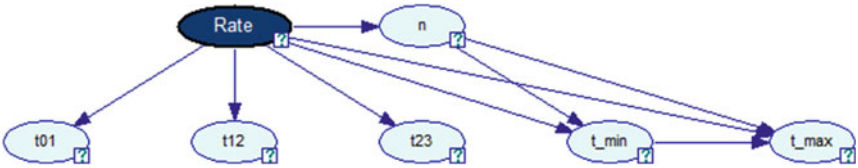


Fig. 1 Risky behavior model on the basis of data about episodes

On Fig. 1, *Rate* is a random variable representing the behavior rateλ, $t_{i,j}$ are random variables characterizing the lengths of the interval between the *i*th and *j*th to the end episodes. With an assumption that behavior was a Poisson random process random variables $t_{i,j}$ were exponentially distributed. The additional information was obtained by including minimum and maximum intervals between episodes ($t_{\min}$ and $t_{\max}$ respectively).

We specified conditional probabilities $\mathbf{P} = \{P(t_{j,j+1}|\lambda), P(t_{01}|\lambda), P(t_{\min}|n,\lambda), P(t_{\max}|n,\lambda,t_{\min}), P(n|\lambda), P(\lambda)\}$, (edges between conditionally dependent nodes) as follows:

$$
\begin{aligned}
P\left(t_{j,j+1}^{l_i j}|\lambda^{(i)}\right) &= e^{-a\lambda^{(i)}} - e^{-b\lambda^{(i)}},\ j = 0,1,2,\\
t_{j,j+1}^{l_i j}[a;b]; P\left(t_{\min}^{l_3}|n,\lambda^{(i)}\right) &= e^{-an\lambda^{(i)}} - e^{-bn\lambda^{(i)}},\ t_{\min}^{l_3} = [a;b];\\
p\left(n\middle|\lambda^{(i)}\right) &= \frac{\left(\lambda^{(i)}T\right)^n}{n!}e^{-\lambda^{(i)}T};\\
p\left(t_{\max}^{(l_4)}\middle|n,\lambda^{(i)},t_{\min}^{(l_3)}\right) &= e^{(n-1)\lambda^{(i)}t_{\min}^{(l_3)}}\left(\left(e^{-\lambda^{(i)}t_{\min}^{(l_3)}} - e^{-\lambda^{(i)}b}\right)^{n-1} - \left(e^{-\lambda^{(i)}t_{\min}^{(l_3)}} - e^{-\lambda^{(i)}a}\right)^{n-1}\right),\\
t_{\max}^{(l_4)} &= [a;b).
\end{aligned}
$$

3 Model Extension

Figure 2 shows extended risky behavior model. The added nodes allow to estimate data given by a respondent.

The nodes $c_{t_{1,2,\min}}$ and $c_{t_{23,\min}}$ represent episode t_{ij} and minimal interval $t_{\min}$ coherence, the nodes $c_{t_{0,1,\max}}, c_{t_{12,\max}}$ and $c_{t_{23,\max}}$ represent episode t_{ij} and maximal interval $t_{\max}$ coherence. We did not consider $c_{t_{0,1,\min}}$, because t_{01} represents an interval between an risky behavior episode and the moment of interview, which is not an observing behavior episode.

In particular, for the node representing the coherence degree with a minimum interval, coherence rate $c_{t_{ij,\min}}$ could take the following three values: the values:t_{ij} and t_{min} were *coherent* ($c^{+}_{t_{ij,\min}}$), values were *incoherent* ($c^{-}_{t_{ij,\min}}$) and values were *undefined* ($c^{?}_{t_{ij,\min}}$). We assumed that the rate $c_{t_{ij,\min}}$ was undefined when both t_{ij} and

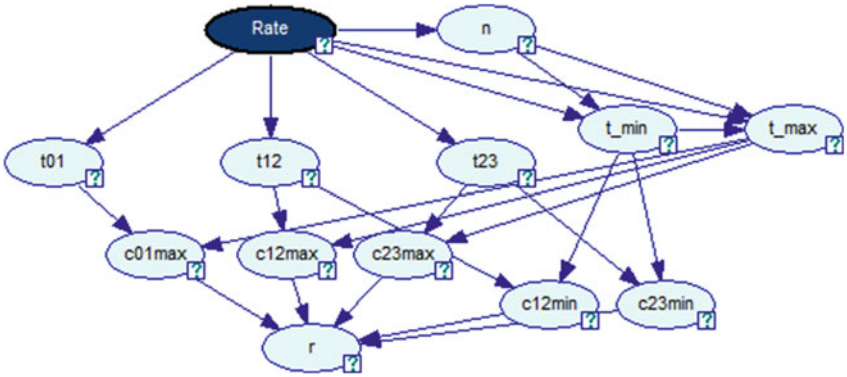


Fig. 2 Extended risky behavior model on the basis of data about episodes

t_{min} belong to the same intervals, i.e. if $t_{ij} \in [a;b)$ and $t_{min} \in [a;b)$ we could not define precisely whether the value t_{min} was smaller than t_{ij} or not.

We specified conditional probabilities of the extended model as follows:

$$P\left(c^{(s)}_{t_{ij,\min}} | t_{ij}, t_{\min}\right) = \begin{cases} \alpha^{(s)}, & t_{ij} > t_{min}; \\ \beta^{(s)}, & t_{ij} < t_{min}; \\ 1 - \alpha^{(s)} - \beta^{(s)}, & t_{ij} = t_{\min}; \end{cases}$$

where $s \in \{+, -, ?\}, \alpha^{(s)}, \beta^{(s)} \in [0;1], \sum \alpha = 1, \sum \beta = 1, \alpha^{(s)} + \beta^{(s)} \leq 1$.

Similarly, we obtained the estimation of the coherence of the random variables t_{ij} corresponding to the intervals between the last episodes realizations with the realization of a random variable $t_{\max}$ ($c_{t_{0,1,\max}}, c_{t_{12,\max}}$ and $c_{t_{23,\max}}$):

$$p\left(c^{(s)}_{t_{ij,\max}} | t_{ij}, t_{\max}\right) = \begin{cases} \alpha^{(s)}, & t_{ij} < t_{\max}; \\ \beta^{(s)}, & t_{ij} > t_{\max}; \\ 1 - \alpha^{(s)} - \beta^{(s)}, & t_{ij} = t_{\max}; \end{cases}$$

where $s \in \{+, -, ?\}, \alpha^{(s)}, \beta^{(s)} \in [0;1], \sum \alpha = 1, \sum \beta = 1, \alpha^{(s)} + \beta^{(s)} \leq 1$.

To estimate respondent reliability ($\boldsymbol{r}$) we added a node connecting all these five new nodes characterizing the pairwise coherence.

To simplify the formulae for conditional probabilities let $c = \left(c_{t_{12,\min}}, c_{t_{23,\min}}, c_{t_{01,\max}}, c_{t_{12,\max}}, c_{t_{23,\max}}\right)$, $c^{+} = \left(c^{+}_{t_{12,\min}}, c^{+}_{t_{23,\min}}, c^{+}_{t_{01,\max}}, c^{+}_{t_{12,\max}}, c^{+}_{t_{23,\max}}\right)$, $c^{-} = \left(c^{-}_{t_{12,\min}}, c^{-}_{t_{23,\min}}, c^{-}_{t_{01,\max}}, c^{-}_{t_{12,\max}}, c^{-}_{t_{23,\max}}\right)$, $c^{?} = \left(c^{?}_{t_{12,\min}}, c^{?}_{t_{23,\min}}, c^{?}_{t_{01,\max}}, c^{?}_{t_{12,\max}}, c^{?}_{t_{23,\max}}\right)$.

Then $p(\boldsymbol{r}^{+}|c) = \frac{\sum c^{+}}{\sum c}, p(\boldsymbol{r}^{-}|c) = \frac{\sum c^{-}}{\sum c}$ and $p\left(\boldsymbol{r}^{?}|c\right) = \frac{\sum c^{?}}{\sum c}$.

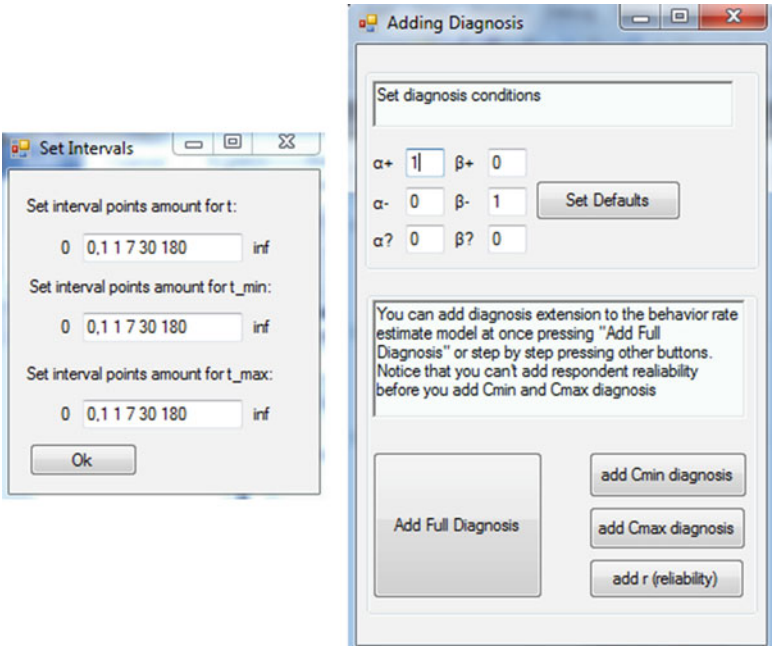


Fig. 3 Setting intervals and adding diagnosis windows

4 Realization

We created software to supplement the risky behavior model with mentioned before diagnosis nodes and for more convenient work with this model. The software was developed by using C# and Smile library [18]. Firstly user defines model: sets intervals for t_{ij}, t_{min} and $t_{\max}$ (Fig. 3); sets $\alpha^{(s)}, \beta^{(s)}$ where $s \in \{+, -, ?\}$ and add diagnosis nodes to the model, it can be made at once or step by step (Fig. 3). After that respondents data can be inserted into the model, input can be made manually then results are shown in the same window (Fig. 4) or from MS Excel file in this case results are saved in a separate file.

5 Example

Let t_{ij} to be divided into these disjunctive intervals: $t^{(1)} = (0; 0, 1)$, $t^{(2)} = [0, 1; 1)$ $t^{(3)} = [1; 7)$, $t^{(4)} = [7; 30)$, $t^{(5)} = [30; 180)$, $t^{(6)} = [180; +\infty)$, for clarity we take the same partition for $t_{\min}$ and $t_{\max}$.

We assumed that the coherence probability was zero, if the data provided by the respondent contradicted each other, and one, if there were no contradictions.

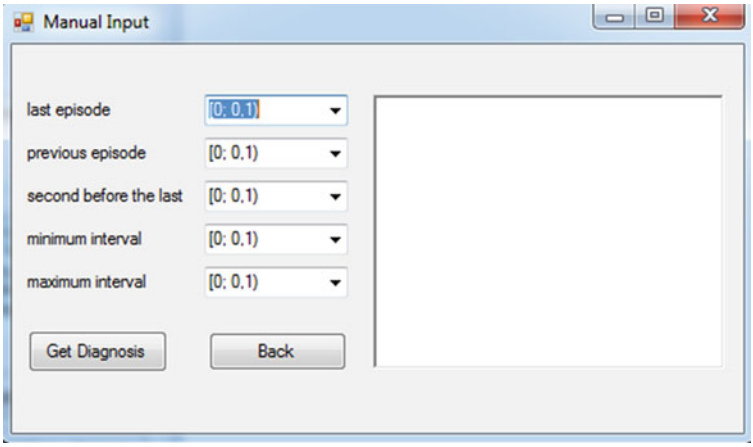


Fig. 4 Manual input window

We considered the example with ten respondents' given data. The data are presented in the Table 1, the first column contains a respondent's id, the other columns contain respondent's evidences about last risky behavior episodes (t_{01}, t_{12}, t_{23}) and minimal and maximal interval evidences ($t_{\min}$ and $t_{\max}$).

Let us have a closer look to the second respondent's data, particularly the coherence estimation of episode t_{12} and minimal interval $t_{\min}$. In this case data is incoherent. The posterior distribution of the coherence random variable is shown in Fig. 5.

After all the coherence estimations were defined, we estimated the respondent's reliability. The second respondent's reliability estimation is presented in Fig. 6.

Table 1 Respondents' data

Respondent's id	t_{01}	t_{12}	t_{23}	$t_{\min}$	$t_{\max}$
1	$[0, 1; 1)$	$[1; 7)$	$[0, 1; 1)$	$(0; 0, 1)$	$[7; 30)$
2	$[7; 30)$	$[0, 1; 1)$	$[7; 30)$	$[1; 7)$	$[30; 180)$
3	$[7; 30)$	$[1; 7)$	$(0; 0, 1)$	$(0; 0, 1)$	$[7; 30)$
4	$[0, 1; 1)$	$[0, 1; 1)$	$[0, 1; 1)$	$(0; 0, 1)$	$[180; +\infty)$
5	$[30; 180)$	$[7; 30)$	$[1; 7)$	$(0; 0, 1)$	$[180; +\infty)$
6	$[0, 1; 1)$	$[0, 1; 1)$	$[0, 1; 1)$	$(0; 0, 1)$	$[0, 1; 1)$
7	$[1; 7)$	$[1; 7)$	$[1; 7)$	$[0, 1; 1)$	$[7; 30)$
8	$[30; 180)$	$[30; 180)$	$[30; 180)$	$[30; 180)$	$[30; 180)$
9	$[180; +\infty)$	$(0; 0, 1)$	$[180; +\infty)$	$[0, 1; 1)$	$[30; 180)$
10	$[7; 30)$	$[7; 30)$	$[30; 180)$	$[1; 7)$	$[180; +\infty)$

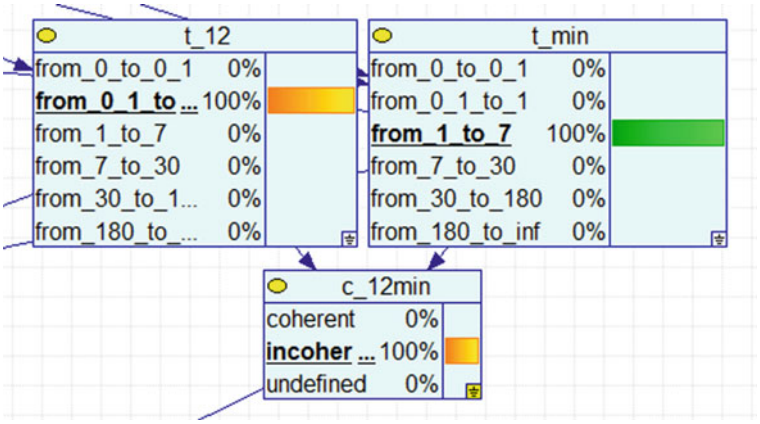


Fig. 5 Example of incoherent data (GeNIe)

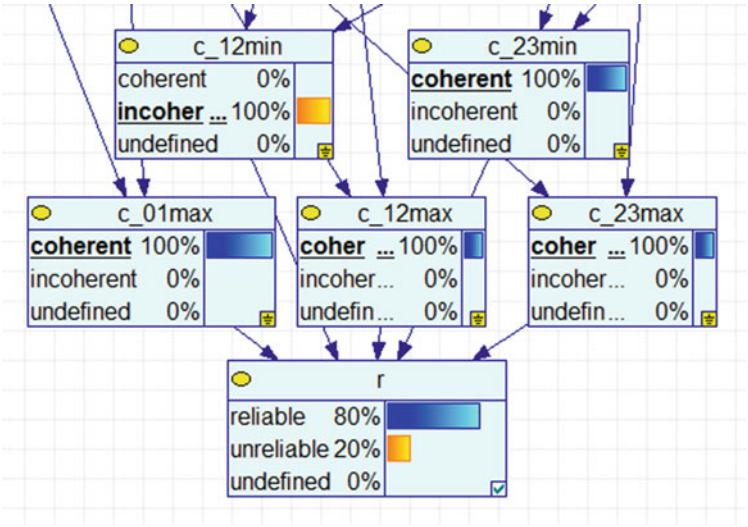


Fig. 6 Reliability estimation for the second respondent (GeNIe)

If the second respondent's data should be considered or excluded from the sample depends on the concrete research problem posed. If we want to use only the data without any contradictions or any uncertainties (all the data is coherent), then we take into account only the data from respondents 1, 4, 5, 7 and 10. Figure 7 shows the reliability estimation with the maximal degree of reliability.

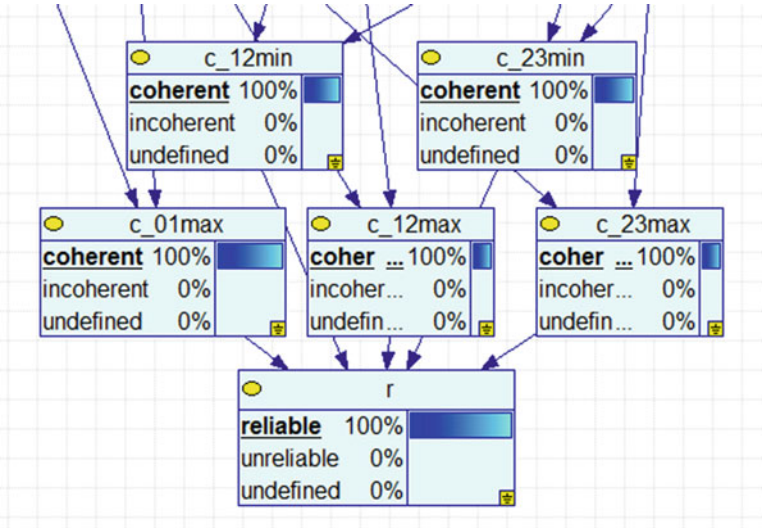


Fig. 7 Reliability estimation of respondent's coherent data (GeNIe)

6 Conclusions

We proposed a method of data coherence diagnosis of the risky behavior model with the data obtained from respondents. For more convenient use of the method software was developed and described. Example of the use of the method was also provided.

The more general cases of the coherence rate distribution (not only coherent-incoherent-undefined), different partition or unequal intervals can be considered.

This coherence diagnosis can be useful to eliminate not reliable respondents' data from the sample. Respondent reliability rate can be used as an analogue of lie scale in psychological test.

Acknowledgments This work was partially supported by the by RFBR according to the research project No. 16-31-00373.

References

1. Kabiri, P., Ghorbani, A.A.: Research on intrusion detection and response: a survey. Int. J. Netw. Secur. **1**(2), 84–102 (2005)
2. Mansour, M.M., Wahab, M.A.A., Soliman, W.M.: Bayesian networks for fault diagnosis of a large power station and its transmission lines. Electr. Power Compon. Syst. **40**(8), 845–863 (2012)

3. Xiong, G., Shi, D., Zhu, L., Duan, X.: A new approach to fault diagnosis of power systems using fuzzy reasoning spiking neural P systems. Math. Prob. Eng. **2013**, Article ID 815352, 13 pp. (2013). doi:10.1155/2013/815352
4. Zhu, Y., Huo, L., Lu, J.: Bayesian networks-based approach for power systems fault diagnosis. IEEE Trans. Power Deliv. **21**(2), 634–639 (2006)
5. Beshelev, S.D., Gurvich, F.G.: Mathematical and Statistical Methods of Expert Estimates, p. 263. Statistics, Moscow (1980)
6. Alonso, S., Chiclana, F., Herrera, F., Herrera-Viedma, E., Alcalá-Fdez, J., Porcel, C.: A consistency-based procedure to estimate missing pairwise preference values. Int. J. Intell. Syst. **23**, 155–175 (2008). doi:10.1002/int.20262
7. Tsyganok, V.V., Kadenko, S.V.: On sufficiency of the consistency level of group ordinal estimates. J. Autom. Inf. Sci. **42**(8), 42–47 (2010)
8. Muller, A., Mitchell, J., Crosby, R., Cao, L., Johnson, J., Claes, L., Zwaan, M.: Mood states preceding and following compulsive buying episodes: an ecological momentary assessment study. Psychiatry Res. **200**, 575–580 (2012)
9. Helouet, L., Marchand, H., Genest, B., Gazagnaire, T.: Diagnosis from scenarios. Discrete Event Dyn. Syst. Theory Appl. **24**(4), 353–415 (2013)
10. Suvorova, A.V., Tulupyev, A.L., Sirotkin, A.V.: Bayesian belief networks in problems of estimating the intensity of risk behavior. J. Russ. Assoc. Fuzzy Syst. Soft Comput. **9**(2), 115–129 (2014)
11. Suvorova, A.V.: Socially significant behavior modeling on the base of super-short incomplete set of observations. Inf.-Meas. Control Syst. **9**(11), 34–38 (2013)
12. Suvorova, A.V., Tulupyeva, T.V., Tulupyev, A.L., Sirotkin, A.V., Pashchenko, A.E.: Probabilistic graphical models socially significant behavior of the individual, taking into account incomplete information. Proc. SPIIRAS **3**(22), 101–112 (2012)
13. Paschenko, A.E., Tulupyev, A.L., Nikolenko, S.I.: HIV infection modeling based on the last episodes of risky behavior. Izvestiya Vuzov – Priborostroenie **11,** 33–34 (2006)
14. Fenton, K.A., Johnson, A.M., McManus, S., Erens, B.: Measuring sexual behaviour: methodological challenges in survey research. Sex. Transm. Infect. **77**, 84–92 (2001)
15. Sunner, L.E., Walls, C., Blood, E.A., Shrier, L.A.: Feasibility and utility of momentary sampling of sex events in young couples. J. Sex Res. **50**(7), 688–696 (2013)
16. Tulupyev, A.L., Nikolenko, S.I., Sirotkin, A.V.: Bayesian Networks: Logical-Probabilistic Approach, 607 pp. Nauka, St. Petersburg (2006)
17. Tulupyev, A.L., Sirotkin, A.V., Nikolenko, S.I.: Bayesian Belief Networks: Logical-probabilistic Inference in the Acyclic Directed Graph, 400 pp. University Press, St. Petersburg. (2009)
18. GeNIe&SMILE: Decisions Systems Laboratory, School of Information Sciences. University of Pittsburg. http://genie.sis.pitt.edu/

Stochastic Computer Approach Applied in the Reliability Assessment of Engineering Structures

K. Frydrýšek and L. Václavek

Abstract The development of computer technology and reliability theory allows for a qualitative improvement of the safety and serviceability assessment of structural components and systems especially in engineering. The random characters of loads, materials, geometries etc. can be considered and evaluated. Using the Simulation-Based Reliability Assessment (SBRA) Method (i.e. Monte Carlo approach based on $>10^6$ of random simulations performed by computers) the application of a fully probabilistic approach to reliability assessment of selected structural components are indicated. Hence, two practical examples are solved (i.e. buckling of a complicated statically indeterminate frame structure and bending of screw implants in bones). Other applications are mentioned. Stochastic transformation models are used in order to express the response of the structural components including the 2nd order theory effects.

Keywords Probability · Reliability assessment · 2nd order theory · Buckling · Frame structures · Beam · Elastic foundation · Medical screws · Engineering · Mechanics · Biomechanics · Simulation-Based Reliability Assessment Method

1 Introduction

Current methods for risk and reliability assessment are mostly deterministic methods (i.e. based on constant inputs and outputs). But in the real world, the typical engineering values as loads, dimensions, material properties etc. are not constant but variable. The developments of computers and reliability theories allow to applied

K. Frydrýšek (✉) · L. Václavek
Faculty of Mechanical Engineering, Department of Applied Mechanics,
VŠB–Technical University of Ostrava, 17. listopadu 15/2172,
708 33 Ostrava, Czech Republic
e-mail: karel.frydrysek@vsb.cz

L. Václavek
e-mail: leo.vaclavek@vsb.cz

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_13

stochastic/probabilistic approaches for a qualitative improvements of the safety and serviceability assessment of structural components and systems especially in the engineering. In structural reliability assessment, the concept of a limit state separating a multidimensional domain of random (stochastic, probabilistic) variables into "safe" and "unsafe" domains has been generally accepted. For more general information about the history and development of stochastic methods in the branch of engineering, see Refs. [1–5]. There are solved structures on elastic foundations, medical problems, movement of Earth plates, design of machines and its parts, experiments etc.

Hence, the SBRA Method is connected with new innovative engineering. Therefore, the probabilistic applications in the engineering are new scientific trends.

This paper presents some computational stochastic and probabilistic solutions (i.e. SBRA—Simulation-Based Reliability Assessment Method) applied in the branch of engineering. The origins and developments of SBRA Method are described in [2]. The main attention is focused on the "fully" probabilistic approach to reliability assessment using SBRA Method as a suitable tool. Let the resistance of the structure be expressed by random variable R and load effect by random variable S. The failure is defined by $RF < 0$ that is

$$RF = R - S < 0. \tag{1}$$

In SBRA Method, the reliability function (1) is analysed using direct Monte Carlo simulation and Anthill computer programme, see [2]. Let N_f be the number of simulation steps when $RF < 0$ and let N be the total number of simulation steps. Then, the probability of failure can be expressed as $P_f = P_{RF<0} = N_f/N$.

At first, the 2D unbraced steel frames are solved (i.e. buckling; mechanics; elastic transformation models are used in order to express the response of the structural components including the 2nd order theory effects; probabilistic reliability assessment). At second, the femoral screw (made of Ti6Al4V and stainless steel) for treatment of proximal femoral neck fractures of humans are solved (i.e. bending; biomechanics, implant rested in bones is solved as a planar beam on elastic foundation including the 2nd order theory effects; probabilistic reliability assessment). At third, other applications are mentioned.

2 Unbraced Planar Frame (Buckling)

The common approach to the stability (buckling) assessment of structural steel components and systems is based so far mainly on the conventional models using buckling length, buckling coefficients, compressive strength and other characteristics, see [6]. However, the computer technology allows for qualitative improvement using probabilistic approach and advanced transformation models based on 2nd order theory.

The unbraced planar steel frame shown in Fig. 1a and Table 1 consists of two cantilevered columns (*1*, *2*), two leaning columns (*3*, *4*) and three crossbars. The supports of cantilevered columns are not rigid, but elastic with stiffness k_j.

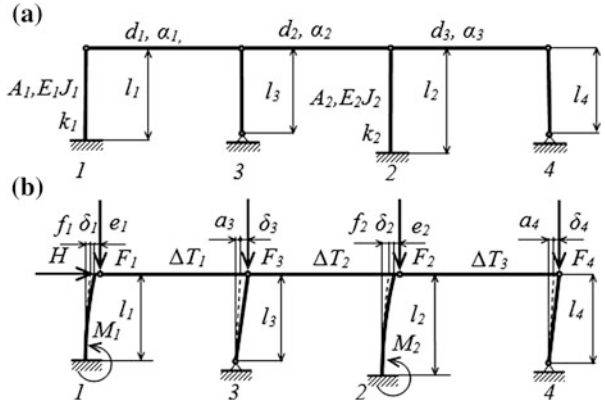


Fig. 1 Planar frame **a** undeformed situation, **b** deformed situation

Table 1 Nomenclature in Sect. 2 (Planar frame)

A_j	Area of cross-section	J_j	Quadratic moment of area
a_i	Initial imperfection	j	Index = 1 and 2
d_i	Length of cross-bar	k_j	Elastic stiffness
E_j	Modulus of elasticity	l_i, l_j	Length of column
e_j	Eccentricity of vertical force	M_j	Reaction bending moment
EQ	Earthquake load	α_i	Thermal coefficient of expansion
F_i	Vertical force	δ_j	Horizontal displacement
f_j	Amplitude of initial crookedness	θ_j	Angle of relative rotation
H	Horizontal force	ΔT_i	Temperature difference
i	Index = 1, 2, 3 and 4	W	Wind load

The frame is loaded by forces F_i, H, and by forced deformation caused by temperature differences ΔT_i (with reference to the temperature during erection), see Fig. 1b. Initial curvatures with amplitudes f_j, were considered. Unavoidable eccentricities of forces F_i are e_i. Imperfections a_i represent the initial deviations.

Analytical transformation model of the frame is expressed by following equations

$$\delta_1 = \frac{H + \sum_{i=3}^{4} \frac{F_i a_i}{l_i} + \sum_{i=1}^{2} \frac{F_i V_i}{G_i l_i} + \frac{F_3}{l_3} d_1 \alpha_1 \Delta T_1 - \frac{F_2}{G_2 l_2} \sum_{i=1}^{2} d_i \alpha_i \Delta T_i + \frac{F_4}{l_4} \sum_{i=1}^{3} d_i \alpha_i \Delta T_i}{\frac{F_1}{G_1 l_1} + \frac{F_2}{G_2 l_2} - \sum_{i=3}^{4} \frac{F_i}{l_i}}; \omega_j = \sqrt{\frac{F_j}{E_j J_j}};$$

$$\delta_2 = \delta_1 + \sum_{i=1}^{2} d_i \alpha_i \Delta T_i; M_j = F_j[(1 + \frac{1}{G_j})\delta_j - \frac{V_j}{G_j} + e_j + f_j]; P_j = 1 - \frac{F_j l_j}{k_j} \frac{\tan(\omega_j l_j)}{\omega_j l_j};$$

$$G_j = \frac{\tan(\omega_j l_j)}{P_j \omega_j l_j} - 1; V_j = e_j[\frac{1}{P_j \cos(\omega_j l_j)} - 1] + f_j \left\{ \left(P_j[1 - (2\omega_j l_j/\pi)^2] \right)^{-1} - 1 \right\}. \quad (2)$$

Table 2 Stochastic input values in Sect. 2 Planar frame

Input random variables		Extreme value or range	Histograms
F_1	Dead load	200 kN	Dead1
	Long lasting load	50 kN	Long2
	Short lasting load	50 kN	Short2
F_2	Dead load	400 kN	Dead1
	Long lasting load	150 kN	Long2
	Short lasting load	150 kN	Short2
F_3	Dead load	200 kN	Dead1
	Long lasting load	100 kN	Long2
	Short lasting load	200 kN	Short2
F_4	Dead load	100 kN	Dead1
	Long lasting load	100 kN	Long2
	Short lasting load	100 kN	Short2
W	Wind	±40 kN	Wind1
EQ	Earthquake	$0.02\Sigma F_i$, or $0.03\Sigma F_i$	Gumbel02
ΔT	Temperature difference	From −21 to +40 °C	Gamma08
Input random variables		Mean value or range	Histograms
f_1	Amplitude of initial curvature, columns *1*, *2*	±20 mm	Normal2
f_2		±25 mm	Normal2
e_1	Eccentricities of forces F_1, F_2	±30 mm	Normal2
e_2		±38 mm	Normal2
a_3	Equivalent geometrical imperfections, columns *3*, *4*	±27 mm	Normal2
a_4		±30 mm	Normal2
A_1	Area of cross section, columns *1*, *2*	13.1×10^2 mm^2	N1-04
A_2		17.1×10^2 mm^2	N1-04
S_1	Section modulus, columns *1*, *2*	138×10^4 mm^3	N1-08
S_2		216×10^4 mm^3	N1-08
I_1	Moment of inertia, columns *1*, *2*	193×10^6 mm^4	N1-08
I_2		367×10^6 mm^4	N1-08
E_1, E_2	Modulus of elasticity, columns *1*, *2*	210 GPa	N1-15
F_{y1}, F_{y2}	Yield stress, columns *1*, *2*	248–500 MPa	A36-m
k_1	Elastic stiffness, flexible support, columns *1*, *2*	8×10^7 Nm/rad	N1-30
k_2		1.5×10^8 Nm/rad	N1-30

Input random variables are summarized in Table 2. Names of histograms are the same as in [2] (i.e. Anthill software). Except EQ, all input random variables are mutually uncorrelated. Force EQ is correlated with forces F_i. Force H is the sum W and EQ, both of them may act to the left or to the right. The coefficient of temperature expansion is $\alpha_i = 12 \times 10^{-6}$ K^{-1}. Dimensions $l_1 = 6$ m, $l_2 = 7.6$ m, $l_3 = 5.4$ m, $l_4 = 6$ m and $d_1 = d_2 = d_3 = 10$ m. Numerical results are summarized

Table 3 Calculated probabilities of failure P_f—safety (carrying capacity, Planar frame)

Failure probability P_f—refered to the onset of yielding			
$EQ = 0.02\Sigma F_i$*Gumbel02		$EQ = 0.03\Sigma F_i$*Gumbel02	
Column *1*	Column *2*	Column *1*	Column *2*
$<10^{-6}$	$<10^{-6}$	23.5×10^{-6}	1×10^{-6}

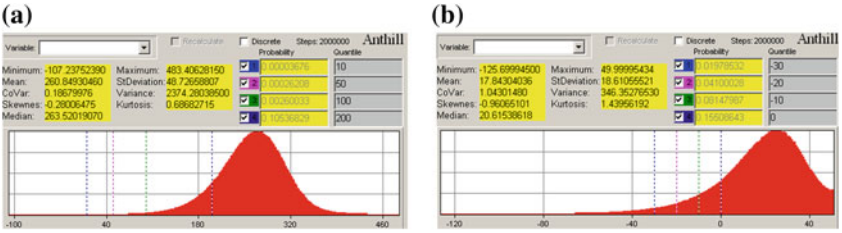


Fig. 2 Histograms of the safety function **a** Carrying capacity, **b** serviceability—Planar frame, Anthill software

in Table 3 and [6]. Two million simulation steps were used for both of the safety and serviceability assessment.

Load carrying capacity is evaluated with reference to the onset of yielding (most exposed fibers of columns *1*, *2* are considered, see Fig. 2). Serviceability assessment of the structure refers to the lateral displacement limit value of the upper end of column *1*. Computed probabilities of failure (probability of exceeding the lateral displacement value, see Fig. 2) are summarized in [6].

To assess the safety and serviceability of this unbraced planar frame using current codes (i.e. different methods LRFD or ASD approach, see [7]) and considering its complexity would not be an easy task, see, e.g. the discussion on the assessment of such system in [8]. However, in this article, the authors offer better stochastic/probabilistic solution which is based on own analytical model applied in Anthill software (SBRA Method).

3 Probabilistic Reliability Assessment of Femoral Screws Intended for Treatment of "Collum Femoris" Fractures (Bending)

Proximal femoral neck fractures, see Fig. 3a, remain a vexing clinical problem in traumatology and orthopaedics (i.e. common type of trauma), see [9–11].

One possible treatment method for femoral neck fractures, is the application of femoral screws made up from stainless steel or Ti6Al4V material, see Fig. 3b, c.

The analytical model for strength analyses of femoral screws is based on the theory of beams on an elastic foundation, where the bone is approximated by the elastic foundation prescribed by stiffness k/Pa/, see [1, 4, 10]. Quite large and

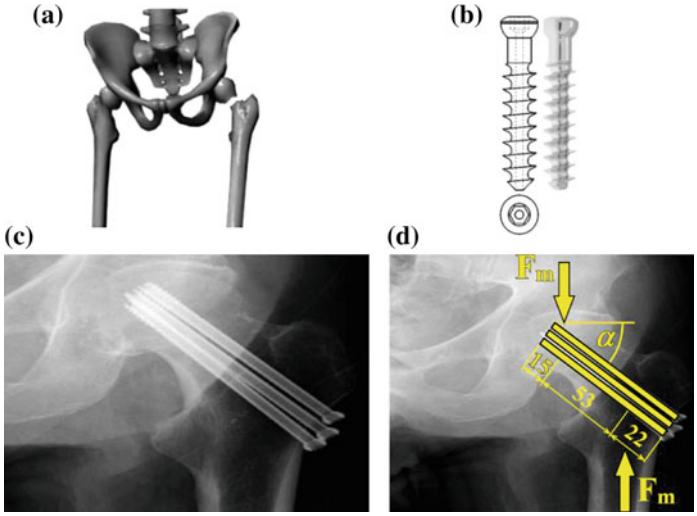


Fig. 3 **a** Femoral neck fracture. **b** Femoral screws. **c** Treatment of fracture—X-ray snapshot. **d** Beams on elastic foundation

complex review about the theory and practice of elastic foundation is performed in author's book [1].

Three screws are applied in parallel positions on the elastic foundation (i.e. in the bone). The force F/N/ acting in one screw can be defined via total loading force F_m/N/, see Fig. 3d, by the equation $F = F_m/n = m\ k_m k_{dyn} g/n$. The variables are as follows: m/kg/ is the entire mass of a patient; k_m/1/ is the coefficient of mass reduction; k_{dyn}/1/ is the dynamic force coefficient; g/ms^{-2}/ is the gravitational acceleration; and n/1/ is the coefficient of inequality in the division of force F_m into three screws. These variables are defined via truncated histograms. The force F can be decomposed into forces $F_1 = F \cos \propto$ and $F_2 = F \sin \propto$, see Fig. 4a. The femoral screw angle α/deg/, which is defined by the limits of adduction and abduction, see Fig. 4b, and the yield limit Re/MPa/ for material, are likewise defined via truncated histograms, see Fig. 5a. According to the 2nd order theory and the theory of beams on an elastic foundation, three linear differential equations for the intervals $x_{1,2,3}$, can be written as $EJ_{ZT}\frac{d^4 v_i}{dx_i^4} + F_2\frac{d^2 v_i}{dx_i^2} + kv_i = 0$ together with 12 boundary conditions. EJ_{ZT}/Nm^2/ is flexural stiffness, v_i/m/ is displacement and x_i/m/ are coordinates. Hence, bending moments M_{oi}/Nm/, and maximum stresses σ_{MAX}/MPa/, see e.g. Fig. 4a, can be calculated.

Probabilistic reliability assessment can be carried out via SBRA Method by means of the reliability function (1), depending on load capacity, compared to the extreme stress values with yield limit. The probability that plastic deformation will occur in the beam is $P_f = 2.51 \times 10^{-5}$ i.e. $P_{f\%} = 0.00251\,\%$ (calculated for 5×10^6 Monte Carlo simulations, see Fig. 5b). Hence, the femoral screws are safe and suitable for patient treatment and the surgeons can use them for treatment.

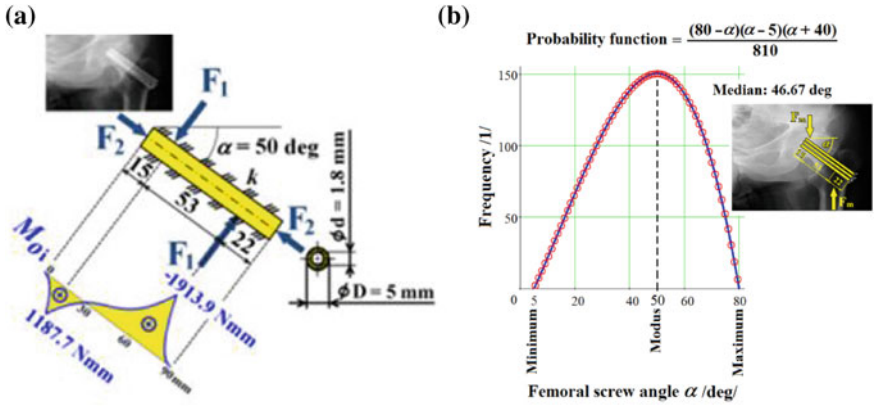
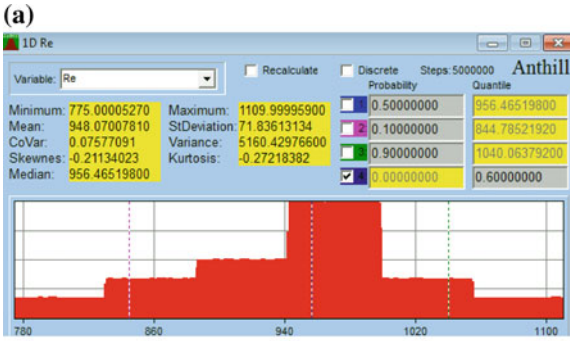
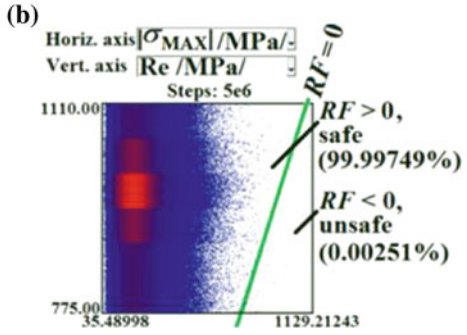


Fig. 4 **a** Beam on elastic foundation. **b** Histogram of femoral angle

Fig. 5 **a** Histogram of yield limit for Ti6Al4V material (Anthill software); **b** Probabilistic assessment of femoral screw (i.e. result of SBRA Method, Anthill software)





4 Conclusion

The transition from the pre-computer era reliability assessment to modern computers technology leads to re-engineering of the entire assessment procedures. The probabilistic SBRA Method is connected directly with this transition. Methods such as SBRA can be considered and applied in order to evaluate the safety and serviceability of structural components and systems taking into the account variabilities of inputs. One of the main prerequisites, related to the application of qualitatively higher reliability assessment methods using the potential of modern computers, is the transition of thinking of designers from deterministic to probabilistic. In case of the stability (buckling) problems (i.e. unbraced planar frame), the current assessment criteria based on buckling length and compressive strength, buckling coefficients, etc. can be gradually replaced by a fully probabilistic approach, using transformation models and 2nd order theory. In case of the design of femoral screws intended for treatment of "collum femoris" fractures in traumatology/orthopaedics (i.e. bending of planar beam rested in femur), the probabilistic assessment criteria can be based on fully probabilistic approach, using transformation models and 2nd order theory. Other examples are mentioned too.

Acknowledgments This work was supported by the Czech projects TA03010804 and SP2016/145.

References

1. Frydrýšek, K., Tvrdá, K., Jančo, R., et al.: Handbook of Structures on Elastic Foundation, pp. 1–1691. VŠB - Technical University of Ostrava, Ostrava, Czech Republic (2013). ISBN: 978-80-248-3238-8
2. Marek, P., Brozzetti, J., Guštar, M., Tikalsky, P., et al.: Probabilistic Assessment of Structures Using Monte Carlo Simulation Background, Exercises and Software, 2nd extended edn. ITAM CAS, Prague, Czech Republic (2003). ISBN: 80-86246-19-1
3. Kala, Z.: Sensitivity analysis of steel plane frames with initial imperfections. Eng. Struct. **33**(8), 2342–2349 (2011)
4. Frydrýšek, K., Čada, R. Probabilistic reliability assessment of femoral screws intended for treatment of "Collum Femoris" fractures. In: BIOMECHANICS 2014—International Conference of the Polish Society of Biomechanics, pp. 61–62. Łódź (2014). ISBN: 978-83-7283-628-1
5. Lokaj, A., Vavrušová, K., Rykalová, E.: Application of laboratory tests results of dowel joints in cement-splinter boards VELOX into the fully probabilistic methods (SBRA method). Appl. Mech. Mater. **137**, 95–99 (2012). doi:10.4028/www.scientific.net/AMM.137.95. ISSN: 1660-9336
6. Václavek, L., Marek, P., Konečný, P.: SBRA approach to the stability assessment of structural components and systems. In: Proceedings of McMat2005—2005 Joint ASME/ASCE/SES Conference on Mechanics and Materials, Article no. 519, pp. 1–6. Baton Rouge, Louisiana, USA (2005)
7. Specification for Structural Steel Buildings: ANSI/AISC 360-05 Standard. AISC Inc., Chicago (2005)

8. Geschwindner, L.F.: A practical approach to the "Leaning" column. Eng. J. **31**, 141–149 (1994)
9. Frydrýšek, K., Jořenek, J., Učeň, O., Kubín, T., Žilka, L., Pleva, L.: Design of external fixators used in traumatology and orthopaedics—treatment of fractures of pelvis and its acetabulum. Procedia Eng. **48**, 164–173 (2012)
10. Frydrýšek, K.: Strength analyses of full and cannulated femoral screws made up from stainless steel and Ti6Al4V, Calculation report, FME VŠB-Technical University of Ostrava, Ostrava, Czech Republic, pp. 1–43 (2014)
11. Farhad, N., Bradley, E.J., Hodgson, S.: Comparison of two tools for the measurement of interfragmentary movement in femoral neck fractures stabilised by cannulated screws. Robot. Comput.-Integr. Manuf. **26**(6), 610–615 (2010)

Part III
Image Processing and Emotion Modeling

Context-Sensitive Image Analysis for Coloring Nature Images

Aleksey A. Alekseev, Vladimir L. Rozaliev, Yulia A. Orlova and Alla V. Zaboleeva-Zotova

Abstract In this article we introduce the method for context-sensitive image analysis for coloring nature images. The method based on Welsh method for coloring grayscale images. We improved Welsh method for nature through changing sample size. In the conclusions we present results of our method and potential improvements.

Keywords Image processing · Coloring images · Context-sensitive image analysis · Signature of image

1 Introduction

Currently, there are a large number of different grayscale images, for example, modern stylized black and white photos, old black and white movies and photographs, most photos that are taken at night time, which are also devoid of color, and grayscale images that are made to save disk space. Often there is a need to restore image colors, which are absent due to the reasons described above.

At present the problem of coloring of grayscale images is not completely solved because of various difficulties. Firstly, when color is removed, information is lost that cannot be accurately recovered. Secondly, it is necessary to understand what is depicted in the image, i.e., the problem of recognition of all objects should be

A.A. Alekseev · V.L. Rozaliev (✉) · Y.A. Orlova
Volgograd State Technical University, Volgograd, Russia
e-mail: vladimir.rozaliev@gmail.com

A.A. Alekseev
e-mail: alekseev.yeskela@gmail.com

Y.A. Orlova
e-mail: yulia.orlova@gmail.com

A.V. Zaboleeva-Zotova
Russian Foundation for Basic Research, Moscow, Russia
e-mail: zabzot@gmail.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_14

solved. There are different approaches to the coloring of grayscale images. The qualitative approach, i.e., manual coloring (using graphical editors), is the most widespread one. However, this method has two major drawbacks: great money and labor costs. For example, it took 3 years to color the movie (a sequence of frames that are images) Seventeen Moments of Spring. The cost of 1 min of colorization amounted to 3000 US dollars.

In this article we localized the area of interest. We used our method only for coloring nature grayscale images.

2 Analysis of Existing Solutions for the Coloring Nature Images

Currently, there are various implementations of this method such as coloring based on the luminance histogram regression, image type selection, coloring based on the selected color image, etc.

2.1 ShiGuang Liu et al. Approach

Grayscale images are automatically colored using the method based on the regression histogram. The color information is got from color image. Locally weighted regression is determined on both grayscale and color images. The distribution function of two images can be obtained.

In this method, it is proposed to reduce these functions by finding and adjusting the zero point of the histogram. When a color corresponding to the luminance is found, the grayscale image is colored. In addition, coloring results are verified [5].

2.2 Yogesh Rathore et al. Approach

This approach use the Welsh color transfer method for coloring. The color source image is selected from the database by the comparison of the images signatures. The color source image is converted to the decorrelated color space YCrCb. Then the luminance histogram is obtained and the image signature is calculated on its basis. The image signature and two color channels (a and b) are stored in the database. If it is necessary to color the target image, its signature is calculated similarly, the best match from the database is found, and the color channel values of the source color image are transferred in accordance with the luminance values to the target image [7].

2.3 Luiz Filipe Menezes Vieira et al. Approach

The Welsh color transfer method is used for coloring [10]. The color source image is selected from the image database by the comparison of image signatures.

Unlike the previous system, this one provides four different types of signatures designed for different types of images (one object in the background, object filled image, etc.). There is also a color space lab [9].

2.4 The Recolored Software

In the Recolored software user should select an area that should be colored and then to specify the color. The software will do everything else including the hue distribution depending on selected object shadows (these also depend on the object's shape).

Recolored makes it also possible to process color images. It is possible to change the color or highlight any part of a grayscale image.

The software contains a library of colors that are grouped in categories such as skin tone, hair color, metal, wood, grass, and cloth.

The software supports the file formats JPEG, PNG, BMP, and RCL as well as RGB and HSL color models [3].

2.5 AKVIS Coloriage Software

AKVIS Coloriage is software for coloring grayscale photos and replacing colors on color images.

The main problem is the separation of areas and the selection of colors. It is necessary to specify areas for recoloring and run the computation. Coloriage recognizes the borders and colors an image by superimposing colors taking textures, luminance, and shadows into account.

The software includes a library of colors to select shades of skin, hair, foliage, sky, and other natural colors. It is possible to save and load different strokes.

The additional Recolor Brush tool makes it possible to adjust the result of automatic coloring. The brush can also be used as a standalone tool for manual coloring.

Immediately after its release in 2005, the plugin version of this software was awarded Best of 2005 Soft of the PC Magazine/RE in the category Graphics and Photos.

AKVIS Coloriage is implemented in the form of a plugin for image processing software and as standalone software.

AKVIS Coloriage supports images in RGB (8/16 bits). The software supports formats JPEG, TIFF, BMP, and PNG. Images in CMYK, Grayscale, Lab, etc. are automatically converted to RGB when loaded to AKVIS Coloriage [8].

2.6 A.D. Varlamov et al. Approach

The coloring is based on the fact that in color images with a uniform content (e.g., sunset, asphalt roadway, etc.) objects of the same type have about the same color. For example, in summer, the forest is covered in green leaves, which turn red and yellow during autumn; in winter, trees are covered with white snow; clouds are white in the blue sky, while storm clouds are gray; etc. A large amount of data obtained as a result of image analysis is used to evaluate the color of each point of the original grayscale image.

The image is colored according to the given type of scene in three steps:

1. Training samples are generated (similar color images are selected).
2. Machine learning is carried out.
3. Similar grayscale images are colored [4].

2.7 Comparison

In the Table 1 is shown comparison of existing approaches. The main problem of existing solution is next: all of this work with image like a set of pixels without understanding what exactly is presented on image.

Table 1 Comparison of exciting approaches

Approach	User intervention	Additional input data
ShiGuang Liu et al.	Input polynoms	Color image
Yogesh Rathore et al.	No	Color image
Luiz Filipe Menezes Vieira et al. approach	No	Color image (or image from DB)
The Recolored Software	Areas with the same colors	No
AKVIS Coloriage Software	Areas with the same colors	No
A.D. Varlamov et al. approach	Type of image	Color image (or image from DB)
Our approach	No	Color image (or image from DB)

3 Coloring Nature Images

3.1 Method for Coloring Images

Coloring grayscale image involves replacing the scalar value of each pixel on the vector (e.g., red, green and blue). T. Welsh introduces the technique for "colorizing" grayscale images by transferring color between a source, color image and a destination, grayscale image.

The Welsh's method based on comparing small pixel neighborhoods on the luminance channels color and grayscale images. In small pixel neighborhood calculate the average and variance. Our modification is to change the size of neighborhood from 5 × 5 to 25 × 25. The color transfer process is presented on Fig. 2.

Figure 1 demonstrates a changing SSIM (structural similarity) index and time of coloring depending on the size of pixel neighborhood. Method of verification based on structural similarity (SSIM) index [7]. The SSIM index is a full reference metric; in other words, the measuring of image quality based on an initial uncompressed or distortion-free image as reference, therefore we must do the following: for test color image get his grayscale variant and colorize it, then we get SSIM for source color image and the result of colorize and multiply by 100 %.

3.2 Signature of Image

The image is colored based on the color image from the images base [2]. Color image is got from base based on comparing their signatures.

The image signature is 128 floating point numbers [1], which are derived from the luminance histogram [6], which is normalized to 1 (Fig. 3).

Two signatures are compared on the basic of correlation (1) between them.

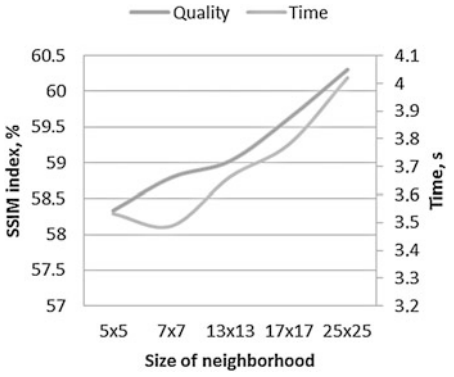


Fig. 1 The dependence the time of coloring and SSIM index of the pixel size of the neighborhood

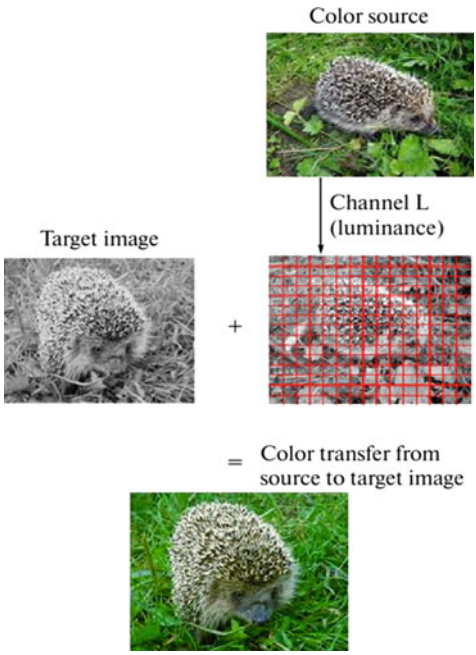


Fig. 2 Color transfer from source color image to target grayscale image

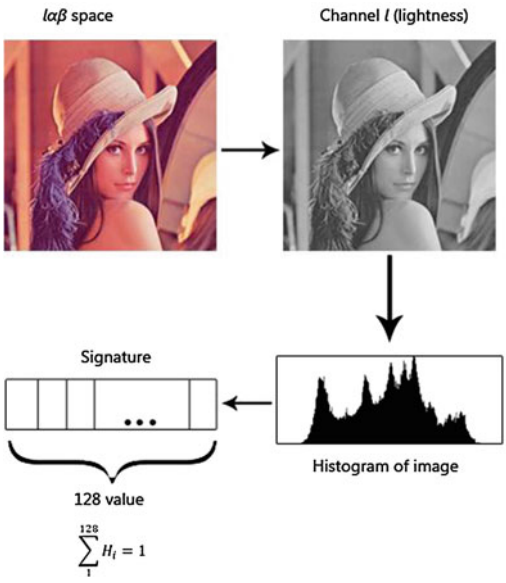
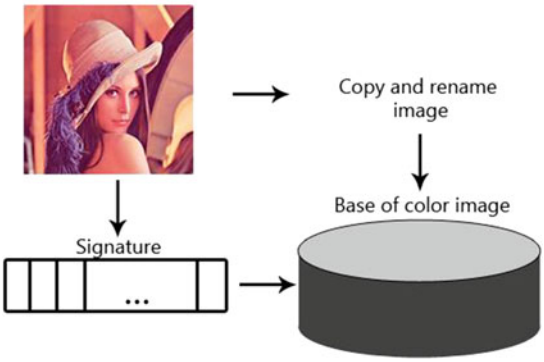


Fig. 3 Signature of image

Fig. 4 Base of color image



$$d_{correl}(H_{1,}H_2) = \frac{\sum_i^N H_1'(i) * H_2'(i)}{\sqrt{\sum_i^N H_1'^2(i) \sum_i^N H_1'^2(i)}} \quad (1)$$

where H_1 and H_2—signatures, N equals number of elements signature (N = 128) and $H_k'(i)$ equals (2):

$$H_k'(i) = H_k(i) - \frac{1}{N}(\sum\nolimits_j^N H_k(j)) \quad (2)$$

3.3 *Base of Color Images*

For coloring grayscale image we need color source image. The base contains color images and pairs of signature and name of the image. Name of image is unique. These pairs provide a fast search required image according its signature. User can add and delete image. Figure 4 illustrate adding image to base.

4 Conclusions

In this paper, computer image processing via coloring of grayscale images was considered. It was shown that coloring of grayscale images is an important unsolved problem. Practice results are considered on Fig. 5.

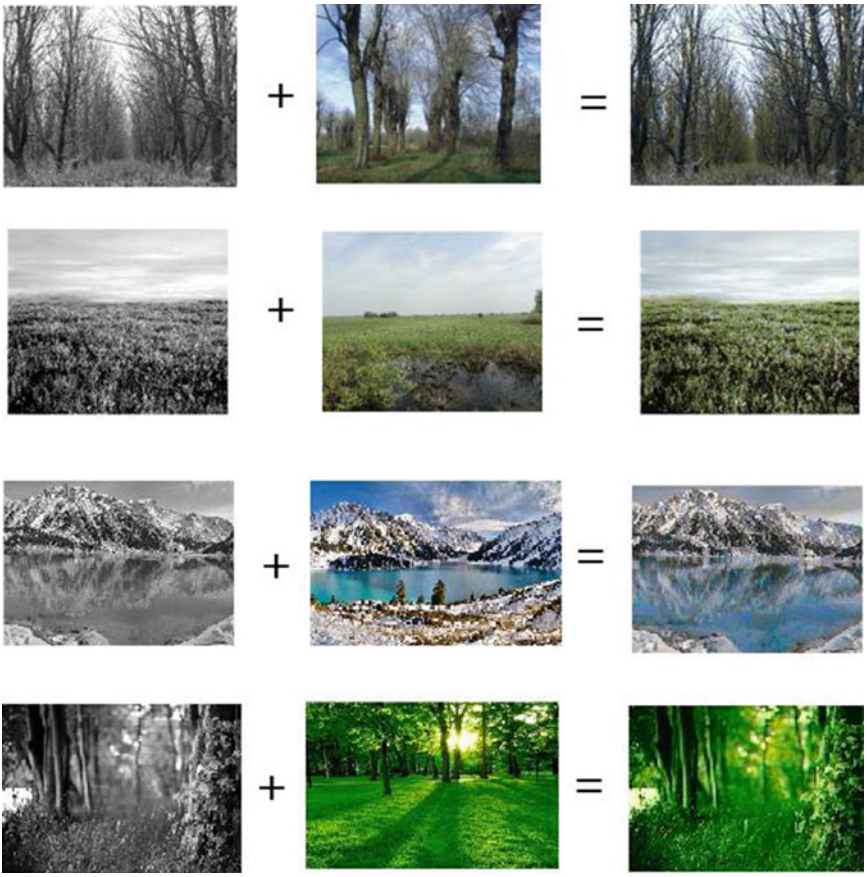


Fig. 5 Results

The following aspects were described:

1. Existing approaches for the coloring of grayscale images were considered. Their advantages and disadvantages were revealed.
2. How to get color source image based on signatures
3. The Welsh method of color transfer to a grayscale image was modified.

Acknowledgments The study was financially supported by RFBR research projects №15-47-02149, 15-07-06322, 15-37-70014, 16-07-00407, 16-07-00453.

References

1. Alekseev, A.: Automatic coloring of grayscale images based on intelligent scene analysis. In: Alekseev, A., Rozaliev, V, Orlova, Y. (eds.) Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications, vol. 25, no. 1, pp. 10–21 (2015)
2. Alekseev, A., Rozaliev, V, Orlova, Y., Zaboleeva-Zotova, A.: Two-stage segmentation method for context-sensitive image analysis. In: Kravets, A., Shcherbakov, M., Kultsova, M. (eds.) Knowledge-Based Software Engineering, JCKBSE 2014. Volgograd, Russia, 17–20 Sept 2014. Communications in Computer and Information Science, vol. 466, pp. 331–340. Volgograd State Technical University [etc.], Tadashi Iijima. Springer International Publishing (2014)
3. Colorize Black and White Photos using Recolored. http://www.recolored.com/
4. Kulibin Software for Coloring Half Tone Images and Black and White Video Films (2011). http://kulibin.org/projects/show/1812
5. Liu, S., Zhang, X.: Automatic grayscale image colorization using histogram regression. Pattern Recogn. Lett. **33**(13), 1673–1681 (2012)
6. Petrov, M.: Computer Graphics: Student's Book for High School. Piter, St. Petersburg (2011) (in Russian)
7. Rathore, Y., et al.: Colorization of Gray Scale Images Using Fully Automated Approach, vol. 7109, pp. 16–19 (2010)
8. The Way to Color BlackandWhite Photos and Color Replacement at Colored Ones by Means of AKVIS Coloriage. http://akvis.com/ru/coloriage/index.php (2010)
9. Vieira, L.F.M., et al.: Fully automatic coloring of grayscale images. Image Vision Comput. **25** (1), 50–60 (2007)
10. Welsh, T., Ashikhmin, M., Mueller, K.: Transferring color to greyscale images. ACM Trans. Graph. **21**(3), 277–280 (2002)

Development of 3D Human Body Model

Vasiliy M. Konstantinov, Vladimir L. Rozaliev, Yulia A. Orlova and Alla V. Zaboleeva-Zotova

Abstract This article describes a software system for scanning and three-dimensional reconstruction of the human body using the two MS Kinect devices. The article describes the main stages of the conversion of partial frames of depth in the general polygon mesh model of the human body. Also, we propose and describe a method for constructing the surface in the "empty zones". We show that using two Kinect devices our scanning system can reconstruct the human body and build the surface in "empty zones". The main idea of our reconstructing approach is to divide the whole surface of human body in two parts: the front part and the back part. Each of that two parts is dividing in three parts with overlaps: the upper part (head, part of the chest, arms), middle part and lower part. Scans of each part aligned to each other with ICP algorithm and stored in two separate point clouds, which represents the front and back part of human body. Finally, two scans reconstructed and "empty zones" between them build with our algorithm based on Bezier curve model. In conclusion represent the final model of human bode with some parameters.

Keywords 3D reconstruction · Human body model · Kinect · ICP · Bilateral filtering · Bezier curve

V.M. Konstantinov · V.L. Rozaliev (✉) · Y.A. Orlova
Volgograd State Technical University, Volgograd, Russia
e-mail: vladimir.rozaliev@gmail.com

V.M. Konstantinov
e-mail: konstantinovr1@gmail.com

Y.A. Orlova
e-mail: yulia.orlova@gmail.com

A.V. Zaboleeva-Zotova
Russian Foundation for Basic Research, Moscow, Russia
e-mail: zabzot@gmail.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_15

1 Introduction

Three-dimensional models of real world objects, in particular models of a human body are applied in many areas, such as applications of the added and virtual reality, video games, in the reengineering process. Such models are created by expensive specialized hardware and software applications.

In this work we offered and described the method of receiving three-dimensional model of a human body with use two MS Kinect devices without additional equipment. The Kinect device is much cheaper than specialized scanners and provides depth stream which can be transformed to a set of three-dimensional points, and then to a polygonal grid of human body.

Now there are a large number of reconstruction systems of complex objects using the Kinect device. However, the reconstruction of the human body imposes a number of additional restrictions on the configuration of the scanning system. So to provide acceptable details level of the surface must be obtained separately the original surface of the scanned body [1, 2]. Some systems use a rotating mechanism to rotate the scanned person. In other multiple devices, which scanning certain areas [3]. Third, these approaches are combined [4].

Another problem is that not all parts of the body surface into the view of the camera. These blank areas are called "empty zones" [5, 6]—areas where the surface is not determined. This problem might be solved in different ways, but most often just an empty area being completed manually in the package of 3D graphics [7].

2 Description of the System Workflow

As entrance data the system accepts two streams from two depth cameras of a Kinect sensor. The system processes each scan (R'_k) in streams which represents the image with the permission 640 × 480 pixels. Each pixel describes distance to a site of the scanned surface. This distance is expressed in millimeters.

Before processing of depth it is necessary to carry out their filtration. It is a problem it is connected with that these depths returned by the device rather strongly noise. These noise have casual additive character statistically independent of a signal. In this case it is necessary to remove noise from shots of depth and in too time to keep the clearness of borders. It is caused by that borders on a shot of depth express transition between surface levels. Smoothing of borders is inadmissible and will lead to loss of information. Use of the bilateral filter allows to smooth the image with preservation of a clear boundary [8].

Further the filtered depth shots (R_k, R_{k-1}) are processed for receiving a set of three-dimensional points (V_k) and normals (N_k) to them. As, resolution of the camera of depth isn't great it makes sense to scan a surface of a body of the person in parts. For combination of partial frames of a surface the variation of algorithm of

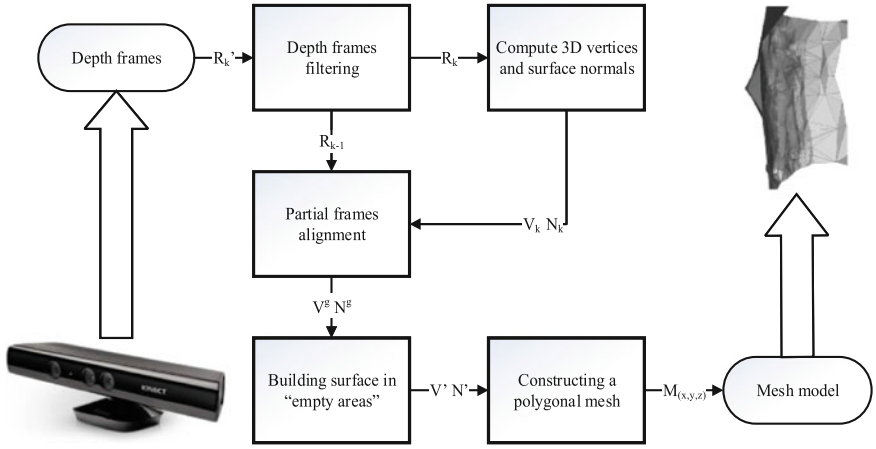


Fig. 1 Overall system workflow

ICP is used. At the exit the full frame of a surface (V^g, N^g) forward and back part of a body of the person turns out.

Further in view of the fact that there are surface sites which can't come into the view of any of depth cameras ("empty zones") it is necessary to complete a surface in these zones. The algorithm of creation of tops on the basis of square Bezier curves is for this purpose developed. At the exit of this stage the general surface (V', N') a human body united by the added tops between forward and a back surface will be received. After calculation of the general surface it is necessary to receive polygonal model $(M_{x,y,z})$. Is for this purpose used the special algorithm of reconstruction based on the solution of the equation of Poisson (Poisson Surface Reconstruction) [9].

As the output data the system provides the *.ply file which contains polygonal model of a surface of a body of the person. In Fig. 1 the scheme of the general functioning of system is submitted.

3 Description of a Creation Human Body Surface Model Method

Before start of system it is necessary to place MS Kinect devices and to allocate space for scanning of a scene. Two devices are located before the person and behind him, and sent to the opposite sides. The distance between two devices depends on growth of the scanned person. Height at which devices are located is equal to a half of growth of the scanned person.

During process of scanning of people has to be motionless precisely in the center (i.e. on identical removal from both devices) the virtual cube limited to the planes of

cameras of MS Kinect devices. Generally, the user can accept any motionless pose. However, thus no part of a lobby or back surface should block the review of the camera.

The viewing angle of the MS Kinect chamber is limited and makes 43° down. It means that the distance of capture of a surface to the utmost of the scanned person is limited. However if conditionally to divide a surface into 3 parts (top, average and lower) with rather big consecutive overlappings the distance can be reduced. It will increase the accuracy of measurement of distance to surface sites as now each partial shot will have permission 640 × 480 pixels, but not a frame of all surface.

After installation and control of system thus the preparatory stage is finished. The scheme of physical placement of system is submitted in Fig. 2.

Before processing of shots of a stream of depth it is necessary to carry out their filtration. In this work for a filtration the bilateral filter [10] is used. As the intensity of the color pixels is the distance to the surface. Thus, the formula of new value of depth (R_k) in concrete pixel (q) is expressed by a formula, where R'_k—value of distance in a concrete point of a shot of depth, u—the central point of level of a surface, $\mathcal{N}_\sigma(t) = \exp(-t^2\sigma^{-2})$—Gauss function, W_p—some weight coefficient.

$$\mathrm{R_k(q)} = \frac{1}{\mathrm{W_p}} \sum\nolimits_{\mathrm{q}\in\mathcal{U}} \mathcal{N}_{\sigma_\mathrm{s}}(||\mathrm{u}-\mathrm{q}||_2)\mathcal{N}_{\sigma_\mathrm{r}}(||\mathrm{R'_k(u)}-\mathrm{R'_k(q)}||_2)\mathrm{R'_k(q)} \qquad (1)$$

After processing the bilateral filter will receive a smoothed shot of depth. For transformation of a frame of depth to a cloud of three-dimensional points it is necessary to calculate the first step a matrix of a perspective projection of the

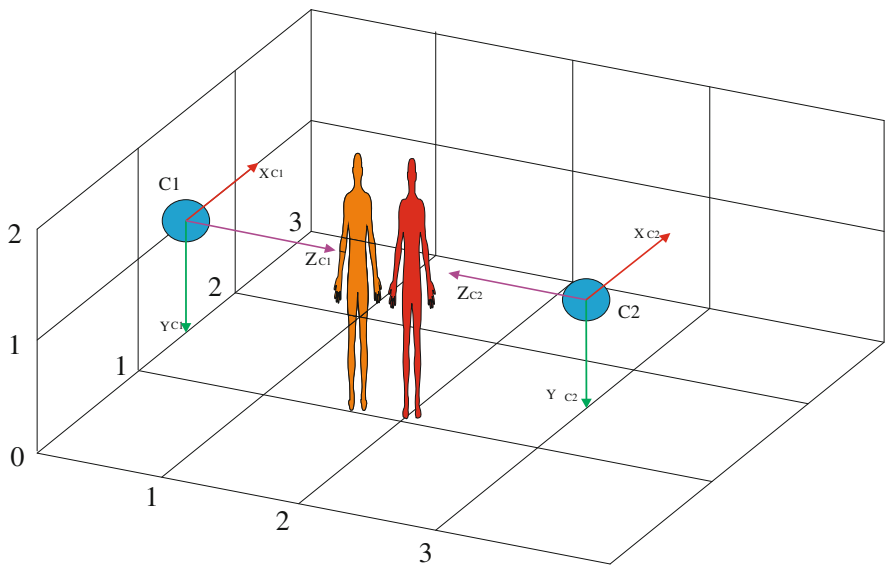


Fig. 2 Illustration of a system setup

camera (K). It can be received on the basis of an internal calibration matrix of the MS Kinect chamber.

Each pixel of a frame of depth represents a vector, where x—the coordinate of "pixel" on width of a shot, y—coordinate on height of a shot, z—value of depth.

$$Q = [x \quad y \quad z]^T \tag{2}$$

For receiving a set of three-dimensional tops (V_k) it is necessary to increase a vector (Q) by the return internal matrix for each pixel of a shot of depth.

$$V_k(q) = D_k(q)K^{-1} \tag{3}$$

The following step—calculation of a normal for each top. The vector of a normal (N_k) is calculated on the basis of coordinates of the next tops.

$$N_k(q) = (V_k(x+1, y) - V_k(x, y)) \times (V_k(x, y+1) - V_k(x, y)) \tag{4}$$

Forward and back surfaces have to be constructed by combination of partial frames. The top, average and lower frames have mutual overlappings therefore for their combination we will use algorithm of ICP (Iterative Closest Point) [11]. The top and average frame, then average and lower frames are consistently combined. On the last step the turned-out frames in a uniform surface are combined (forward and back). The scheme of overlappings in partial frames is submitted in Fig. 3.

The first step of algorithm consists in finding of coinciding tops in two partial frames. The current frame (R_k) and the previous frame are compared (R_{k-1}). If the matrix of transformation of the camera from the previous step (the previous partial

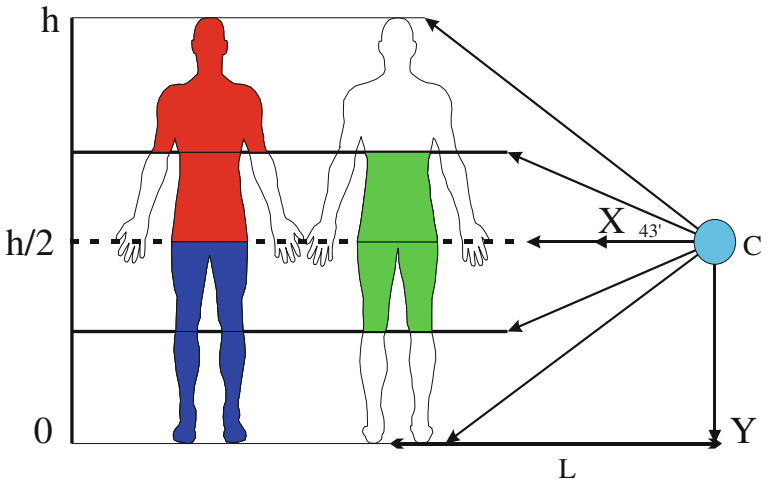


Fig. 3 Illustration of a system setup

shot of depth) is known it is possible to project tops in global coordinates (V_{i-1}). As the Kinect chamber is fixed in one point and has only one degree of freedom (rotation round axis X) the matrix of transformation has an appearance

$$T_{i-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_{i-1} & -\sin\varphi_{i-1} \\ 0 & \sin\varphi_{i-1} & \cos\varphi_{i-1} \end{bmatrix} \quad (5)$$

Thus the formula of a projection of tops to the card of depth (p) has an appearance

$$V_{i-1} = T_{i-1}^{-1} V_{i-1}^{g} \quad (6)$$

Further, if the projection contains tops (V_i), then it is necessary to calculate new values of a normal (n) for each top.

$$n = T_i N_i(p) \quad (7)$$

Coordinates of tops don't need to be recalculated as the camera in a new shot isn't displaced on spatial axes. If threshold value (ts) that is exceeded the current point will be recognized as the combined. If isn't present, then the point is considered new (from a frame of a new shot of depth).

$$n \times n_{i-1}^{g} < ts \quad (8)$$

Value of a threshold (ts) determines by the following formula

$$\text{argmin} \sum\nolimits_{u} ||(TV_i(u) - V_{i-1}^{g}(u)) \times n_{i-1}^{g}(u)||^2 \quad (9)$$

At the exit of this stage back and forward surfaces in the form of clouds of three-dimensional points and normals to them will be received.

After paired combination of partial frames we will receive separate back and forward surfaces of a body. Between these surfaces will be there is a thin strip of "empty zones". These zones are formed because of inaccessibility to scanning in view of a physical arrangement of cameras.

The main idea of a method of filling the empty zones is represented in Figs. 4 and 5.

For finding of boundary tops threshold value of depth by means of the operator Sobel (3 × 3) is used [12]. The top (P_1) consistently gets out of a set of boundary tops for a forward surface (P_f). For each such top the next top (P_2) gets out of a set for a back surface (P_b). Further is under construction lines of a normal for both tops and the point of intersection (P_c) pays off. On the basis of the received tops the square Bezier curve on the following formula, where B(t)—auxiliary top, t—parameter [0, 1].

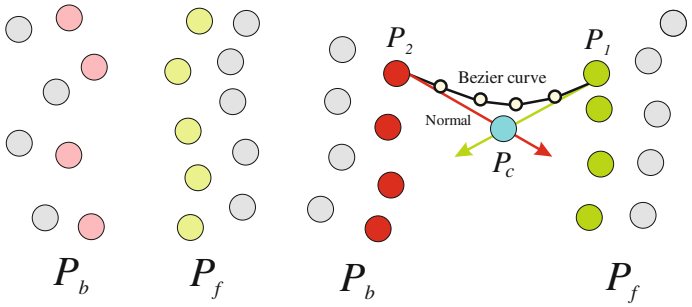


Fig. 4 Illustration of the method of constructing the surface in "empty areas"

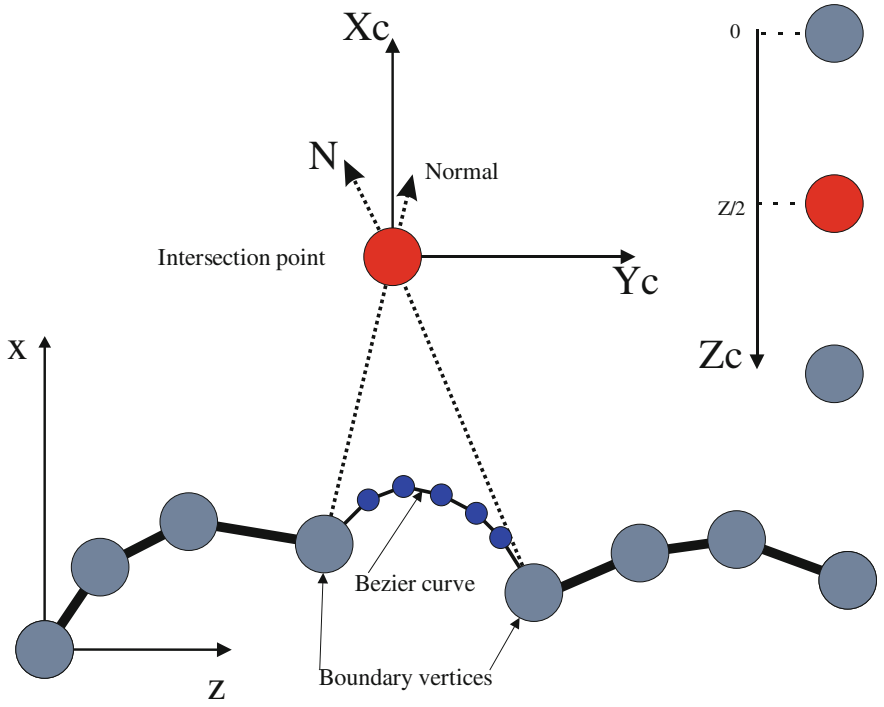


Fig. 5 Building the square Bezier curve

$$B(t) = (1-t)^2 P_1 + 2t(1-t)P_c + t^2 P_2 \tag{10}$$

Thus, "empty zones" will be filled with a set of tops of B(t), as shown in Fig. 5.

The output will be obtained overall model of the human body in the form of three-dimensional vertices reciprocally clouds. The last stage—the construction of

the mesh-method, based on the solution of the Poisson equation. The output is a polygon mesh, which is described geometry of the human body in the format * .ply.

4 Results

On the basis of this method we have developed a program for the reconstruction of surface geometry of the human body. For the implementation of reconstruction used Poisson library PCL (Point Cloud Library). The method of combining the partial surface of the frame is realized on the basis of the library Kinect Fusion. The architecture developed by the program is shown in Fig. 6. In the nearest future development of the system provided for adding functionality to animate models using BVH skeleton.

Figure 7 in left part shows the partial frames depths, which will be processed by our system. On the right represents the reconstructed body model after filtering and alignment.

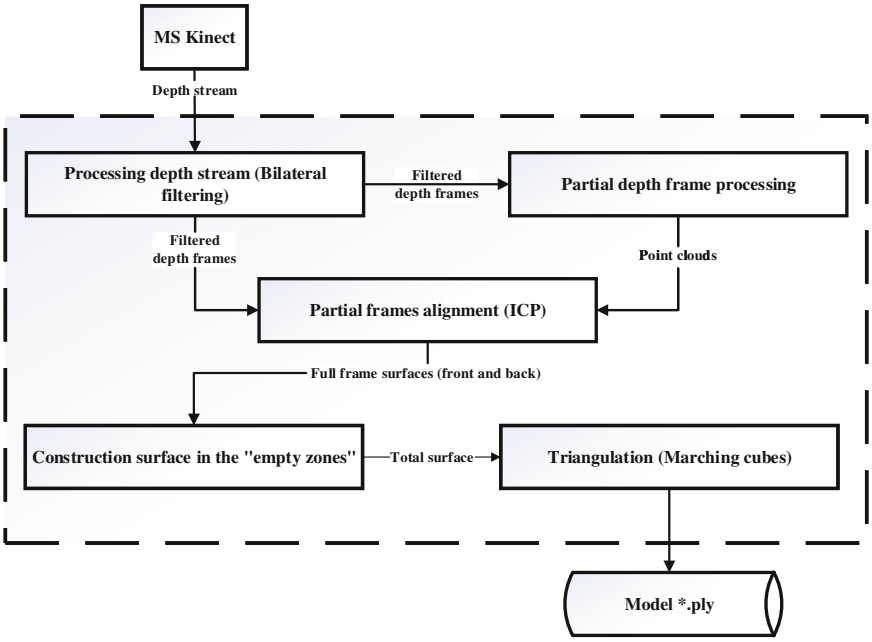


Fig. 6 Architecture

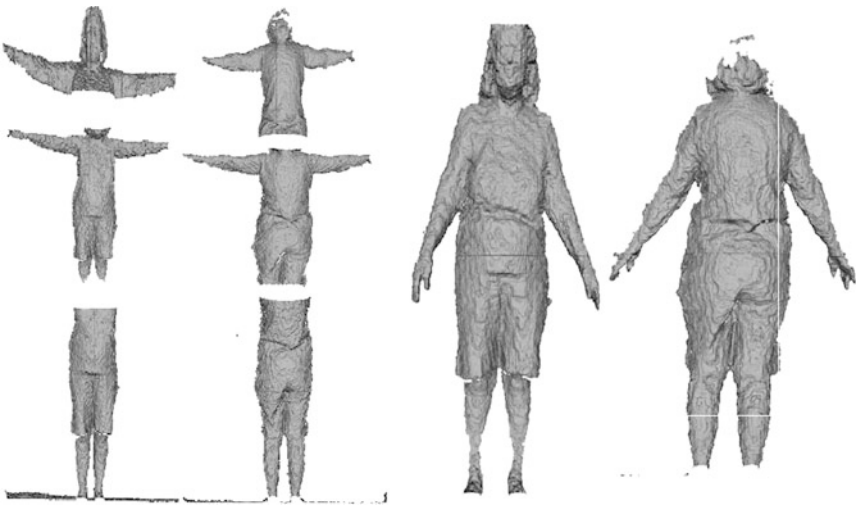


Fig. 7 Reconstructed woman body example

Acknowledgments The study was financially supported by RFBR research projects №15-47-02149, 15-07-06322, 15-37-70014, 16-07-00407, 16-07-00453.

References

1. Huang, X.Q., Adams, B., Wicke, M., Guibas, L.J.: Non-rigid registration under isometric deformations. Comput. Graph. Forum **27**(5), 1449–1457 (2008)
2. Weiss, A., Hirshberg, D., Black, M.J.: Home 3D body scans from noisy image and range data. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1951–1958 (2011)
3. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., et al.: KinectFusion: real-time dense surface mapping and tracking. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011 pp. 127–136 (2011)
4. Rozaliev, V.L., Orlova, Y.A.: Recognizing and analyzing emotional expressions in movements. In: Isaias, P. (ed.) E-Learning Systems, Environments and Approaches. Theory and Implementation, Part II, § 9, pp. 117–131. Springer (2015)
5. Shum, H.P.H., Ho, E.S.L., Jiang, Y., Takagi, S.: Real-time posture reconstruction for Microsoft Kinect. IEEE Trans. Cybern. **43**(5), 1357–1369 (2013)
6. Li, H., Vouga, E., Gudym, A., Luo, L., Barron, J.T., et al.: 3D self-portraits. ACM Trans. Graph. **32**(6) (2013)
7. Sanna, A., Lamberti, F., Paravati, G., Rocha, F.D.: A Kinect-based interface to animate virtual characters. J. Multimodal User Interfaces **7**(4), 269–279 (2013)
8. Alekseev, A., Rozaliev, V., Orlova, Y.: Automatic coloring of grayscale images based on intelligent scene analysis. Pattern Recogn. Image Anal. (Advances in Mathematical Theory and Applications) **25**(1), 10–21 (2015)
9. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Eurographics Symposium on Geometry Processing, pp. 61–70 (2006)

10. Cui, Y., Chang, W., Nöll, T., Stricker, D.: KinectAvatar: fully automatic body capture using a single kinect. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7729 LNCS (PART 2), pp. 133–147 (2013)
11. Lim, H., Lee, S.O., Lee, J.H., Sung, M.H., Cha, Y.W., et al.: Putting real-world objects into virtual world: fast automatic creation of animatable 3D models with a consumer depth camera. In: Proceedings—2012 International Symposium on Ubiquitous Virtual Reality, ISUVR 2012, pp. 38–41 (2012)
12. Chen, Y., Dang, G., Cheng, Z.Q., Xu, K.: Fast capture of personalized avatar using two Kinects. J. Manufact. Syst. **33**(1), 233–240 (2014)

Fourfold Symmetry Detection in Digital Images Based on Finite Gaussian Fields

Alexander Karkishchenko and Valeriy Mnukhin

Abstract This paper considers an algebraic method for symmetry analysis of digital images, based on the interpretation of such images as functions on "Gaussian fields". These are finite fields $\mathbb{GF}(p^2)$ of special characteristics $p = 4k + 3$, where $k > 0$ is an integer. It is shown that such fields may be considered as "finite complex planes" and some properties of such fields are studied. The concept of logarithm in Gaussian fields is introduced and used to define a "log-polar"-representation of digital images. Next, an algorithm for fourfold rotational symmetry detection in gray-level images is proposed.

Keywords Image analysis · Fourfold symmetry · Digital image · Gaussian numbers · Finite fields · Log-polar coordinates · Polar representation

1 Introduction

Symmetry is a central concept in many natural and man-made objects and plays a crucial role in visual perception, design and engineering. That is why symmetry detection and analysis is a fundamental task in such fields of computer science as machine vision, medical imaging, pattern classification and image database retrieval [1, 2]. Numerous applications have successfully utilized symmetry for model reduction [3], segmentation [4], shape matching [1], etc.

Discussing image symmetries, the crucial difference between *continuous* and *discrete* cases must be taken into account. Indeed, a continuous object can be characterized by its symmetry group, which is invariant to rotation, scale and

A. Karkishchenko · V. Mnukhin (✉)
Southern Federal University, 105/42, Bolshaya Sadovaya,
344006 Rostov-na-Donu, Russia
e-mail: mnukhin.valeriy@mail.ru

A. Karkishchenko
e-mail: karkishalex@gmail.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_16

translation transformations. At the same time, for digital images we may talk only about some *measure of symmetry*, which depends on rotations and scaling.

When developing algorithms for image analysis, it is quite common to proceed on the assumption of continuity of images. Then powerful tools of continuous mathematics, such as complex analysis and integral transforms, can be efficiently used. However, its application to digital images often leads to systematic errors associated with the inability to adequately transfer some concepts of continuous mathematics to the discrete plane. As examples we can point to such concepts as rotation in the plane and the polar coordinate system. Being natural and elementary in the continuous case, they lose these qualities when one tries to define them accurately on a discrete plane. As a result, formal application of continuous methods to digital images could be complicated by systematic errors [5, 6]. This raises the issue of the development of methods initially focused on discrete images and based on tools of algebra and number theory.

One of such methods is considered in this paper. It is based on the interpretation of digital images as functions on "Gaussian fields". These are finite fields $\mathbb{GF}(p^2)$ of special characteristics $p = 4k + 3$, where $0 \leq k \in \mathbb{Z}$. We show that elements of such fields may be considered as "finite complex planes" and non-negative functions on such fields may be considered as digital images of prime 'sizes' $p = 7, 11, \ldots, 71, \ldots, 127, 257$, etc. (Note that this is not a limitation since every image can be trivially extended to an appropriate size.)

The significance of such an approach is based on the fact that finite Gaussian fields inherit some properties of the continuous complex field. In particular, we show that the concept of principal value complex logarithm can be transferred to Gaussian fields. As an immediate result, we derive "log-polar"-representation of digital images and use it for the symmetry analysis in the same fashion as for continuous images in [2, 5–7] and for digital images in [12, 15–18]. In this paper we consider the most simple case of fourfold rotational symmetry detection, which, nevertheless, has important practical applications.

2 Finite Fields of Gaussian Integers

Let $\mathbb{Z}$ and $\mathbb{C}$ be the ring of integers and the complex field respectively, let $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ be a residue class ring modulo an integer $n \geq 2$, and let $\mathbb{GF}(p^m)$ be a Galois field with p^m elements, where p is a prime and $m > 0$ is an integer.

In number theory [9, Chap. 1.4] a *Gaussian integer* is a complex number $z = a + bi \in \mathbb{C}$ whose real and imaginary parts are both integers. Note that within the complex plane the Gaussian integers may be seen to constitute a square lattice.

Gaussian integers, with ordinary addition and multiplication of complex numbers, form the subring $\mathbb{Z}[i]$ in the field $\mathbb{C}$. Unfortunately, lack of division in these rings significantly restricts its applicability to image processing problems [14]. So it is natural to look for finite fields, whose properties would be in some respect similar

to properties of $\mathbb{C}$. Note, that if $p = 4k + 3$ is a prime, then the polynomial $x^2 + 1$ is irreducible over $\mathbb{Z}_p$. As an immediate corollary, the next definition follows.

Definition 1 Let $p \geq 3$ be a prime number such that $p \equiv 3 \pmod 4$. Then the finite field

$$\mathbb{C}(p) \stackrel{\text{def}}{=} \mathbb{Z}_p[x]/(x^2 + 1) \simeq \mathbb{GF}(p^2)$$

will be called Gaussian field. Elements of $\mathbb{C}(p)$ will be called discrete Gaussian numbers.

Thus, Gaussian fields have p^2 elements, where

$$p = 3, 7, 11, 19, 23, 31, 43, 47, 59, 67, 71, 79, 83, 103, \ldots .$$

In particular, there are 87 fields $\mathbb{C}(p)$ for $3 \leq p < 1000$. Elements of Gaussian fields are of the form $z = a + bi$, where $a, b \in \mathbb{Z}_p$ and i denotes the class of residues of x, so that $i^2 + 1 = 0$. The multiplication and addition in $\mathbb{C}(p)$ is straightforward, just as notions of the *conjugate* to z number $z^* = a - bi \in \mathbb{C}(p)$ and the *norm* $N(z) = zz^* = a^2 + b^2 \in \mathbb{Z}_p$. (Note that in $\mathbb{C}(p)$ the concept of modulus $|z| = \sqrt{N(z)}$ is not defined.) It is easy to show that $N(z_1 z_2) = N(z_1)N(z_2)$ and $N(z) = 0 \;\Leftrightarrow\; z = 0$.

3 Polar Decompositions of Gaussian Fields

Let us use the analogy between $\mathbb{C}$ and $\mathbb{C}(p)$ to represent elements of $\mathbb{C}(p)$ in an "exponential form". For this, let us recall the algebraic method of introducing log-polar coordinate system onto a continuous complex plane.

Let $\mathbb{C}^*$ be the multiplicative group of complex numbers and $\mathbb{R} = \langle \mathbb{R}, + \rangle$ be the additive group of reals. Note that the correspondence

$$0 \neq z = re^{i\theta} = e^{\ln r + i\theta} \;\leftrightarrow\; (l, \theta), \quad \text{where} \quad l = \ln r \in \mathbb{R} \quad \text{and} \quad 0 \leq \theta < 2\pi,$$

between non-zero complex numbers z and its log-polar coordinates (l, θ) may be considered as an isomorphism

$$\mathbb{C}^* \simeq \mathbb{R} \times (\mathbb{R}/2\pi\mathbb{Z}). \tag{1}$$

In fact, *any* direct product decomposition of $\mathbb{C}^*$ produces a representation of complex numbers.

Let us transfer the previous construction to $\mathbb{C}(p)$. For this, note that since $\mathbb{C}(p)$ is a finite field, its multiplicative group $\mathbb{C}^*(p) = \mathbb{C}(p) \setminus \{0\}$ is cyclic [10, p. 314] and is generated by a primitive element g. For example, it is easy to check that $g = 2 + 7i$ and $p = 1 + 19i$ are primitive in $\mathbb{C}^*(71)$ and $g = 1 + 5i$ is primitive for $p = 251$.

We need the following elementary verifying result.

Lemma 1 *For every* $p = 4k + 3$, *numbers* $m = (p - 1)/2 = 2k + 1$ *and* $n = 2(p + 1) = 8(k + 1)$ *are relatively prime,* $\gcd(m, n) = 1$.

Note that $mn = p^2 - 1 = |\mathbb{C}^*(p)|$ and so for $\mathbb{C}^*(p)$ there appears [10, p. 163] the following analogue of the decomposition (1).

Theorem 1 *For every finite Gaussian field* $\mathbb{C}(p)$, *its multiplicative group is decomposed into direct product of cyclic groups of orders* $m = (p - 1)/2$ *and* $n = 2(p + 1)$,

$$\mathbb{C}^*(p) = \langle g \rangle \simeq \mathbb{Z}_m \times \mathbb{Z}_n. \tag{2}$$

We will call (2) the *polar decomposition* of $\mathbb{C}(p)$.

The polar decomposition can be used to transfer onto $\mathbb{C}(p)$ the concept of complex logarithm. For this, fix any primitive element g and define the mapping $\mathrm{Exp}_g : \mathbb{Z}_m \times \mathbb{Z}_n \to \mathbb{C}^*(p)$ as follows:

$$\mathrm{Exp}_g(l, \theta) = g^{nl + m\theta} = z \in \mathbb{C}^*(p), \quad \text{where} \quad (l, \theta) \in \mathbb{Z}_m \times \mathbb{Z}_n. \tag{3}$$

Thus, Exp_g is an isomorphism between the additive group of the ring $\mathbb{Z}_m \times \mathbb{Z}_n$ and the multiplicative group of the Gaussian field $\mathbb{C}(p)$.

Definition 2 The mapping $\mathrm{Exp}_g : \mathbb{Z}_m \times \mathbb{Z}_n \to \mathbb{C}^*(p)$ is called the modular exponent to base g, and its inverse mapping $\mathrm{Ln}_g : \mathbb{C}^*(p) \to \mathbb{Z}_m \times \mathbb{Z}_n$ is the modular logarithm to base g. Its domain $\mathbb{Z}_m \times \mathbb{Z}_n$ is called the polar domain.

The "*basic logarithmic identity*" follows immediately:

$$\mathrm{Ln}_g(z_1 z_2) = \mathrm{Ln}_g(z_1) + \mathrm{Ln}_g(z_2). \tag{4}$$

Note that $\mathrm{Ln}_g(0)$ is not defined, and to evaluate $(l, \theta) = \mathrm{Ln}_g(z)$ for any $z = g^s \in \mathbb{C}^*(p)$ one just needs to solve the Diophantine equation $nx + my = s$ and set

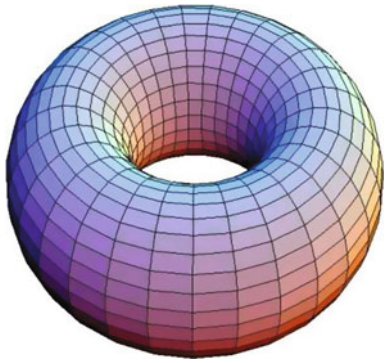
$$(l, \theta) = (x \bmod m,\ y \bmod n) \in \mathbb{Z}_m \times \mathbb{Z}_n. \tag{5}$$

Example 1 Let $g = 1 + 2i \in \mathbb{C}^*(7)$ and $z = g^2 = 4 + 4i$. Then $s = 2$, $m = 3$, $n = 8$ and the equation $8x + 3y = 2$ has the evident solution $x = 1$, $y = -2$. Hence, $\mathrm{Ln}_g(4 + 4i) = (1 \bmod 3,\ -2 \bmod 8) = (1, 6) \in \mathbb{Z}_3 \times \mathbb{Z}_8$.

Note that depending on g, either $\mathrm{Ln}_g(i) = (0, n/4)$ or $\mathrm{Ln}_g(i) = (0, 3n/4)$.

We will consider the pair $(l, \theta) \in \mathbb{Z}_m \times \mathbb{Z}_n$ as "log-polar coordinates" of the corresponding discrete Gaussian integer z. It is useful to consider $\mathbb{Z}_m \times \mathbb{Z}_n$ as a $m \times n$-discrete torus, see Fig. 1.

Fig. 1 A discrete torus



4 Digital Images as Functions on Gaussian Fields

Let $\mathbb{C}(p)$ be any Gaussian field of a characteristic $p = 4k + 3$ and let $f(z) : \mathbb{C}(p) \to \mathbb{R}$ be any real-valued function on $\mathbb{C}(p)$. Due to our geometric interpretation of Gaussian fields, we call $f(z)$ a *digital gray-level image of size* $p \times p$, or just $p \times p$-*image* briefly.

Another way to describe images is based on polar decompositions of Gaussian fields. Namely, we may associate with a $p \times p$-image $f(z)$ an image ψ of size $m \times n$.

For this, fix any primitive element $g \in \mathbb{C}^*(p)$ and define a function $\psi : \mathbb{Z}_m \times \mathbb{Z}_n \to \mathbb{R}$ such that

$$\psi(\mathrm{Ln}_g(z)) = f(z), \quad 0 \neq z \in \mathbb{C}^*(p). \tag{6}$$

Definition 3 The transform $\mathscr{P}_g[f] = \psi$ will be called log-polar transform to base g of the image *f*, or just its polar transform $\mathscr{P}$ briefly. The image ψ will be called the polar form of *f*.

Thus, the polar form of *f* may be considered as some arrangement of all its pixels except $f(0, 0)$ in the form of a $m \times n$-matrix, as it is shown in Fig. 2. Thus, the image *f* is "almost" recovered by its polar form ψ. In order to achieve full recoverability, let us formally agree to extend the polar domain by an extra element ∞, assuming that $\psi(\infty) = f(0, 0)$. Note that in such an *extended polar domain* $\mathbb{Z}_m \times \mathbb{Z}_n \cup \{\infty\}$ the polar transform $\mathscr{P}$ becomes invertible. The linearity of $\mathscr{P}$ is obvious. (Just as discrete torus is an intuitive interpretation of the polar domain, we may interpret the extended polar domain as a torus with a small sphere inside.)

As the following statement shows, the transform $\mathscr{P}$ indeed can be regarded as a discrete analogue of the transition to the log-polar coordinate system.

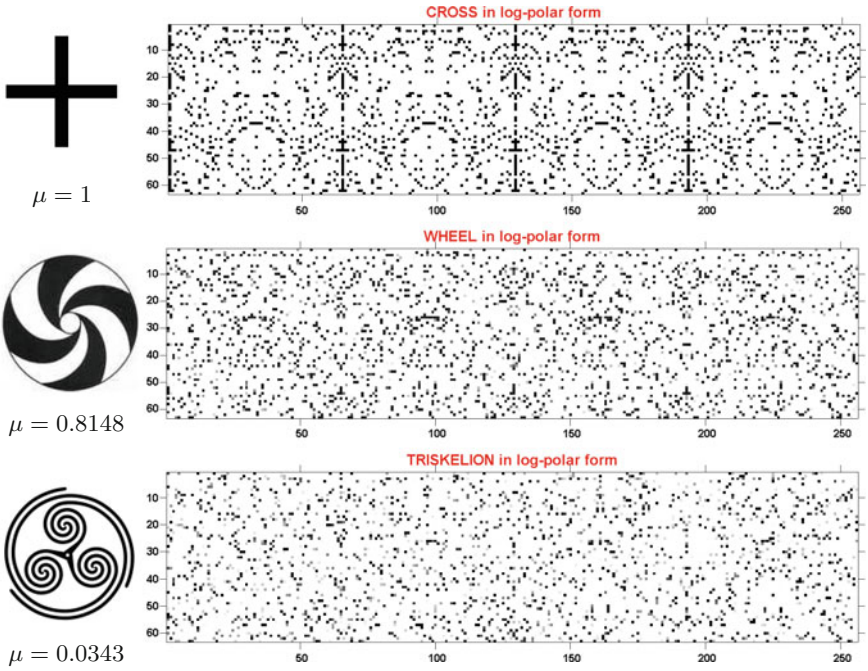


Fig. 2 Images, their polar forms and measures of 4-symmetry

Proposition 1 *If* $\mathscr{P}_g[f(z)] = \psi(l, \theta)$, *then*

$$\mathscr{P}[f(wz)] = \psi(l - l_0, \theta - \theta_0) \tag{7}$$

where $0 \neq w \in \mathbb{C}(p)$ *and* $\mathrm{Ln}(w) = (l_0, \theta_0)$.

We will apply the previous relation (7) to symmetry analysis. As is known, a continuous object is said to have *r-fold rotational symmetry* with respect to a point C if a rotation by an angle of $2\pi/r$ around C does not change the object. Unfortunately, this definition does not work for digital images because digital rotation is much more hard to define (see, for example, [8, p. 377]). As a result, for digital images we may talk only about some *measure of symmetry*, which depends on rotations and scaling.

The geometric interpretation of Gaussian fields immediately implies the following definition.

Definition 4 A digital image f is said to have fourfold central rotational symmetry if and only if $f(iz) = f(z)$.

Let $\psi = \mathscr{P}[f]$ be the polar form of a $p \times p$-image f. We may consider ψ as an $m \times n$-matrix, where $n = 8(k+1)$, $m = 2k+1$ and $k = (p-3)/4 \in \mathbb{Z}$. Since n is multiple of 4, decompose ψ into four blocks ψ_t, $t = 0, 1, 2, 3$, of equal size $m \times n/4$.

Proposition 2 *An image f is fourfold central symmetric if and only if its polar form ψ can be decomposed into four equal blocks, $\psi_0 = \psi_1 = \psi_2 = \psi_3$.*

Proof Since the polar transform is invertible, it follows from Definition 4, that f is fourfold symmetric if and only if $\mathscr{P}[f(\omega z)] = \mathscr{P}[f(z)] = \psi(l, \theta)$. But it have been noted in Sect. 3 that $\mathrm{Ln}_g(\omega) = (0, \pm n/4)$ for every primitive $g \in \mathbb{C}^*(p)$. Hence, fourfold symmetry of f is equivalent to the condition

$$\psi(l, \theta \pm n/4) = \psi(l, \theta) \quad \text{forall} \quad l \in \mathbb{Z}_m,\ \theta \in \mathbb{Z}_n,$$

or just $\psi_0 = \psi_1 = \psi_2 = \psi_3$.

Evidently, for real-world images one can expect only approximate equalities $\psi_0 \simeq \psi_1 \simeq \psi_2 \simeq \psi_3$, so that the problem to choose an appropriate measure of symmetry $\mu(f)$ for an image f arises. One of possible ways is the following. For any normalized polar form matrix $\tilde{\psi} = \psi / \max\{\psi\}$ of an image f we introduce

$$\mu(f) = \exp(-\alpha x^{\beta}), \quad \text{where} \quad x = \max_{0 \le i < j \le 3} \left\{ \frac{||\tilde{\psi}_i - \tilde{\psi}_j||}{||\tilde{\psi}_i \mathrm{P} + \mathrm{P}\tilde{\psi}_j||} \right\}. \tag{8}$$

Here $||\cdot||$ stands for any matrix norm, and α, β are nonnegative reals, whose precise values could vary depending on the practical problem to be solved. Then $\mu(f)$ measures the central fourfold symmetry of f.

Note 1. Since the polar form ψ depends on the primitive element $g \in \mathbb{C}^*(p)$, one may ask if $\mu(f)$ also depends on g. It is an easy matter to use the "change of base" formula [15, Prop. 7] to prove invariance of $\mu(f)$ under change of g.

Example 2 To illustrate the method we consider images f_1, f_2, f_3 shown in Fig. 2 on the left. These are square-sampled 127×127-images, binary the first and gray-level the last two. Polar form matrices ψ_i of the images, evaluated for $g = 1 + 8i \in \mathbb{C}^*(127)$, are in the right side of Fig. 2. The evaluated symmetry measures are the following:

$$\mu(f_1) = 1.000, \quad \mu(f_2) = 0.8148, \quad \text{and} \quad \mu(f_3) = 0.03434,$$

where the Frobenius matrix norm was used in (8) and the parameters were taken to be $\alpha = 24$, $\beta = 4$.

Note that for fixed p, the polar transform is based on precomputations and do not require any arithmetic operations. Jointly with any of the sliding window methods, the introduced approach can be used to detect centers of local fourfold symmetry in images.

Example 3 The upper part of Fig. 3 demonstrates the flag-throwing competition at the Swiss National Day celebration in Lausanne (flag twirling is one of the oldest national sports of Switzerland). A sliding 127×127-window algorithm based on the discussed above method, has been applied to this 659×659-grayscale image. As the result, 659×659-matrix $M = (\mu_{ij})$ has been evaluated, where μ_{ij} is the

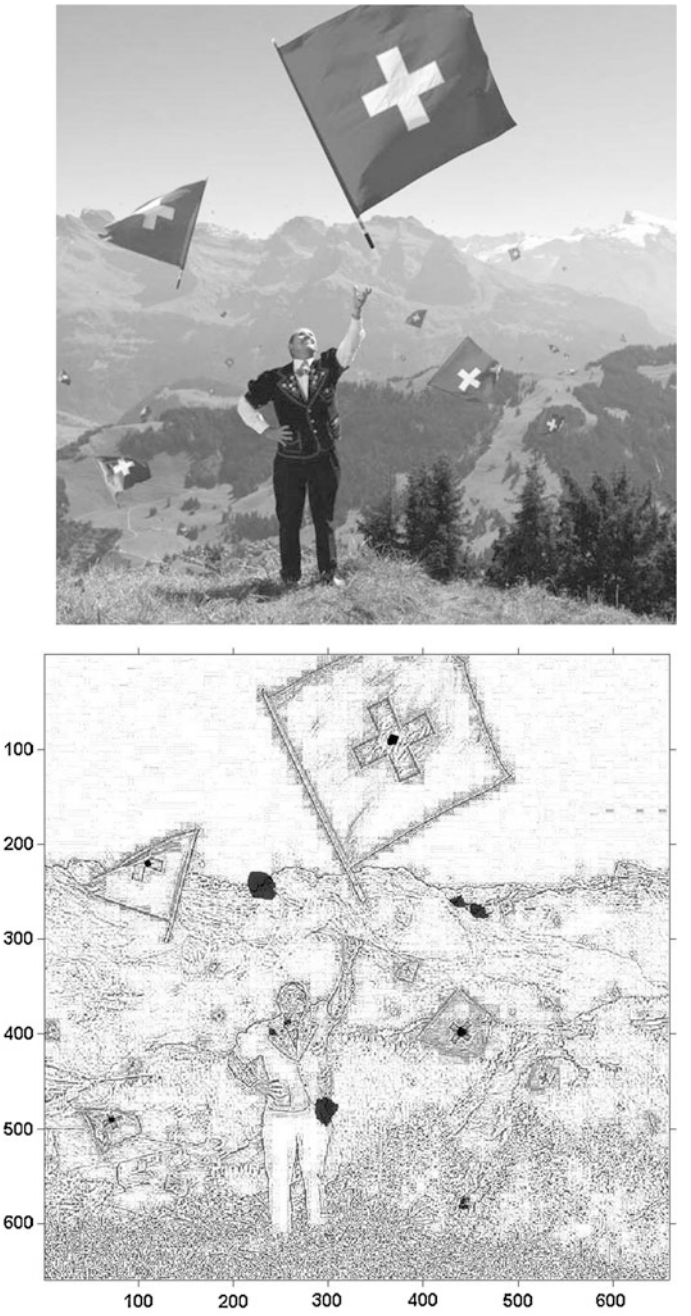


Fig. 3 Sliding window test result

measure of 4-symmetry with the center at (i,j), for $64 \le i,j \le 596$, and $\mu_{ij} = 0$ outside the diapason. Then M has been cutted-off at 0.78 rate and combined with the "pensil-drawn" transform of the original image. The result is shown in the lower part of Fig. 3.

As it is easy to notice, centers of all the four biggest flags have been detected. Measures of symmetry for smaller flags occurred to be below the cutoff level due to the large size of the sliding window. At the same time, few details of mountain ridges jointly with some elements of the man's suit produce significant responses. It indicates necessity of further improvements; however, these are beyond the scope of the paper.

5 Conclusion

This paper proposes an algebraic method for the processing of digital images, based on their representation as functions on special finite fields, called "Gaussian fields". The concept of logarithm in such fields is introduced and used to define a "log-polar"-representation of digital images. An algorithm for fourfold symmetry detection in gray-level images is proposed. However, apart from this work, there remain issues such as the detailed study of properties of the introduced polar transformation, the study of possibilities of its generalizations, as well as the details of its practical application. Authors hope to return to the study of these issues in the future.

Acknowledgements This research has been partially supported by the Russian Foundation for Basic Research grant no. 16-07-00648.

References

1. Gool, L., Moons, T., Ungureanu, D., Pauwels, E.: Symmetry from shape and shape from symmetry. Int. J. Robot. Res. **14**(5), 407–424 (1995)
2. Martinet, A., Soler, C., Holzschuch, N., Sillion, F.: Accurate detection of symmetries in 3D shapes. ACM Trans. Graph. **25**(2), 439–464 (2006)
3. Mitra, N.J., Guibas, L.J., Pauly, M.: Partial and approximate symmetry detection for 3D geometry. ACM Trans. Graph. **25**(3), 560–568 (2006)
4. Thrun, S., Wegbreit, B.: Shape from symmetry. In: Proceedings of the International Conference on Computer Vision (ICCV), vol. 2, pp. 1824–1831 (2005)
5. Chertok, M., Keller, Y.: Spectral symmetry analysis. IEEE Trans. Pattern Anal. Mach. Intell. **32**(7), 1227–1238 (2010)
6. Derrode, S., Ghorbel, F.: Shape analysis and symmetry detection in gray-level objects using the analytical fourier-mellin representation. Signal Process. **84**(1), 25–39 (2004)
7. Karkishchenko, A.N., Mnukhin, V.B.: Symmetry recognition in the frequency domain (in Russian). In: 9th International Conference on Intelligent Information Processing. pp. 426–429. TORUS Press, Moscow (2012)

8. Karkishchenko, A.N., Mnukhin, V.B.: Topological filtration for digital images recognition and symmetry analysis. J. Mach. Learn. Data Analysis. **1**(8), 966–987 (2014). (in Russian)
9. Ireland, K., Rosen, M.: A Classical Introduction to Modern Number Theory. Springer, New York (1982)
10. Dummit, D.S., Foote, R.M.: Abstract Algebra. Wiley (2004)
11. Campello de Souza, R.M., de Oliveira, H.M. Silva, D.: The *Z* Transform over Finite Fields. arXiv preprint 1502.03371 published online 11 February 2015
12. Campello de Souza, R.M., Farrell, R.G.: Finite field transforms and symmetry groups. Discret. Math. **56**, 111–116 (1985)
13. Bandeira, J., Campello de Souza, R.M.: New trigonometric transforms over prime finite fields for image filtering. In: VI International Telecommunications Symposium (ITS2006), pp. 628–633, Fortaleza-Ce, Brazil (2006)
14. Baker, H.G.: Complex gaussian integers for gaussian graphics. ACM Sigplan. Notices **28**(11), 22–27 (1993)
15. Mnukhin, V.B.: Transformations of digital images on complex discrete tori. Pattern Recognit. Image Anal.: Adv. Math. Theory Appl. **24**(4), 552–560 (2014)
16. Mnukhin, V.B.: Digital images on a complex discrete torus. Mach. Learn. Data Anal. **1**(5), 540–548 (2013). (in Russian)
17. Karkishchenko, A.N., Mnukhin, V.B.: Applications of modular logarithms on complex discrete tori in problems of digital image processing. Bull. Rostov State Univ. Railway Transp. (ISSN 0201-727X). **3**, 147–153 (2013) (in Russian)
18. Mnukhin V.B.: Fourier-Mellin transform on a complex discrete torus. In: Proceedings of the 11th International Conference "Pattern Recognition and Image Analysis: New Information Technologies" (PRIA-11-2013). pp. 102–105, The Russian Federation, Samara (2013)
19. Varitschenko, L.V., Labunets, V.G., Rakov, M.A.: Abstract Algebraic Systems and Digital Signal Processing (in Russian). Naukova Dumka, Kiev (1986)

Inpainting Strategies for Reconstruction of Missing Data in Images and Videos: Techniques, Algorithms and Quality Assessment

Viacheslav Voronin, Vladimir Marchuk, Dmitriy Bezuglov and Maria Butakova

Abstract This paper describes a method and algorithm for spatially-temporally consistent image and video completion. We propose modification of an image inpainting algorithm based texture and structure reconstruction. Proposed method allows to remove static and dynamic objects and restore missing regions using spatial and temporal information from neighboring frames. This paper also focuses on a machine learning approach for no-reference visual quality assessment for image and video inpainting. Experimental comparisons to state-of-the-art inpainting methods demonstrate the effectiveness of the proposed approaches.

Keywords Image processing · Inpainting · Reconstruction · Quality assessment

1 Introduction

Nowadays all type of the communication techniques has been done with the help of the multimedia image and video data. Image inpainting or "image completion" is one of the important topics in image processing, which can be applied in many areas, from the automatic restoration of damaged photographs to removal of unwanted objects in images. Reconstruction methods use information from outside of the damaged area for interpolating missing pixels. The application of inpainting are numerous: text removing, old photo scratch restoring and different artistic film effects. The main purpose of video inpainting is filling-in holes in a video sequence using spatial and temporal information from neighboring regions. The holes may correspond to missing parts or removed objects from the scenes (logos, text, etc.). Video inpainting techniques find a wide range of applications in video processing

V. Voronin (✉) · V. Marchuk
Don State Technical University, Rostov-on-Don, Russia
e-mail: voroninslava@gmail.com

D. Bezuglov · M. Butakova
Rostov State Transport University, Rostov-on-Don, Russia
e-mail: bezuglovda@mail.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_17

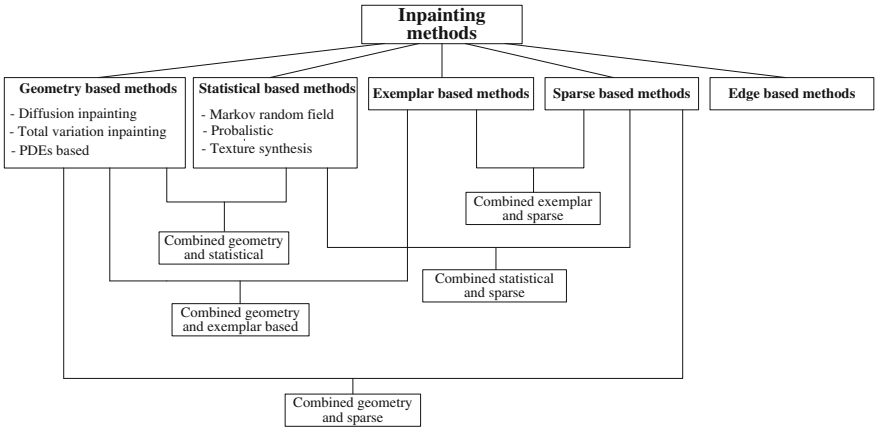


Fig. 1 The classification of inpainting methods

where an unknown region in the scene has to be estimated in a way providing visually coherent results. For instance, error/loss concealment which is a post-processing technique usually performed at the decoder side to recover missing blocks of decoded streams may represent an interesting application of video inpainting techniques.

The image inpainting methods can be classified into the several groups: based on geometry [1, 2], statistics [3, 4], sparsity [5], exemplars [6, 7] and edges [8] methods. The classification of inpainting methods is shown in Fig. 1. This classification is based on two-dimensional representations of various signals in the context of image properties. Analysis of the literature shows that each of the methods has specific application associated with the class of image model and uses the definition of a smooth, local or homogeneous region.

The geometry-based methods for image reconstruction use partial differential equations (PDE) and restore missing pixels by a diffusion process. These methods have some drawback comprising structure and texture blurring. The PDEs methods more suitable for removal scratches and small defects on the images. Statistical-based methods use Markov random field or texture synthesis approach. These algorithms show good results for a stochastic texture and fail in curved edges reconstruction. Sparsity-based methods often blur texture and structure in a recovery of large missing areas. Another problem for this group of inpainting methods is computationally complexity. The exemplar-based approaches use a non-parametric sampling. The basic idea is searching similar patches and coping them from the true image. The main drawback is limitation to inpaint curvy structures. In many cases cannot be found similar patches because the patch size is large or singular.

Recently, various image inpainting algorithms based on combinational edge and texture reconstruction have been developed. They show high quality results, but re-quire a significant computational time to inpaint large missing areas.

One of the current trends is the development of methods based on contour models, that is, the boundaries of the interpolation, for visual analysis is first necessary to correctly restore the most informative part of the image—object structure.

An analysis of this classification leads to the conclusion that at the present time are promising methods, which use combined local and global properties of the image. These properties include texture and contour characteristics. In the method based on a combination of structural recovery and various local characteristics (stochastic textures) does not currently offer combined approaches that combine the best properties of these methods groups.

In this paper we introduce a novel algorithm for automatic image and video inpainting based on 2D autoregressive texture model, exemplar-based and structure curve construction.

2 Image and Video Inpainting

2.1 An Image Model

For image and video we use geometric image model which was proposed in [9]. In the most cases any image has several types of regions with different the local geometric features (Fig. 2a). There are texture, non texture and contours of objects. A simplified mathematical model of the original image can be represented as follows:

$$Y_{i,j} = (1 - M_{i,j}) \cdot S_{i,j} + M_{i,j} \cdot R_{i,j}, \quad i = \overline{1,I}, \quad j = \overline{1,J}, \tag{1}$$

where $S_{i,j}$ are 'true' image pixels; $M = \left|M_{i,j}\right|$ is a binary mask of distorted values of pixels (1—corresponds to the missing pixels, 0—to the true pixels); $R_{i,j}$ are missing

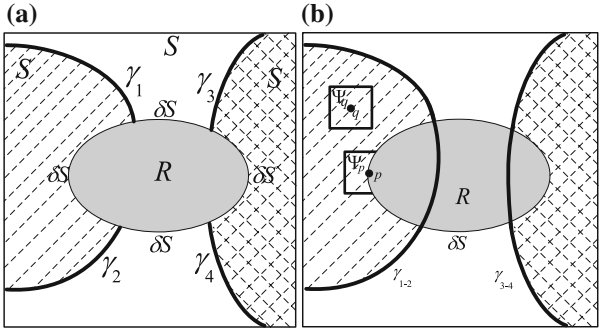


Fig. 2 The image model

pixels; δS is the boundary of the recovery region; $\gamma_1, \gamma_2 \ldots \gamma_k \ldots \gamma_L$, $k = \overline{1, L}$ are the edges crossing an area with a missing pixels R; I and J is the number of rows and columns correspondence.

2.2 Proposed Method

In our previous work [9] we have proposed the image inpainting method. This approach based on the separate structure and texture prediction with a high accuracy. In this paper we propose the algorithm and its competitive analysis with state-of-the-art methods using proposed non reference image quality assessment.

We summarize proposed algorithm in the following scheme (Fig. 3).

1. *Segmentation using Chan–Vese (CV) model*
   For the image segmentation around missing pixels in proposed algorithm we use the Chan–Vese (CV) approach based on solves the minimization of the energy functional (Fig. 4) [10].

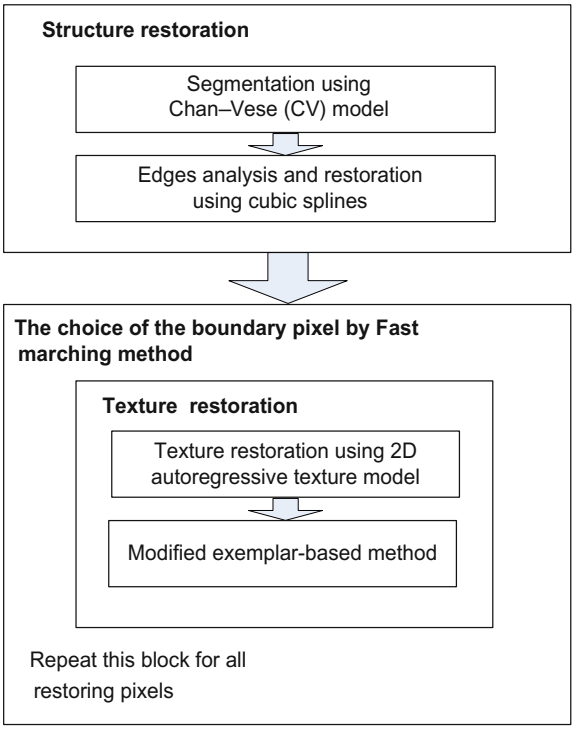


Fig. 3 The proposed inpainting algorithm

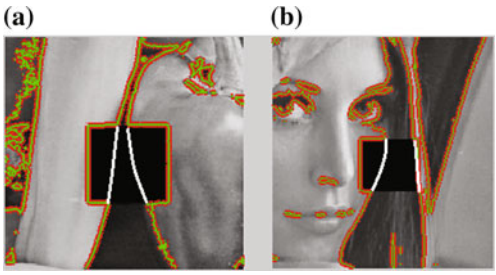


Fig. 4 Examples of segmentation and structure curve construction

$$E^{CV}(c_1, c_2, C) = \mu \cdot Length(C) + \lambda_1 \cdot \int\limits_{inside(C)} |u_0(x,y) - c_1|^2 dxdy + \lambda_2 \cdot \int\limits_{outside(C)} |u_0(x,y) - c_2|^2 dxdy, \tag{2}$$

where μ , λ_1 and λ_2 are positive constant, usually fixing $\lambda_1 = \lambda_2 = 1$, c_1 and c_2 are the intensity averages of u_0 inside C and outside C, respectively.

2. *Edges analysis and restoration using cubic splines*
On this step for edges reconstruction we use the cubic spline interpolation for each of the two parts of the contours. We calculate the cubic spline Hermite for points P_k and P_l on the edges and non-zero tangent vectors $\mathbf{Q}_k$ and $\mathbf{Q}_l$ using the vector equation in form:

$$\mathbf{B}(t) = (1 - 3t^2 + 2t^2)P_k + t^2(3 - 2t)P_l + t(1 - 2t + t^2)\mathbf{Q}_k - t^2(1 - t)\mathbf{Q}_l, 0 \le t \le 1. \tag{3}$$

3. *The choice of the boundary pixel by fast marching method*
The fast marching method is used in order to select a restored pixel in the area R, based on the solution of Eikonal equation, where the solution of the equation is the distance map of ψ_p pixels to the boundary δS (Fig. 2b) [11].
4. *Texture restoration using 2D autoregressive texture model*
The texture reconstruction algorithm is a modification of the example-based method proposed by Criminisi et al. [6]. Before copy-past procedure we propose to use patch prediction step based on the autoregressive (AR) model [9]:

$$\hat{\Psi}_p = \sum\nolimits_{m \in s(o,p]} \varphi_m \cdot \hat{\Psi}_{s-m} + \sum\nolimits_{n \in s(o,q]} \vartheta_n \cdot \eta_{s-n} + \eta_s, \tag{4}$$

where $(\varphi_m)_{m \in s(o,p]}$ and $(\vartheta_n)_{n \in s(o,q]}$ denotes, respectively the autoregressive and moving average parameters with $\varphi_0 = \vartheta_0 = 1$, and η_s denotes a sequence of distributed centered random variables with variance σ^2.
Model parameters are estimated by Yule-Walker method.

5. *Modified exemplar-based method*
   After a texture reconstruction for every $\hat{\Psi}_p$ on the image S we find a patch Ψ_q, for which the distance is minimal (Fig. 2b):

$$D_E(\hat{\Psi}_p, \Psi_q) = \sqrt{\sum (\hat{\Psi}_p - \Psi_q)^2} \rightarrow \min. \quad (5)$$

On last step the pixels for Ψ_q is restored by copying the pixels from $\hat{\Psi}_p$ within the spline interpolated boundaries.

2.3 Inpainting Quality Assessment

Inpainting image quality is not trivial task because in most cases the reference image is absent. Typically, expert-based approaches are involved but they are expensive and time consuming procedure. The alternative approach is objective image quality assessment based on the machine learning. The problem of inpainted image quality assessment is highly related to the human visual system modeling problem. So in order to design a good inpainting quality assessment one should take into account its different properties.

The visual attention is important in human visual perception. At any time, the human eye sees clearly only a small part of the scene, at the same time much more significant area of the scene is perceived as blur. This "diffuse information" sufficient to assess the importance of different areas of the stage and draw attention to important areas of the visual field.

In [12], we have proposed the inpainting quality assessment technique based on a machine learning approach. The algorithm of proposed method is presented on the Fig. 5.

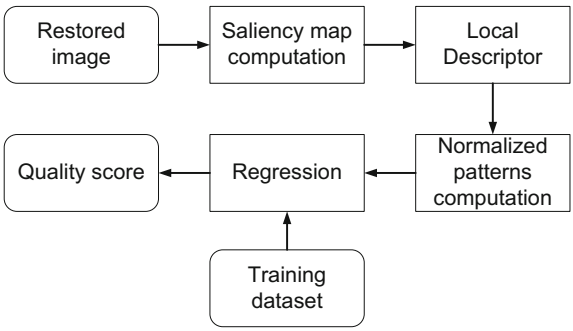


Fig. 5 Overall algorithm block scheme

To select the areas of interest are encouraged to use the following expression:

$$GD_{out} = \frac{1}{\|R\|} \cdot \sum_{p \in R} S'(p). \tag{6}$$

where S' is saliency map for the reconstructed image, $S'(p)$ is saliency map for the corresponding pixel p; R is amount if pixels for $S' > 0$. We have used GD_{out} value as a threshold level at the next step and calculate the assessment only for those recovered areas for which $S(p) > GD_{out}$.

The spectral representation is used as low-level feature of local areas. It is further proposed analysis of the following basis Fourier, Walsh, Haar using vector efficiency. For correct calculation of the components of the system efficiency criterion in the presence some distortions requires the use of statistical averaging.

The criterion of minimum average risk is used for signal processing algorithms and systems synthesis. In general, a simplified mathematical model of the image transformation is represented as:

$$\bar{S}_{i,j} = T\left[Y_{i,j}\right], i = \overline{1,I}, j = \overline{1,J}, \tag{7}$$

where $\overline{S}_{i,j}$ is estimate of the image; T is a transformation operator. The probability for each of the subclasses P_ℓ assumed to be known for two-dimensional processes $\{Y_\ell(i,j)\}_{\ell=1}^{L}$, $Y_\ell(i,j) = \left\{y_\ell^{(k)}(i,j)\right\}_{k=1}^{\infty}$, where index ℓ is number of the subclasses, k is number of the realization in every subclass ℓ. Let's compare some set of basic functions $\left\{\varphi_{n,m}^{(\mathrm{h})}(i,j)\right\}_{n,m=1}^{\infty}$, $\quad \mathrm{h} = 1, 2 \ldots H$.

The Fourier transform for image is represented in the form:

$$y_{\ell h}^{(k)}(i,j) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{n,m} \varphi_{n,m}^{h}(i,j), \tag{8}$$

where $a_{n,m}$ are the Fourier coefficients.

If number of series is limited then approximation error presented as:

$$\varepsilon_{\ell h}^{(k)} = \rho\left[y_\ell^{(x)}(i,j); \quad \tilde{y}_\ell^{(k)}(i,j;K)\right], \tag{9}$$

where K is the number of Fourier coefficients.

For estimate computational complexity of the Fourier transform we propose cost function $S_h(\ell, K)$. The total cost function which includes both approximation error and complexity represented in the form:

$$W_{\ell h}^{(k)} = w\left[\varepsilon_{\ell h}^{(x)}(\ell, K); \quad S_h^{(k)}(\ell, K)\right]. \tag{10}$$

The value of the conditional risk depends from the subclass of signals and basis. It is calculated by averaging the cost function (10):

$$r_{\ell h} = < W_{\ell h}^{(k)} >_k = \int w\left[\varepsilon_{\ell h}^{(x)}(\ell, N); \quad S_h^{(k)}(\ell, N)\right] f(x)dx, \tag{11}$$

where $f(x)$ is the probability density of the analyzed signal and noise.

The average risk is determined by averaging the conditional risk for every subclass signals:

$$R_h = < r_{\ell h} > = \sum_{\ell=1}^{L} P_\ell \cdot r_{\ell h}, \tag{12}$$

where P_l is the probability for every subclass signal.

In accordance with the criterion of minimum average risk chosen the basis H for which the average risk is minimal.

To evaluate the effectiveness of the imaging quality criteria we considered test image textures obtained by Gaussian field model with predetermined correlation functions. As an example, in the paper we studied different basis function as Fourier, Walsh and Haar for different probability subclasses ($P_1 = P_2 = P_3 = 0.33$; $P_1 = 0.5;\ P_2 = 0.3;\ P_3 = 0.2;\ P_1 = 0.2;\ P_2 = 0.3;\ P_3 = 0.5$).

We choose cost function in the form:

$$W_{qm}^{(k)} = \varepsilon_{mq}^2 H_m. \tag{13}$$

In Table 1 presents the average risk for different basis and correlation functions. The comparative analysis shows what the basis Haar have the smaller average risk then Walsh and Haar.

In Table 2 presents the average risk for different probability subclasses.

Analysis of the results presented in Table 2 shows that for all image models of the probability distributions Haar basis is characterized by the lowest average risk. Based on this analysis we choose the Haar basis as local features for detected areas on image.

After calculating frequency histogram for inpainted regions feature we apply machine learning stage. For this purpose uses Support Vector Regression (SVR) as a learning method. The experimental method for the subjective quality assessment was chosen the Mean Opinion Score (MOS), which values received from

Table 1 The average risk for different correlation functions

Basis	Correlation functions		
	$K_1(\tau_x, \tau_y)$	$K_2(\tau_x, \tau_y)$	$K_3(\tau_x, \tau_y)$
Fourier	1.73	0.15	0.29
Walsh	0.45	0.13	0.23
Haar	0.095	0.015	0.018

Table 2 The average risk for different probability subclasses

Basis	Type of the probability		
	1	2	3
Fourier	0.71	0.97	0.53
Walsh	0.26	0.3	0.24
Haar	0.043	0.06	0.03

participants elevation test dataset. For every restored image used the quality scale as: 5—Artifacts are Imperceptible; 4—Artifacts are perceptible buy not annoying; 3—Artifacts are slightly annoying; 2—Artifacts are annoying; 1—Artifacts are very annoying.

3 Experimental Results

In our experiments we use database of 300 images (texture, structure and real images). For every image we apply the mask with missing pixels (black square blocks) and reconstructed by different approaches. One of the image inpainting result presented at Fig. 6 (a—images with missing pixels, b—images reconstructed

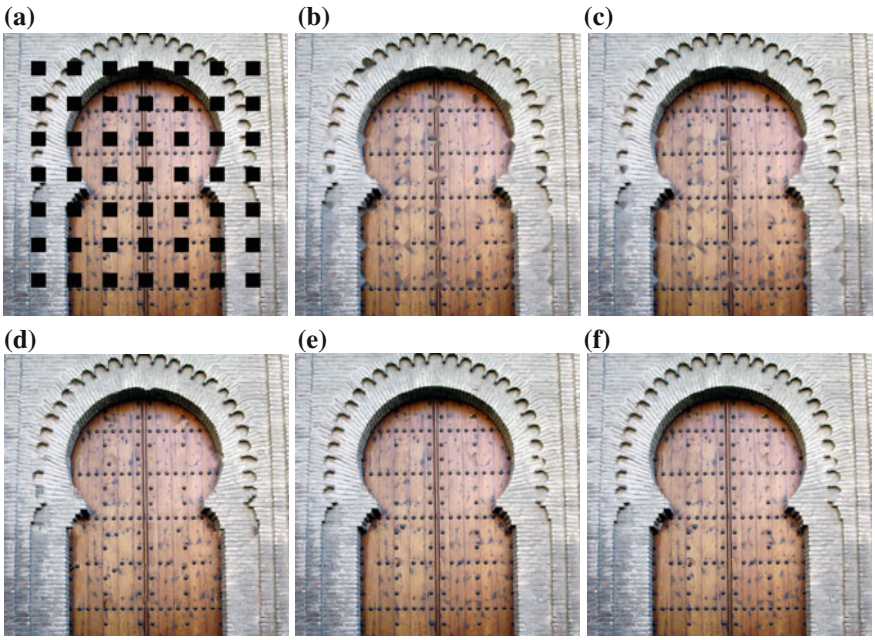


Fig. 6 Examples of image restoration

Table 3 Results of numeric comparison

Methods	Type of image	*MOS*	$\overline{MOS}$	Proposed metric
Smoothing	Texture	2.21	2.08	0.62
	Structure	1.58		
	Image	2.45		
Navier-Stokes	Texture	3.15	2.69	0.72
	Structure	1.73		
	Image	3.21		
Telea	Texture	3.08	2.78	0.75
	Structure	2.02		
	Image	3.23		
Exemplar-based	Texture	4.41	3.69	0.81
	Structure	3.13		
	Image	3.54		
Proposed method	Texture	4.52	4.28	0.9
	Structure	4.11		
	Image	4.21		

by the Smoothing, c—images reconstructed by the Navier-Stokes, d—images reconstructed by the Telea, e—images reconstructed by the exemplar-based, f—images reconstructed by the proposed method).

The objective quality was assessed by 10 human observers for each inpainted image. The received results were statistically analyzed by MOS estimate. Results of numeric comparison are presented in the Table 3. Proposed objective quality metric shows strong correlation with perceived by human quality. Thus, our approach is quite efficient to estimate quality of the inpainted images.

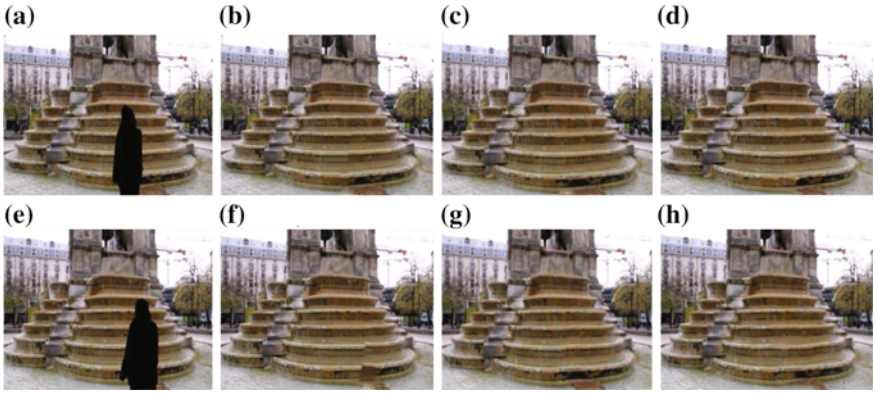


Fig. 7 Examples of video completion

We also present some results of video inpainting on the test video scene (Fig. 7). After applying the missing mask, all frames have been inpainted by the proposed method and state-of-the-art methods [13, 14]. In Fig. 7 examples of video completion (a, e—the image with a missing pixels, b, f—the restoration by the Wexler method, c, g—the restoration by the Newson method, d, h—the restoration by the proposed method) are shown.

4 Conclusions

This paper describes a framework for spatially-temporally consistent image completion and no-reference quality assessment. The proposed inpainting method is modification of the patch-based approach with ability texture synthesis and curved structure interpolation. The novel no-reference inpainting quality assessment technique is based on a machine learning approach. We have demonstrated that predicted quality value highly correlates with a qualitative opinion in a human observer study. Experimental comparisons to state-of-the-art inpainting methods demonstrate the effectiveness of the proposed approaches.

Acknowledgments The reported study was funded by RFBR according to the research project 16-07-00888-a.

References

1. Bertalmio, M., Bertozzi, A., Sapiro, G.: Navier-Stokes, fluid dynamics, and image and video inpainting. In: Proceedings IEEE Computer Vision and Pattern Recognition (CVPR), Hawaii, pp. 213–226 (2001)
2. Masnou, S., Morel, J.-M.: Level lines based disocclusion. In: Proceedings International Conference Image Process, pp. 259–263 (1998)
3. Perez, P.: Markov random fields and images. CWI Q. **11**(4), 413–437 (1998)
4. Chan, T.F., Shen, J.: Mathematical models for local non-texture inpainting. SIAM J. Appl. Math. **62**(3), 1019–1043 (2001)
5. Guleryuz, O.G.: Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising. Part I: Theory IEEE Transactions on Image Processing vol. 15, no. 3 (2006)
6. Criminisi, A., Perez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Trans. Image Process. **13**(9), 28–34 (2004)
7. Ružić, T., Pižurica, A.: Texture and color descriptors as a tool for context-aware patch-based image inpainting. In: Proceedings SPIE Electronic Imaging, vol. 8295 (2012)
8. Cao, F., Gousseau, Y., Masnou, S., Pérez, P.: Geometrically guided exemplar-based inpainting. SIAM J. Imaging Sci. **4**(4), 1143–1179 (2011)
9. Voronin, V.V., Marchuk, V.I., Petrosov, S.P., Svirin, I., Agaian, S., Egiazarian, K.: Image restoration using 2D autoregressive texture model and structure curve construction. In: Proceedings SPIE 9497, vol. 949706 (2015)

10. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process. **10**(2), 266–277 (2001)
11. Sethian, J.A.: Level Set Methods and Fast Marching Methods, 2nd edn. Cambridge University Press, Cambridge, UK (1999)
12. Frantc, V.A., Voronin, V.V., Marchuk, V.I., Sherstobitov, A.I., Agaian, S., Egiazarian, K.: Machine learning approach for objective inpainting quality assessment. Proc. SPIE **9120**, 91200S (2014)
13. Wexler, Y., Shechtman, E., Irani, M.: Space-time completion of video. IEEE Trans. Pattern Analysis Mach. Intell. (PAMI), 463–476 (2007)
14. Newson, A., Almansa, A., Fradet, M., Gousseau, Y., Pérez, P.: Towards fast, generic video inpainting. In: Proceedings of the 10th European Conference on Visual Media Production, pp. 7:1–7:8 (2013)

The System for the Study of the Dynamics of Human Emotional Response Using Fuzzy Trends

Natalya N. Filatova, Konstantin V. Sidorov, Sergey A. Terekhin and Gennady P. Vinogradov

Abstract The article discusses the features of a multi-channel system "EEG/S". The system has been used to build an interpretation model of certain characteristics of the emotions caused by visual, acoustic or olfactory stimuli. Monitoring of emotion characteristics changes is based on a periodic evaluation of attractors' properties. The attractors are reconstructed using EEG signals or a subject's speech. For this purpose we used the rules for increment evaluation of fuzzy characteristics of attractors by constructing an additional index scale. The scale allows generating a numerical evaluation for each element of a fuzzy set using a term index and a membership function.

Keywords Human emotions · Speech analysis · Attractor · Emotion interpreter · Fuzzy set · Fuzzy signs

1 Introduction

The objective evaluation of human emotional states is an important problem. It affects both efficiency and quality of communications as well as medical diagnostics. It is obvious that construction of interpretation models of nonverbal information transmitted by voice signals can have a huge impact on the progress of our civi-

N.N. Filatova (✉) · K.V. Sidorov · S.A. Terekhin · G.P. Vinogradov
Department of Information Technologies, Tver State Technical University, Afanasy Nikitin Street 22, Tver 170026, Russia
e-mail: nfilatova99@mail.ru

K.V. Sidorov
e-mail: bmisidorov@mail.ru

S.A. Terekhin
e-mail: rabeenovich69@mail.ru

G.P. Vinogradov
e-mail: wgp272ng@mail.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_18

lization in the context of improving communication not only between individuals, but also between humans and members of other species in the biosphere.

Currently, is possible to recognize human states by voice if these states are characterized by heightened tension (anxiety) or euphoria. There are a lot of different software tools to solve the problem of recognizing an emotion sign based on the voice messages analysis. However, the accuracy of such estimates decreases with decreasing emotions intensity. Existing databases with emotional speech (ES) example collections are not able to solve the problem of identifying the emotional development with the time. The problems of forecasting the changes in the direction of emotional development (change of the emotion sign) and the rate of emotional development also remain unsolved.

The paper [1] proposes an interpretation model of three main characteristics of emotion (a sign, a level and a direction of progress); it is specified by the rules of assessment of a sign and a level of emotional response [2]. The rules are based on experimentally established correlation between emotion characteristics and morphology of attractors which are reconstructed from the samples of speech or EEG signals. The demonstration of the same patterns in attractors, which are reconstructed according to various bioelectric signals and recorded in different channels in parallel, experimentally confirms the validity of using a dedicated set of quantitative features to interpret the valence and the level of emotion according to speech patterns.

The report focuses on centered on the research results that allow formulating the rules for assessing emotion dynamics. They were also a base for creating an algorithm for estimating increments of fuzzy features that characterize the geometry of EEG or speech attractors. All the results are obtained experimentally using a special automated system "EEG/S" (electroencephalogram/speech signal) [3].

2 The Technical and Methodological Support for Experiments

2.1 The Features of Technical Support

The "EEG/S" (electroencephalogram/speech signal) multi-channel system is designed to investigate human emotional responses to external stimuli [1, 2]. The structure of the system allows using from one to three channels for to present stimuli and two channels to record the reactions of the subject. The output channels are divided into as follows: a working channel for recording acoustic signals (a speech of a subject); a control channel for recording electrical activity of a brain (EEG). Such structure allows analyzing examples of speech signals, which are registered only after confirming a change in the emotional state of the subject. Changing the state of the subject objectively is confirmed by using EEG signals. The "EEG/S" system is able to identify weak manifestations of human emotional responses to external stimuli.

2.2 The Features of an Experimental Technique

A testee uses an acoustic, visual or olfactory analyzer to percept stimuli. The experimental technique contains some restrictions:

- voice signal recording (X(t)) is continuous throughout the experiment, alongside with EEG signals;
- X(t) recording should include a complete experimental record (a control text that includes reactions of a testee, comments with stimuli self-assessment, an audio track containing external stimuli (for visual or audio channels), and start and end markers of olfactory stimuli);
- the final selection of the channel for presentation of external stimuli should be based on the results of a preliminary assessment of an analyzer sensitivity of a testee;
- the reactions of a testee should be recorded using control messages that do not have to contain a particular meaning, but they should include the basic vowel phonemes of the Russian language; the duration of the control message is 20,000 samples;
- the duration of the experiment should not exceed 20–30 min; otherwise a testee can experience unanticipated negative emotions such as fatigue or boredom.

In addition to visual and acoustic stimuli, the "EEG/S" system includes a channel for olfactory stimuli presentation. In accordance with the stereochemical theory of olfaction [4], a person has a seven-component system of odor recognition that is based on the distinction of seven primary odors: musk, camphor, floral, ethereal, minty, pungent and putrefactive. It is known that the part of the brain, which analyzes the pulses from the receiver cells in a nose, is closely connected with the limbic system, the so-called primitive brain or an "olfactory brain". This part of the brain regulates emotions, mood and memory [5]. The experiments used essential oils. These oils were added to a water chamber of an aromatizer-humidifier (oil burner) or used to moisten a woolen cloth. An oil burner intensifies an odor by heating the solution. It was used to present weak flavors (e.g. tangerine, lavender, etc.), and a woolen cloth was used to present strong flavors (e.g. lemon etc.). The source of smell was located at a distance of 20 cm from the olfactory analyzer, alongside of a testee (so as not to fall into his field of view). The time of stimulus presentation varied. Experiments with olfactory stimuli alternated with a background phase which is necessary for restoring a neutral state of an olfactory analyzer.

This channel proved its efficiency when constructing complex (integrated) stimuli. Combining non-verbal stimuli (music and smell) it is possible to minimize cognitive processes (which appear when using visual information) and to enhance the emotional response at the same time.

3 The Study of the Dynamics of Emotions

The aim of a bioelectric signal (BES) analysis was an objective evidence of changes in the emotional state of the testee. These signals have been recorded using a control channel. In previous studies [6] the authors noted the changes in the power spectrum of EEG signals during the transition of emotions from positive to negative. On the basis of this information the "EEG/S" system pre-localized a recording interval which showed the changes in the emotional state of the testee.

The studies have confirmed the previously noted conclusion which stated that to clarify the boundaries of the time interval with emotions of one sign, it is not appropriate to use power spectra and corresponding linear transformations of signals. It would be more effective to use methods of nonlinear dynamics [7]. In particular, the authors use the idea of multidimensional phase portrait reconstruction, which shows the phase space of a dynamic biosystem [8].

The "EEG/S" system has attractors reconstruction software for a selected part of a bioelectrical signal (X(t)). Emotion and morphological features of attractors' projections are connected on the basis of experimental results. It is shown that there are three types of signs to interpret the sign and level of emotions: the length of maximum vectors R describing the contours of two-dimensional projections of attractors; γ density of the trajectories in the vicinity of the center of the two-dimensional projections of attractors and the sparsity degree of a density matrix [9].

The paper [1] presents an emotion interpreter model which includes three components:

$$\boldsymbol{Int\_em} = \langle \boldsymbol{Z}, \boldsymbol{U}, \boldsymbol{D} \rangle \tag{1}$$

where Z is the sign of emotions (positive/negative), U is the level of emotions (low, medium, high), D is the emotional response dynamics (emotion amplifies/attenuates).

To assess literals (1) we created special rules that use a BES pattern model, which is based on the reconstruction of the attractor following the model of speech or EEG pattern:

$$\boldsymbol{Mod\_sig} = \langle \bigcup_{j} (\boldsymbol{f}(\boldsymbol{R_i}), \boldsymbol{M\gamma}, \boldsymbol{\gamma s}, \boldsymbol{k_{0\_i}}) \rangle \tag{2}$$

where R_i is the length of the vector recovered from the origin of coordinates to the i-th point of the attractor, $f(R_i)$ represents linear transformations, j is the number of projections of the attractor, $M\gamma$ is a "density" matrix with the points of attractor's trajectories, k_0 is the number of zero cells in $M\gamma$.

Given that the nature of the attractor's trajectories depends on the time series plots selected for its reconstruction, the analysis of X(t) includes a sliding window which saves chaos signs in the analyzed signal. The attractor reconstruction is carried out for each window, as well as determination of its characteristics. Average

values of the characteristics over all calculation windows are integral estimates for the entire BES.

Taking into account the size of the control phrase, voice recording parameters and dimensions of the calculation window (L = 20,000 samples), no more than 3 calculation windows was enough to cover the entire speech pattern.

Let us consider the results of experiments in studying the development of a positive emotional reaction to periodic external stimuli. The paper [3] describes the methods of experiments.

The density estimates in the center of the matrix $M\gamma$ according to the second calculation window 2_PO demonstrate the greatest sensitivity to external stimuli (Fig. 1). We highlighted the plots graphs with coincidence of trends for all three calculation window. It means matching in the reactions of the testee to the second and the third positive stimuli.

The detailed analysis of the results of the second calculation window (2_PO) shows that for all speech patterns registered after the presentation of positive stimuli (4–9) the density estimates are smaller than that for patterns after neutral stimuli (Fig. 2). If we highlight the range of the density for the neutral state, then all points with positive stimuli will be below this area.

The points showing the response of the testee to one stimulus are connected by a one solid line of a corresponding color. Transitions from the tj-th time state to the (tj + del)-th time state where del is the time of the next stimulus presentation are shown in dashed lines. The graph (Fig. 2) illustrates the possibility of using "density" feature to monitor the level of emotions while precepting external stimuli.

To analyze the development of emotional reactions over time (Fig. 3) it is necessary to take into account a weak reproduction of quantitative estimates of characteristics of the attractors reconstructed from speech signals, as well as their dependence on the type of phonemes.

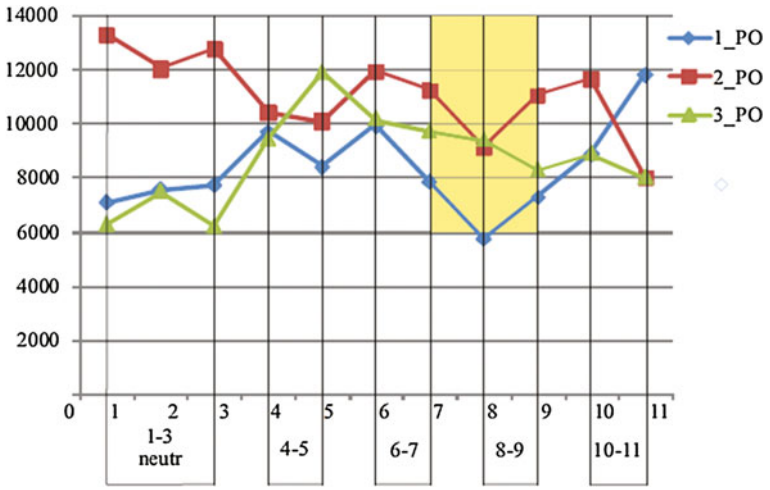


Fig. 1 Changes in density during experiment using two calculation windows

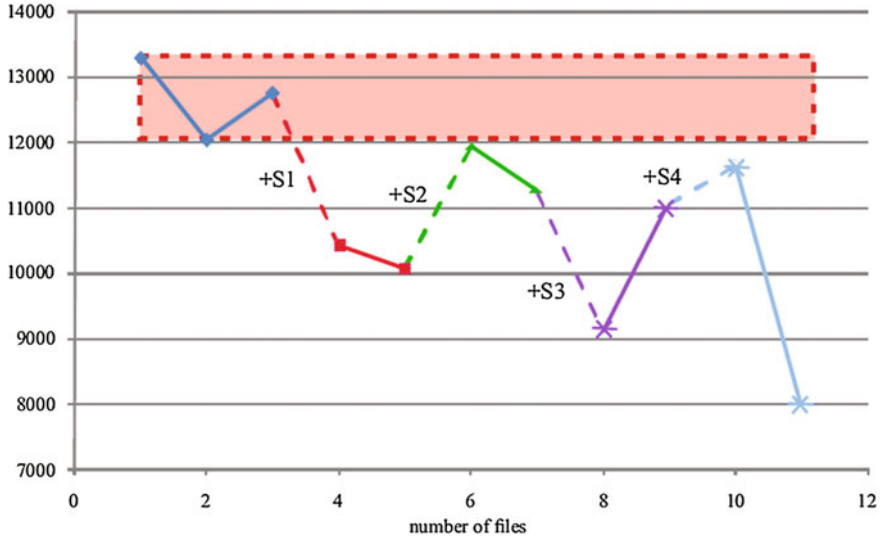


Fig. 2 Changes in the density of the settlement in the second window (20 000–40 000) during the experiment with the presentation of positive incentives (+S1, …, +S4)

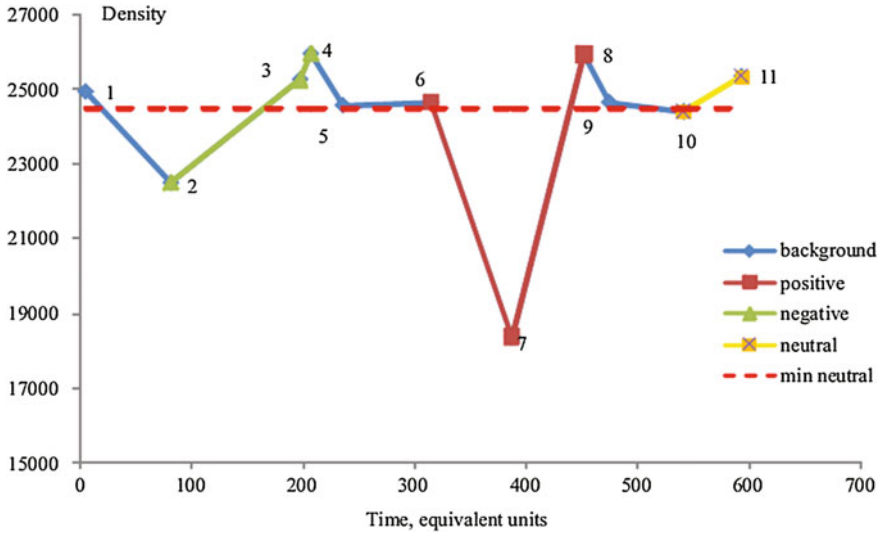


Fig. 3 Change the central density in the course of the experiment, taking into account the duration of the perception of emotional significance of an external stimulus

Multiple independent external and internal factors affect a subjective perception of information by a testee. Therefore, it is necessary to transit to fuzzy features. The phasification of model parameters (2) made it possible to create a model for interpreting a sign and a level of emotions [2].

4 The Assessment of Changing Development of Emotions

The problem of determining the direction of emotions can be viewed as the task of forming a trend for the dependency type γ(t) and $k_0(t)$.

Procedure fuzzification γ(t) and k0(t) leads to ambiguities in the determination of ordinates graphs (Figs. 2, 3). In practice, we get version of fuzzy time series. In order to identify a trend the authors suggest using a fuzzy segment matrix, which was proposed in [10] for one of the tasks of medical diagnostics.

Let us consider this method on the example of a γ(t) trend analysis. Coordinate values along the time axis t for all points of graphs γ(t) are accurate and correspond to the times of recording BES patterns (for determining γ(t) values). Values on the ordinate axis are represented using a fuzzy variable, i.e. as the element of two fuzzy sets T_j and T_{j+1}:

$$\boldsymbol{\mu_j/Y_i \in T_j, u\ \mu_{j+1}/Y_{i+1} \in T_{j+1}} \tag{3}$$

where Y_i is the assessment of the relevant characteristic of the attractor (γ(t) and $k_0(t)$) at t_i-th time interval of the base scale, T_j, T_{j+1} are elements of a term set (*TY*) for the linguistic variable PV (e.g., PV "The density of trajectories near the center of the attractor on the interval t_i", μ_j,μ_{j+1} are normalized membership functions of *TY* terms.

Applying the rule of fuzzy set phase-association for each point of the graph γ(t) we obtain a density estimate of the linguistic scale $\mu_j/Y_i \in T_j$.

In order to calculate the trend of emotions in the limited section of the graph γ(t) it is necessary to perform arithmetic operations on elements of fuzzy sets. Therefore, we propose to move from a term-set *TY* to a new "index" scale using the following transformation [10]:

$$\boldsymbol{W_j(\gamma_k) = ind_{T_i} + \mu_{T_{i+1}}(\gamma_k)} \tag{4}$$

where $W_j(\gamma_k)$ is a numeric representation of a density linguistic estimate, ind_{Ti} is the term index with a fuzzy set intersecting with the left extension of the fuzzy set corresponding to the term T_{i+1}, $\mu_{Ti+1}(\gamma_k)$—membership function of density (γk) of the fuzzy set T_{i+1}.

The comparison of (3) and (4) shows that the transition to the index scale in numerical representation of $W_j(\gamma_k)$ saves all information from a linguistic scale. The first fuzzy set is represented by its term index. The corresponding estimation of a membership function is determined by $\mu_i = 1 - \mu_{i+1}$. The second term is determined

by a simple increment of a parameter ind_{Ti}. In this case the membership function is the fractional part in the estimation $W_j(\gamma_k)$.

The resulting numeric set $\{W_j(\gamma_k)\}$ allows determining topological characterization of the time dependence of the density estimates of the attractor. Thus, we are able to create the rules for assessing the dynamics of emotions.

As the function is monotone, in order to assess the direction of the graph γ(t) segments we are able to use three terms, such as: "Descending (TO1)", "Smooth (TO2)", "Increasing (TO3)".

When using a base scale for assessing γ, we is determine the direction of each segment in the graph γ(t) in accordance with the sign of increment γ for the interval $(\gamma(t_{i+1}) - \gamma(t_i))$. For example, NP = {Increasing}, if $\gamma(t_{i+1}) - \gamma(t_i) > 0$.

Since the estimates γ(t) are considered as the elements of the corresponding fuzzy sets (*TY*), then in order to check the conditions (5) we use the estimates of the index scale:

$$\textit{If } \boldsymbol{W}(\gamma_i) - \boldsymbol{W}(\gamma_{i+1}) < 0, \textit{then NP=T0}_1 \tag{5}$$

$$\textit{If } \boldsymbol{W}(\gamma_i) - \boldsymbol{W}(\gamma_{i+1}) < 0, \textit{then NP=T0}_2 \tag{6}$$

$$\textit{If } \boldsymbol{W}(\gamma_i) - \boldsymbol{W}(\gamma_{i+1}) < 0, \textit{then NP} = \textit{T0}_3 \tag{7}$$

In order to determine a fuzzy set, which includes a segment, formulated rules (Table 1).

The segment corresponds to one of fuzzy sets according to (Table 1).

All segments of the graph γ(t) belong to fuzzy sets that define their slope. The membership recognition results of these segments can be presented as a square matrix *Mn*. Each element (m_{ij}) is the estimate of a fuzzy set for the segment which begins at t_i and ends at t_j timepoint (or m_{ij} changes in a slope of a segment). The matrix *Mn* presents many options of a fuzzy polygonal curve approximation that can be built on a set of points Y.

Table 1 The rules for calculating the values of the membership functions of fuzzy sets $(T0_j)$ of segments

Rule №	Estimation of increment (increment characteristics of the attractor) on the index'es scale	Membership function for a segment (μ_{T0})
Rule 1	$\boldsymbol{W}_{T0} \leq -1$	$\boldsymbol{\mu}_{T0_1} = 1$
Rule 2	$\boldsymbol{W}_{T0} > -1$	$\boldsymbol{\mu}_{T0_1} = \boldsymbol{W}_{T0}$
Rule 3	$\boldsymbol{W}_{T0} < 1$	$\boldsymbol{\mu}_{T0_3} = \boldsymbol{W}_{T0}$
Rule 4	$\boldsymbol{W}_{T0} \geq 1$	$\boldsymbol{\mu}_{T0_3} = 1$
Rule 5	$-1 \leq \boldsymbol{W}_{T0} < 0$	$\boldsymbol{\mu}_{T0_2} = 1 - \boldsymbol{\mu}_{T0_1}$
Rule 6	$0 < \boldsymbol{W}_{T0} \leq 1$	$\boldsymbol{\mu}_{T0_2} = 1 - \boldsymbol{\mu}_{T0_3}$
Rule 7	$\boldsymbol{W}_{T0_{1,3}} = 0$	$\boldsymbol{\mu}_{T0_2} = 0$

	1	2	3	4	5	6	7	8	9	10	11	
1		D	Inc	Inc	S	S	D	Inc	S	S	Inc	
2	D		**Inc**	Inc	Inc	Inc	D	Inc	Inc	Inc	Inc	
3	Inc	Inc		**Inc**	D	D	D	Inc	D	D	S	
4	Inc	Inc	Inc		**D**	D	D	S	D	D	S	
5	S	Inc	D	D		**D**	D	Inc	S	S	Inc	
6	S	Inc	D	D	S		**D**	Inc	S	S	Inc	
7	D	D	D	D	D	D		**Inc**	Inc	Inc	Inc	
8	Inc	Inc	Inc	S	Inc	Inc	Inc		**D**	D	D	
9	S	Inc	D	D	S	S	Inc	D		**S**	Inc	
10	S	Inc	D	D	S	S	Inc	D	S		**Inc**	graphic segments
11	Inc	Inc	S	S	Inc	Inc	Inc	D	Inc	Inc		

Fig. 4 Matrix changes in inclination of the segments for graphic γ (t), shown in Fig. 3

The interpretation of the analyzed graph is based on the elements ($m_{i,i+1}$) located to the right of the main diagonal line.

For the given example (Figs. 3, 4) the segments 2–3 and 3–4 belong to the fuzzy set $T0_3$ "Increasing". Therefore for the range from t_2 to t_4 there is a gradual increase in negative emotions (the membership function is defined by Rule 3, $\mu_{T03} = 0.7$). In accordance with the rules (Table 1) almost all the segments, which form the part of the diagram 4–5–6–7, are defined as the elements of the fuzzy set TO_1 "Decreasing".

Thus, in the interval from t4 to t7 there is a gradual increase of positive emotions ($\mu_{T01} = 0.9$).

The interpretation of the dynamics of emotions can be based on the analysis of the chart $k_0(t)$, however the characteristic of density estimate γ(t) is more sensitive to changes of emotion levels.

5 Conclusion

The experimental validation of the proposed model and the algorithm showed high accuracy of interpreting a sign and a level of emotions (according to EEG samples the error is 17 %, for the Russian speech samples the error is 4 %, for the German speech samples the error is 8 %).

The results of the evaluation of the dynamics of emotions are preliminary and will be elaborated in new experiments with the "EEG/S" system.

The proposed approach is useful for application in the sector of industry to monitor the state of operators in different technical systems.

Acknowledgments This work is supported in part by the Russian Foundation for Basic Research (project no. 14-01-00719 "Interpreter of Emotions Manifested in the Natural Speech").

References

1. Filatova, N.N., Sidorov, K.V.: An Interpretation Model of Emotions in Speech. In: Proceedings. of the Congress on Intelligent Systems and IT "IS&IT'13", vol. 4. Moscow, Fizmatlit Publ. (2013) (in Russ., Trudy Kongressa po intellektualnym sistemam i informatsionnym tekhnologiyam "IS&IT'13", vol. 1, pp. 98–105. (2013))
2. Filatova, N.N., Sidorov, K.V.: An Assessment Model of Some Emotional Features Based on Analysis of Speech Samples or EEG Signals. In: Integrated Models and Soft Computing in Artificial Intelligence. Proceedings. of the 8th International Science and Techology Conference, vol. 2 Moscow, Fizmatlit Publ. (2015) (in Russ., Sbornik nauchnykh trudov VIII-y mezhdunar. NTK Integrirovannye modeli i myagkie vychisleniya v iskusstvennom intellekte, vol. 2, pp. 727–737. (2015))
3. Filatova, N.N., Terekhin, S.A.: Bioengineering System for Research on Human Emotional Response to External Stimuli. In: IEEE: International Conference on Biomedical Engineering and Computational Technologies (SIBIRCON), pp. 13–17. Novosibirsk, Russia, 28–30 Oct 2015 (2015). doi:10.1109/SIBIRCON.2015.7361841
4. Amoore, J.E., Johnston, Jr. J.W., Rubin, M.: The Stereochemical Theory of Odor. Scientific American, vol. 210, no. 2, pp. 42–49 (1964)
5. Baars, B.J., Gage, N.M.: Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience, 2nd edn. Elsevier, pp. 653 (2010)
6. Lapshina, T.N.: EEG correlates of emotion. Int. J. Psychophysiol. **61**(3), 328 (2006)
7. Mekler, A.A.: Calculation of EEG correlation dimension: large massifs of experimental data. Comput. Methods Programs Biomed. **92**(1), 154–160 (2008)
8. Takens, F.:. Detecting strange attractors in turbulence. In: Dynamical Systems and Turbulence. Springer, Heidelberg, pp. 366–381 (1981)
9. Filatova, N.N., Khaneev, D.M.: Use of neurolike structures for automatic generation of hypotheses for classification rules. Fuzzy Syst. Soft Comput. **8**(1), pp. 27–44. (2013) (in Russ., Nechetkie sistemy i mjagkie vychislenija)
10. Filatova, N.N., Milovidov, A.A.: Analysis and interpretation of audiograms in the space of fuzzy attributes. Fuzzy Syst. Soft Comput. **6**(1), pp. 55–66 (2011) (in Russ., Nechetkie sistemy i mjagkie vychislenija)

Part IV
Hybrid Expert Systems and Intelligent Decision Support Systems in Design and Engineering

Intelligent Technology for Integrated Expert Systems Construction

Galina V. Rybina, Victor M. Rybin, Yury M. Blokhin and Elena S. Sergienko

Abstract Development of the integrated expert systems with the problem-oriented methodology and the problems of the intellectualization of AT-TECHNOLOGY workbench are reviewed. Intelligent technology for integrated expert systems construction is described, in particular for dynamic integrated expert systems construction. Intelligent planning methods applied for integrated expert systems development plan generation are described with usage of the intelligent planner, reusable components, typical design procedures, and other intelligent program environment components. Discussed technology was used for building tutoring IES prototypes for different courses/disciplines and two industrial dynamic IES prototypes: "Management of medical forces and resources for major traffic accidents" and "Resource management for satellite communications system between regional centers".

Keywords Artificial intelligence · Integrated expert systems · Problem-oriented methodology · AT-TECHNOLOGY workbench · Intelligent software environment · Automated planning · Automated planner

1 Introduction

Trends towards integration of research in different fields of artificial intelligence had most clearly manifested at the turn of the XX and the XXI centuries and made it necessary to combine semantically different objects, models, methodologies, concepts and technologies. It created new classes of problems and new architectures of intelligent systems. The systems based on knowledge or expert systems (ES) (they were originally intended to support the decision unformalized problems (UF-problems)), are an essential part of a significant number of static and dynamic

G.V. Rybina (✉) · V.M. Rybin · Y.M. Blokhin · E.S. Sergienko
National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Kashiskoe Sh. 31, 115409 Moscow, Russian Federation
e-mail: galina@ailab.mephi.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_19

intelligent systems, as the analysis of development experience [6, 13, 15–17, 25] has shown. At the same time, in conjunction with ES, methods of applied mathematics, soft computing and versatile means of control data are used for solving tasks with practical complexity and importance. It leads to the necessity of combining ES into a single architecture with such diverse components as: a database comprising engineering, production and management information; software packages with the developed calculating, modeling and graphics tools; educating systems, including built on technology training and learning systems; simulation modeling systems for dynamic applications, etc.

A new class of intelligent systems emerged—integrated expert systems (ES). Their scalable architecture is represented in the form of $ES + K$. Here K is some software that provides decision support for the formalized part of the problem, and function of finding solutions to the UF-tasks relevant to the components of the poorly structured problem falls on the ES.

To address IES as an independent object of research, the IES terminological basis was formed and the basic definition of the IES as a systematically organized set of components ES and K was introduced. Their structure and nature of the interaction determine the type of the IES relevant and solvable problems—its functionality. The concept of "static IES" and "dynamic IES" was also introduced [13]. Automated problem-oriented methodology [13] was created for the construction of IES, with powerful functionality and scalable architecture. It is actively used and constantly develops. The core idea is based on the conceptual modeling of the IES architecture at all levels of the integration processes in the IES and focusing on the modeling specific types of UF-tasks that are relevant to the technology of traditional ES. In the laboratory "Intelligent Systems and Technologies" Department of Cybernetics National Research Nuclear University MEPhI several generations of software tools for problem-oriented methodology were created, details can be found in [13, 15–17, 24]. They are united under intelligent program environment of AT-TECHNOLOGY workbench.

It is necessary to point out that in the current work we are focused on two IES classes: dynamical and tutoring integrated expert systems. Dynamical IES in context of problem-oriented methodology of IES construction are expansion of static IES connected with coming to executing in dynamical problem areas and dynamical problems solving (monitoring, diagnosys, planning, control and others). Thus, in dynamical IES architecture special components modelling external environment and allowing to interact with real-world hardware and perform reasoning in conditions of frequently changing data and knowledge.

A large part of the problems is linked to the high complexity phases of design and implementation of the IES, as showed by practical experience of creating a series of static, dynamic, and educating IES through the use of problem-oriented methodology and AT-TECHNOLOGY workbench [13–17]. Specifics of a particular area of concern and the human factor provide a significant impact. Therefore, the need to develop intelligent software environment and its basic component—intelligent planner [13, 15] for the further development of problem-oriented

methodology and AT-TECHNOLOGY workbench with the aim of creating intelligent technology to build specific classes of IES has become urgent.

To date, multiple versions of intelligent planners for the AT-TECHNOLOGY workbench were created. They were developed by combining predictive models and methods of planning methods used in IES, details are in [13, 15, 19–22].

This paper is focused on a new research phase connected with further development of intelligent planner and other components of the intelligent program environment. The purpose for the new research is to increase the degree of automatic (intellectualization) planning and project management for the creation of a broad class of IES.

2 Some Aspects of Current Research in the Field of Automated Planning

Today automatic generation of plans by a software and hardware system is often meant as the intelligent planning, but the term intelligent planning has no clear definition. For example, the term intelligent planning is more commonly used in Russian literature [10], while automated planning is used in English [8]. However, in both cases the plan generation process done by the computer is meant. Accordingly, plan is a glimpse of future behavior in the context of intelligent planning. In particular, the plan is usually a set of steps with some restrictions (e.g., temporal) for the execution of some agent or agents [8].

Software system that generates plans based on a formal description of the environment, the initial state of the environment and assigned to the planner purpose is meant under planner in practical application. In some works, planner is called the agent [10]; in other cases, the planner also contains the designer carrying out the allocation of resources for the implementation of building plans.

A significant number of methods, approaches, formalisms, etc. was developed by now in the field of intelligent planning. Among them should be highlighted: planning with propositional logic; planning in a space plans; planning in space of conditions; planning as constraint satisfaction problems; planning on the basis of precedents; broadcast to the other problem; temporal planning; planning in non-deterministic and probabilistic areas; hierarchical planning (HTN-formalism), and others. Detailed reviews can be found in [8, 10], and others, as well as in the works of authors [19–21].

Methods of the intelligent planning are widely used in a number of applications. The most popular fields of intelligent planning applications are: management of autonomous robots [8, 26]; semantic web and web services composition [27]; computer-aided learning (in particular for the construction of individual education plans [4, 22]); calibration of equipment [11]; control of conveyor machines [3]; resource-scheduling [2]; resource allocation in computing systems [7], logistics [5], the automation of software development [9] and others.

PDDL (Planning Domain Definition Language) developed under the McDemorts leadership is the current norm in the field of the intelligent planning with the three main versions. We can detect the following trends in the context of development and use of various intelligent planning formalisms and related languages: the unification and standardization of planning languages on the basis of PDDL; the emergence of task-specific planning languages; a shift in emphasis towards nondeterministic research planning (RDDL languages and PPDDL); the emergence of complex mechanisms to meet the preferences and limitations, and others.

However, in general, application of intelligent planning for the automated support processes of building intelligent systems is a poorly investigated area, and it is possible here to refer mainly to the experience gained in the creation of applied IES based on the problem-oriented methodology and AT-TECHNOLOGY workbench, in particular, the development and use of educational and dynamic IES [18, 22, 23]. Let us consider the basic concepts of intellectual AT-TECHNOLOGY workbench software in more detail.

3 Model of Intelligent Software Environment and Its Components

Significant place in the framework of the problem-oriented IES constructing methodology (basic points are reflected in [13]) is given to the methods and means of intelligent software support for the development processes. It is general concept of "intelligent environment". Complete formal description of the intellectual environment model and methods of the individual components implementation is presented in [13], so here only a brief description of the model in the form of quaternion is presented.

$$M_{AT} = \langle KB, K, P, TI \rangle \tag{1}$$

KB is a technological knowledge base (KB) on the composition of the project, and typical design solutions used in development of IES. $K = \{K_i\}$, $i = 1...m$—set of current contexts K_i, consisting of a set of objects from the KB, editing or implementing on the current control step. P a special program—an intelligent planner that manages the development and IES testing process. $TI = \{TI_i\}$, $i = 1...n$—many tools TI_i, applied at various stages of IES development.

A component of the KB is a declarative basis of intellectual support for the development of IES, acting as data storage in a given environment and defined as

$$KB = \langle WKB, CKB, PKB \rangle \tag{2}$$

WKB is a KB containing knowledge of the standard design procedures (SDP), describing the sequences and methods of using various tools to create applied IES and a sequence of steps for creating IES. *CKB*—is KB comprising knowledge about the use of SDP and re-used components (RUC), including fragments of previously created IES prototypes. PKB (optional)—is a KB containing specific knowledge used at various stages of creating IES prototype for solving problems that require innovative approaches. The current context K_i is represented as set of $K_i = \langle KD, KP \rangle$. *KD* here is a declarative context for storing static declarative information about the structure of the project, the knowledge engineer and the current user. *KP* is a procedural context, which includes objects clearly affecting the further planner steps (LC system phase, currently edited or executable object, the current target, the current executor, the global development plan, etc.).

The main procedural (operational) component is intelligent planner. This model generally describes it.

$$P = \langle SK, AF, Pa, Pb, I, GP \rangle \tag{3}$$

SK here is the state of the current context, in which the scheduler was activated. $AF = \{AF_i\}$, $i = 1 \ldots k$ is a set of functional modules AF_i, a part of planner. Pa is a selection procedure for the current target based on the global development plan. *Pb* is a selection procedure for the best executive function module from the list of possible candidates. *I*—procedures to ensure the interface with the corresponding components of the AT-TECHNOLOGY workbench; *GP*—operating procedures for the IES global development plan.

Any SDP can be represented as triples

$$TPP_i = \langle C, L, T \rangle \tag{4}$$

where *C*—is the set of conditions under which the SDP can be implemented; *L*—script implementation described in the describing internal language actions of the SDP; *T*—set of parameters initialized by intelligent planner at SDP inclusion in the development plan of a IES prototype. Each RUC, involved in the development of an IES prototype, is represented as a tuple:

$$RUC = \langle N, Arg, F, PINT, FN \rangle \tag{5}$$

N in this model is the name of the component, by which it is registered in the complex. $Arg = \{Arg_i\}$, $i = 1 \ldots l$—set of arguments containing current project database subtree serving the input parameters for the functions from the set. $F = \{F_i\}, i = 1 \ldots s$—a variety of methods (RUC interfaces) for this component at the implementation level. *PINT*—a set of other kinds of RUC interfaces, used by the methods of the RUC. $FN = \{FNi\}$, $i = 1 \ldots \upsilon$—set of functions names performed by this RUC. Prototypes IES model of the functioning process is represented as tuple:

$$MA = \langle Sc, C, Cl \rangle \tag{6}$$

Sc is a scenario of IES prototype work. *C*—set of IES prototype subsystems that can be divided into 2 categories (standard subsystem (RUC) from the RUC repository and prototype's IES subsystem implemented by developers). *Cl* is a ratio of "control transfer", describing the procedure for interaction between the IES prototype's subsystems.

Let us briefly examine the methods and approaches used in the implementation of the intellectual supportive environment for the development of IES model. The main components of this IES are the technological KB on the composition of IES project, SDP and RUC, and the intelligent planner managing the process of plans construction and implementation for the development of IES prototypes. These are the main purposes why it is necessary to use different types of knowledge in the process of developing a IES prototype: checking referential integrity of the project on the development of IES; automated construction of components diagrams; layout synthesis of IES prototype architecture; planning a series of steps to create a prototype of IES-specific features and tasks; determining a set of the most relevant sub-tasks for each of the stages (steps) in the development of IES prototype and others.

4 IES Prototype Project and Its Components

The task of creating a plan for developing an applied IES prototype is a complete task from the field of artificial intelligence, because it requires involvement of a variety of knowledge about models and methods for solving typical problems [13], a technology on design and development of IES, on how to integrate with external databases, packaged applications and other programs. Therefore, a project to develop IES-based problem-oriented methodology and data on the problem being solved is stored in some format on physical carrier body of knowledge. Based on this data the intelligent planner is running the process of prototyping for applied IES.

It should be noted that the implementation of the current intelligent planner version [20] is a hybridization of approaches based on the use of HTN-formalism and flexible mechanisms to find solutions used in the IES. It allows the use of a declarative way of describing knowledge of the development, in this case the production-type knowledge representation language [13].

The implementation of managing the IES prototype development is a basic process, and requires the use of certain types of knowledge and use of intelligent planner. A brief description of some of the basic models for this process in accordance with [13] is given.

The model of the IES prototype is presented as a tuple:

$$PRJ = \langle PN, KB, Solver, Pd, PDFD, PPIK, PCOMP \rangle \quad (7)$$

PN—project name, *KB*—IES prototypes KB, *Solver*—output machine (agents) for the IES prototype. *PD*—project data, i.e. the different types of information (knowledge, data, individual parameters, text, etc.), used by the intelligent planner in the process of developing a IES prototype and to generate the finished prototype, wherein the main project data include records of interviewing experts, lexical dictionary, fragments of the knowledge field, KB in the different knowledge languages representation, the type of problem being solved, as well as various service information (the profile of the current user, the name of knowledge engineer who created the project, the date of commencement and the anticipated completion, etc.). *PDFD*—enhanced information and logical model of the IES prototype architecture presented as a hierarchy of advanced data flow diagram (ADFD), which is one of the most important components of the project because its structure largely determines the composition of the prototype and its functionality (hierarchy ADFD built using just a single relationship—decomposition, i.e. the upper levels operation is detailed using ADFD of the lower level hierarchy, and it is possible that only one ADFD that is not part of the operation exists, such as "contextual data flow diagram"—a set of PIC). $PCOMP = \{PCOM\,P_i\}$, $i = 1..problem - orientedmethodolog.l$ is the collection of various IES subsystems developed by means of AT-TECHNOLOGY workbench and external applications.

Basing on described models and current project state a plan of IES prototyping is generated with use of automated planning methods. Plan can be decomposed into global and detailed. Now some details about plan generation and automated planning usage will be discussed.

Thus, the main task of intelligent planner is a dynamic support knowledge engineer operations at all life cycle stages of building prototypes as shown in Fig. 1. Dynamic support is done by generating IES development plans for the current IES

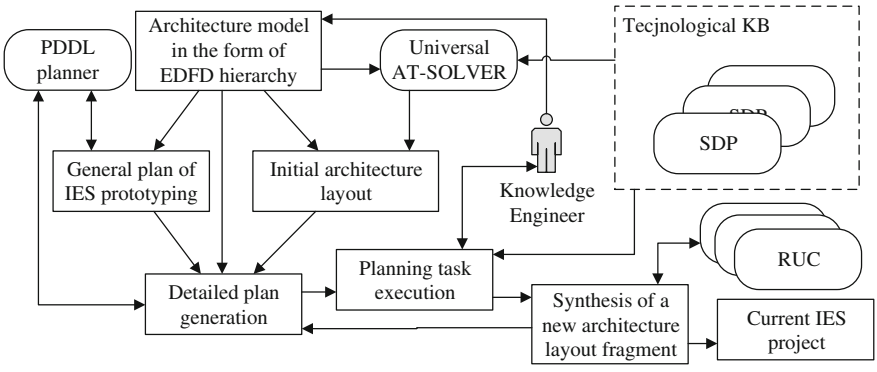


Fig. 1 Planning scheme

prototypes and allowing the specific plans execution (made either automatically or interactively). It should be noted that detailed plans and global IES prototyping generation and architecture model synthesis is based on the integration of IES with planning methods.

5 Planning of IES Prototyping Steps

Planning is performed in a projects state space [8] (in the early studies, the concept of design space was used). State space is formed by a plurality of project parameters possible values. This approach was chosen because of its popularity and the large number of tools that can be effectively integrated into AT-T-ECHNOLOGY workbench intelligent planner.

Planning to build an IES prototype is a complete project development task, and it is a reason why elements of the plan have to be tied to the time. According to this [1] temporal approach with an explicit modeling of time was chosen, as it is planned to adapt the intelligent planner under the team development management in the further research. It is also another argument in favor of temporal planning. Described above formalism and language PDDL is used as a basis for combining these approaches. It allows taking into account time and using real variables.

Methods of automated planning are used to solve described above problems with help of intelligent planner. These methods are deterministic classical planning in state space model with the trivial time (to generate a global plan) and temporal planning with the action cost (to generate a detailed plan). These methods are implemented with the help of external programs—PDDL-planners, and the integration is based on file sharing. For PDDL-planners the task is created in the PDDL language version 2.1 [1] with additional enhancements that are further addressed in a separate process.

Separation to the global and detailed plan allows effective optimization to the intelligent planner operations, retaining its response time, as the number of states even for a small planning problem can be quite high [12]. The task of planning for PDDL-planner must be submitted in the form of quaternary.

For example let us describe planning connected with fragment of SDP for supporting Combined Knowledge Acquisition Method [13, 15].

In the Fig. 2 a fragment of projects state diagram is shown. The project itself contains a lot of sub state, formed by its components. The example of such sub-component—knowledge base is show in the same figure. The knowledge base has a few states, and many possible transitions. Each transition is linked with some action of knowledge engineer. Such actions are planned with automated planning technology.

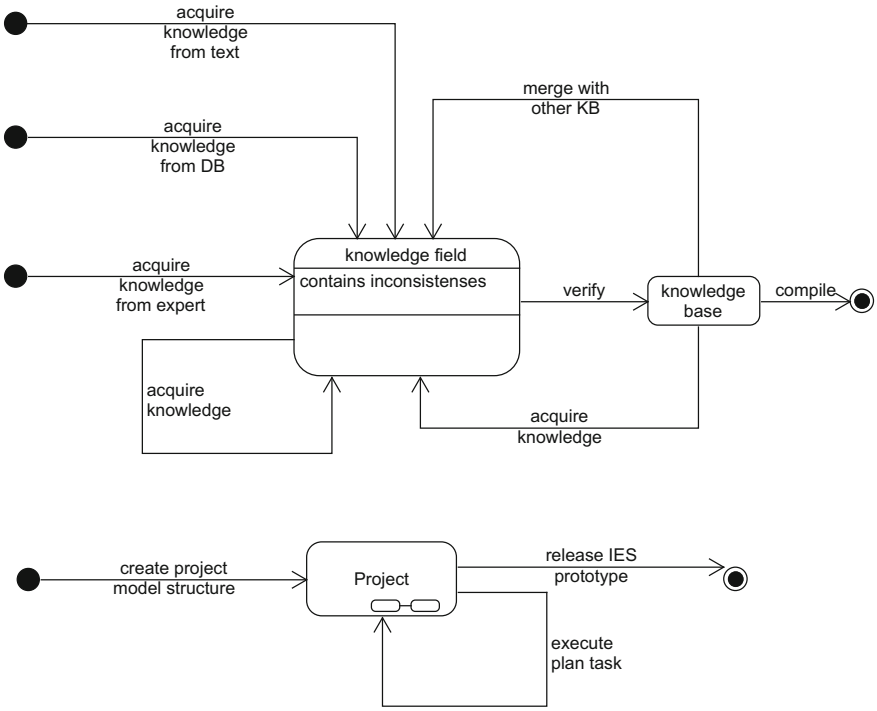


Fig. 2 A fragment of UML state diagramm of IES prototype project

It is necessary to point out, that described domain is dynamic, because planning is not performed offline and the planner knows about execution results. Thus, the plan is generated each time some planning task completed.

6 Conclusion

Currently, an experimental software study of the current version of the intelligent planner during educational IES prototyping on various courses is carried out. This study is carried out in particular, for the collective development of IES prototypes with limited resources. Research and development related to the use of intelligent software environment for building software applications for two dynamic IES prototypes (“Management of medical forces and resources for major traffic accidents” and “Resource management for satellite communications system between regional centers”) and tutoring IES prototypes for different courses/disciplines are presented.The work was supported by the Russian Foundation for Basic Research support (project â„– 15-01-04696).

References

1. Benton, J., Coles, A.J., Coles, A.: Temporal planning with preferences and time-dependent continuous costs. In: Proceedings of the Twenty-Second International Conference on Automated Planning and Scheduling, ICAPS 2012, Atibaia, São Paulo, Brazil (2012). 25–19 June 2012
2. Burkov, V., Korgin, N., Novikov, D.: Theory of Organizational System Management Introduction. Librokom, Mosow (2009)
3. Do, M.B., Lee, L., Zhou, R., Crawford, L.S., Uckun, S.: Online planning to control a packaging infeed system. In: Proceedings of the Twenty-Third Conference on Innovative Applications of Artificial Intelligence, San Francisco, California, USA (2011). 9–11 Aug 2011
4. Garrido, A., Morales, L., Serina, I.: Using AI planning to enhance e-learning processes. In: Proceedings of the Twenty-Second International Conference on Automated Planning and Scheduling, ICAPS 2012, Atibaia, São Paulo, Brazil (2012). 25–19 June 2012
5. Gerevini, A., Haslum, P., Long, D., Saetti, A., Dimopoulos, Y.: Deterministic planning in the fifth international planning competition: PDDL3 and experimental evaluation of the planners. Artif. Intell. **173**(5–6), 619–668 (2009). doi:10.1016/j.artint.2008.10.012
6. Grosan, C., Abraham, A.: Intelligent Systems—A Modern Approach, Intelligent Systems Reference Library, vol. 17. Springer (2011). http://dx.doi.org/10.1007/978-3-642-21004-4
7. Kolesov, N., Tolmacheva, M., Yuhta, P.: Computation planning timecost and complexity assesment in distributed systems. In: Proceedings of 4th Russian multiconference, Taganrok. pp. 312–314. TTI YFU, Taganrok (2011)
8. Nau, D.S.: Current trends in automated planning. AI Magazine **28**(4), 43–58 (2007)
9. Novoseltsev, V., Pinzjin, A.: Implementation of efficient algorithm of linear functional programm synthesis. Bull. Tomsk Polytechn. Univ. **312**(5), 32–35 (2008)
10. Osipov, G.S.: Artificial Intelligence Methods. Fizmatlit, Moscow (2011)
11. Parkinson, S., Longstaff, A., Crampton, A., Gregory, P.: The application of automated planning to machine tool calibration. In: Proceedings of the Twenty-Second International Conference on Automated Planning and Scheduling, ICAPS 2012, Atibaia, São Paulo, Brazil (2012). 25–19 June 2012
12. Russell, S.J., Norvig, P.: Artificial Intelligence—A Modern Approach, 3rd edn. Pearson Education (2010)
13. Rybina, G.V.: Theory and Technology of Construction of Integrated Expert Systems. Monography. Nauchtehlitizdat, Moscow (2008)
14. Rybina, G.V.: Practical use of problem-oriented methodology of integrated expert system construction (a review of applications in static and dynamic problem fields). Devices and Systems. Control, monitoring, diagnostics **12**, 10–28 (2011)
15. Rybina, G.V.: Intelligent Systems: from A to Z. Monography Series in 3 books, vol. 1. Knowledge-based systems. Integrated expert systems. Nauchtehlitizdat, Moscow (2014)
16. Rybina, G.V.: Intelligent Systems: from A to Z. Monography Series in 3 books, vol. 2. Intelligent dialogue systems. Dynamic intelligent systems. Nauchtehlitizdat, Moscow (2015)
17. Rybina, G.V.: Intelligent Systems: from A to Z. Monography Series in 3 books, vol. 3. Problem-oriented intelligent systems. Tools for intelligent system developing. Dynamic intelligent systems. Nauchtehlitizdat, Moscow (2015)
18. Rybina, G.V., Blokhin, Y.M.: Distributed knowledge acquisition control with use of the intelligent program environment of the at-technology workbench. Commun. Comput. Inf. Sci. 150–159 (2014)
19. Rybina, G.V., Blokhin, Y.M.: Use of intelligent program environment for construction of integrated expert systems. Life Sci. J. **11**(8), 287–291 (2014)
20. Rybina, G.V., Blokhin, Y.M.: Modern automated planning methods and tools and their use for control of process of integrated expert systems construction. Artif. Intell. Decis. Mak. **1**, 75–93 (2015)

21. Rybina, G.V., Blokhin, Y.M., Danyakin, I.D.: Intelligent technology for integrated expert system construction. Inf. Meas. Control Syst. **1**, 3–19 (2015)
22. Rybina, G.V., Blokhin, Y.M., Ivashenko, M.G.: Some aspects of intelligent technology for integrated expert system construction. Devices Syst. Control Monit. Diagn. **4**, 27–36 (2013)
23. Rybina, G.V., Mozgachev, A.V.: The use of temporal inferences in dynamic integrated expert systems. Sci. Techn. Inf. Process. **41**(6), 390–399 (2008)
24. Rybina, G.V., Mozgachev, A.V.: At-technology workbench for integrated expert systems construction support: common description and developing prospects. Devices Syst. Control Monit. Diagn. **11**, 17–40 (2011)
25. Tyler, A.R.: Expert Systems Research Trends. Nova Science Publishers, New York (2007)
26. Yakovlev, K., Baskin, E.: Graph models for solving 2d path finding problems. Artif. Intell. Decis. Mak. **1**, 5–12 (2013)
27. Zou, G., Chen, Y., Xu, Y., Huang, R., Xiang, Y.: Towards automated choreographing of web services using planning. In: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, Ontario, Canada. (2012). 22–26 July 2012

Multi-method Technology for Multi-attribute Expert Evaluation

Alexey B. Petrovsky

Abstract The paper describes a new multi-method technology PAKS-M (Progressive Aggregation of the Classified Situations by many Methods) for multi-attribute expert evaluation of complex systems and processes. This technology provides reducing the dimension of the attribute space, constructing several hierarchical systems of composite criteria and an integral quality index, which aggregate initial attributes as well as the classification and/or ordering of multi-attribute objects using several decision making methods. The technology is based on knowledge of experts and/or preferences of decision maker. It significantly reduces the time and complexity of solution of multiple criteria tasks, and allows analyzing and explaining the results.

Keywords Multi-attribute decision making · Reduction of dimension of the attribute space · Hierarchical aggregation of attributes · Composite criteria · Integral index · Expert knowledge · Decision maker preferences

1 Introduction

A choice of object with large number of numerical and/or verbal attributes, which characterize the object properties, is one of the widespread problems of multi-attribute decision making. In these cases, it is very difficult for an expert and/or decision maker (DM) to select the best object, rank or classify objects, because, as a rule, the objects formally are not comparable with each other according to their attributes. The known methods [1–7, 17–20] are not suitable for solving the multiple criteria tasks in the attribute space of large dimension because they require significant labor costs in order to receive, process and present big

A.B. Petrovsky (✉)
Institute for Systems Analysis, Federal Research Center Informatics and Control, Russian Academy of Sciences, Prospect 60 Let Octyabrya, 9, Moscow 117312, Russia
e-mail: pab@isa.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_20

volumes of information about objects, knowledge of experts and/or the preferences of DM.

The paper describes a new multi-method technology PAKS-M (Progressive Aggregation of the Classified Situations by many Methods) for comparing, ordering, and classifying multi-attribute objects based on knowledge of experts and/or the preferences of DM [15]. This technology provides a reduction of the dimension of the attribute space by a consecutive aggregation of a large number of initial (numerical, symbolical, or verbal) characteristics of objects into a small number of composite criteria or a single integral index of quality with the verbal scales. The grades of criteria scales are constructed using different methods of verbal decision analysis. The final criteria present the initial characteristics of objects in a compact form, and are used for solving the considered choice task by several techniques of decision making. The multi-method technology PAKS-M allows one to compare the lists of criteria with each other, select the preferable criteria, analyze the obtained results, and assess the quality of the choice.

The PAKS-M technology can be applied in various problem areas where it is necessary to evaluate a collection of objects by a small number of criteria or an integral index based on knowledge of experts and/or the preferences of DM. Examples of the successful solution of such tasks are the multi-aspect evaluation of efficiency of research projects [16] and the multicriteria selection of a prospective computing complex [14].

2 Basic Stages of Technology

Generally speaking, the task of multicriteria choice is formulated as follows. A collection of objects (alternatives, options) $A_1,\ldots,A_p$, which are evaluated by one or several experts upon many criteria $Q_1,\ldots,Q_m$, is given. Each criterion Q_i has a scale $X_i = \{x_i^1,\ldots,x_i^{gi}\}$, $i = 1,\ldots,m$, the discrete numerical or verbal grades of which are ordered in some cases. Based on the knowledge of experts and/or preferences of DM, it is required: (1) to select one or several best objects; (2) to order all objects; and (3) to distribute all objects by several classes (categories). Let us consider the basic stages of solving a multicriteria choice task using the PAKS-M technology.

Firstly, an expert and/or DM forms the set $K_1,\ldots,K_m$, $m \geq 2$ of initial characteristics of objects that reflect the basic properties of the given objects. These characteristics can either be determined in advance or created in the process of analyzing the specifics of the choice task. The scale $X_i = \{x_i^1,\ldots,x_i^{gi}\}$, $i = 1,\ldots,m$, which has numerical (point, interval) or verbal evaluation grades, is constructed for each initial indicator. The scale grades can be prepared specially or used usually in practice.

Further, the dimension of the attribute space is reduced by constructing a hierarchical system of criteria in accordance with the knowledge of expert and/or

preferences of DM. The various combinations of initial attributes (evaluation tuples) are aggregated into smaller sets $L_1,\ldots,L_n$, $n < m$ of new attributes (composite criteria) formulated by the expert or DM. The attributes are aggregated consecutively step by step, i.e. the groups of criteria are combined serially in new criteria of the next level of hierarchy. The formation of the rating scale $Y_j = \{y_j^1,\ldots, y_j^{hj}\}$, $j = 1,\ldots,n$ of a composite criterion is considered as a special task of ordinal classification. It means that combinations of the grades of attributes (tuples of criteria estimates in the space $X_1 \times \cdots \times X_m$) are the classified objects, and the grades of the composite criterion are the decision classes (categories) $L_1,\ldots,L_n$ with ordered scales $Y_1,\ldots,Y_n$. Thus, each grade on the scale of composite criterion corresponds to any combination of the grades of the initial attributes and has the concrete semantic content for the expert or DM.

In the PAKS-M technology, rating scales of composite criteria can be constructed with different methods [2, 3, 8]. The simplest method is the tuple stratification, in which the multi-attribute space is cut with parallel hyper-planes. Each layer (stratum) consists of combinations of the homogeneous (for example, with the fixed sum of grade numbers) initial estimates and represents any generalized grade on the scale of composite criterion. The initial estimates can be combined into the generalized grades, for example, as follows: the best initial estimates form the best generalized grade, middle initial estimates form the middle generalized grade, and the worst initial estimates form the worst generalized grade. The maximal number of layers is equal to $M = 1 - m + \Sigma_i g_i$. The number of classes is equal to $N \le L$.

The ZAPROS (the abbreviation of Russian words: CLOsed PRocedures nearby Reference Situation) or ORCLASS (ORdinal CLASSification) methods of the verbal decision analysis [5] are more complicated because they operate on the set of possible tuples in the attribute space that is formed by the Cartesian product of the grades on the scales of all criteria. In these cases, the all-possible combinations of initial estimates in the attributes space is equal to $P = \Pi_i g_i$. The ZAPROS method allows constructing a joint ordinal scale of composite criterion from initial estimates. The ORCLASS method builds a complete and consistent classification of all tuples of initial estimates where classes form an ordinal scale of composite criterion. In order to simplify the construction of scales of composite criteria and reduce the influence of the method features, it is proposed to form the scales with a small number of verbal grades applying different methods and/or their combinations at various stages of aggregation.

Combining attributes into composite criteria, the expert or DM constructs a hierarchical system of criteria. The suggested procedure of building the hierarchical system of criteria consists of unified blocks selected by the expert or DM depending on the problem specifics. Each classification block of *i*th hierarchical level includes any attribute set and a single composite criterion. The classified objects are tuples of initial attribute estimates. Decision classes of *i*th level are grades of scale of the composite criterion. In a classification block of the next hierarchical level, composite criteria of the *i*th level are considered as new attributes. Tuples of their scale

grades represent new classified objects in the reduced attribute space, whereas decision classes of the $(i + 1)$th level will be now scale grades of new composite criterion. As a rule, the criteria aggregation is a multi-stage procedure. Therefore, at each stage of aggregation, it is recommended to combine a small number of attributes into a composite criterion. The upper level of the system can be several final criteria or, if it is necessary, a single integral index, scale grades of which characterize a quality of the objects considered.

The construction of hierarchical systems of criteria and formation of verbal rating scales of criteria are subjective non-formalized procedures. In the PAKS-M technology, several hierarchical systems of composite criteria with different rating scales, that variously aggregate initial characteristics, are constructed. An expert or DM, based on his/her knowledge, experience and intuition, defines the structure, number and content of the criteria in the aggregation-tree. The person establishes at each level of hierarchy including the highest level, which of the attributes are considered as the independent criteria and which are combined into any particular composite criterion.

These procedures have the following peculiarities. When the object has many characteristics, it is rather difficult to say in advance, which of the characteristics will be important for choosing the best object. This becomes clear only when the task will be solved. Nevertheless, depending on the specifics of objects it is recommended to determine previously, which of initial indicators can be the final criteria. In addition, combining a large number of indicators in a composite criterion can create difficulties while the explanation of the results of choice.

The increase of the number of grades on rating scales of criteria enlarges the number of evaluation combinations in aggregating attributes into a composite criterion, which in turn complicates the criteria aggregation and increases the duration of choice. However, even a slight change in the range of variables in any grade of the criteria can impact significantly on the grades of the scale of composite criteria of the next levels and the final index. These issues also dictate the need to build several hierarchical systems of criteria that differ from each other.

At the conclusive stage, a multicriteria choice task is solved for each hierarchical system of criteria. It is convenient to consider such result as an expression of the viewpoint of any expert/DM. In other words, while using several criteria systems, different methods for constructing the scales of composite criteria and the integral index, the initial task of choice transforms into the task of a collective choice. In order to increase the validity of the final decision, we use several methods of group multicriteria choice, for instance, the ARAMIS (Aggregation and Ranking Alternatives nearby the Multi-attribute Ideal Situations) method, the lexicographic ordering by grades of the evaluations, the weighted sums of ranks, Borda procedure, and other techniques [2, 8, 10–13]. Then the expert/DM analyzes the results, which are obtained using different techniques, and makes the conclusive choice: selection of the best objects, ranking objects or division of objects into classes (groups).

One can be evaluate the efficiency of the PAKS-M technology in different ways depending on the type of the solved tasks of multicriteria choice. For example, in the ranking tasks, the efficiency is understood as the ratio of the numbers of

not-comparable alternatives before and after the reduction of the space dimension. In the classification tasks, the efficiency is evaluated by the comparison of numbers of calls to an expert/DM, which are necessary for the construction of consistent classifications in the initial and new space of the attributes. However, in the case of classification of large dimension, it is simply not possible to construct a fully consistent classification in the initial attribute space.

3 Evaluation of Research Efficiency

The Russian Foundation for Basic Research (RFBR) is the federal agency that organizes and funds basic research, exams their practical application [16]. In RFBR, there is the special expertise for multi-expert and multicriteria assessment of grant applications and completed projects. Each project is evaluated independently by several experts without the consent of their views. To assess the content of application and the obtained results experts use specific qualitative criteria with detailed verbal rating scales. In addition, the expert express his/her opinion on whether to support the project (at the competition stage), continue the project (at the implementation stage), and what is scientific and practical value of the obtained result (at the final stage). On the basis of expert recommendations, the Expert Board of RFBR decides to approve, continue or reject the project, as well as evaluates the project efficiency.

According to the expertise rules of RFBR, completed goal-oriented R&D projects and the obtained results are evaluated upon the following indicators (initial criteria): K_1 "Degree of the problem solution", K_2 "Scientific level of results", K_3 "Results protection", K_4 "Prospects of results application", K_5 "Result correspondence to the project goal", K_6 "Goal achievement", K_7 "Difficulties of the project performance", K_8 "Collaboration with users". Each criterion has 2- or 3-point rating scale $X_i = \{x_i^1, x_i^2, x_i^3\}$, $i = 1,\ldots,8$ of ordered verbal grades. For example, the scale X_1 of the criterion "Degree of the problem solution" looks as follows: x_1^1—'the problem is solved completely', x_1^2—'the problem is solved partially', x_1^3—'the problem is not solved'. The "Goal achievement" is estimated as x_6^1—'really', x_6^2—'non-really'. For simplicity, the verbal notations of the criteria grades are replaced by the numerical symbols as follows: 0 designates the best grade x_i^1, 1—the middle (or the worst) grade x_i^2, 2—the worst grade x_i^3.

The evaluation of research efficiency is one of the practical problems, which require the development of integrated index. In order to evaluate the feasibility of effective practical use of the obtained results, the notion of "Project efficiency" has been formalized with the PAKS-M technology [15]. Construction of the integral index of project efficiency was examined as the multiple criteria classification problem in the reduced multi-attribute space. The classified alternatives were multicriteria estimations $(x_1^{e1}, x_2^{e2},\ldots,x_8^{e8})$ of projects in the attribute space $X_1 \times \cdots \times X_8$. DM can form the integral index of project efficiency in different ways, compare the constructed indexes and analyze the obtained results.

Firstly, the initial indicators K_1, K_2, K_3 were aggregated into a composite criterion L_1 "Project results", the indicators K_5, K_6, K_7 into a composite criterion L_2 "Project realization", and the indicators K_4, K_8 into a composite criterion L_3 "Application of project results". The composite criteria L_1, L_2, L_3 have ordinal scales $Y_j = \{y_j^1, y_j^2, y_j^3\}$, $j = 1,2,3$. The names of the verbal estimates (y_j^1—high, y_j^2—middle, y_j^3—low) designate decision classes of the first hierarchical level, depend on a content of corresponding composite criterion and can be replaced, as above, by the numbers 0,1,2. All tuples $(y_1^{e1}, y_2^{e2}, y_3^{e3})$ of estimate grades on the scales of composite criteria L_1, L_2, L_3 present now new classified objects of the next hierarchical level in the space $Y_1 \times Y_2 \times Y_3$. Then these tuples are divided into five ordered classes $D_1,\ldots,D_5$, corresponding to rating grades of scale $Z = \{z^1, z^2, z^3, z^4, z^5\}$ of the top level index "Project efficiency", where z^1—'superior', z^2—'high', z^3—'middle', z^4—'low', z^5—'unsatisfactory'. The hierarchical frame of building the composite criteria, integral index, and composing rating scales is presented in Fig. 1.

Various techniques can be used for constructing scales of the composite criteria L_1, L_2, L_3 and the integral index D in accordance with DM preferences. The grades of the integral index D "Project efficiency" were formed with the following different methods of verbal decision analysis: M_1—ORCLASS method on all levels of the criteria hierarchy (OC); M_2—stratification of tuples on all levels of the criteria hierarchy (ST), M_3—stratification of tuples on the lower level of the criteria hierarchy, and ORCLASS method on the upper level of the criteria hierarchy (ST + OC); M_4—ORCLASS method on the lower level of the criteria hierarchy, and stratification of tuples on the upper level of the criteria hierarchy (OC + ST).

While applying, for instance, the ORCLASS method, the class 'High project results' (grade $y_1^1 = 0$) of the composite criteria L_1 includes the following estimate

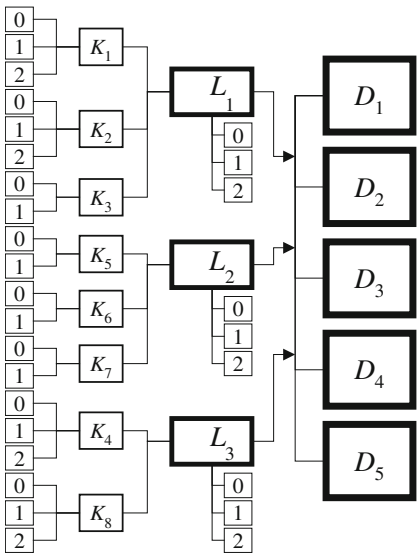


Fig. 1 Hierarchical frame of building composite criteria and an integral index

combinations: (000), (001), (010), (100). The class 'Middle project results' (grade $y_1^2 = 1$) includes the estimates (011), (021), (101), (111), (201), (110), (200), (020), (210), (120). The class 'Low project results' (grade $y_1^3 = 2$) includes the estimates (121), (211), (112), (220). The class 'High project realization' (grade $y_2^1 = 0$) of the composite criteria L_2 includes the best estimates (000). The class 'Middle project realization' (grade $y_2^2 = 1$) includes the average estimates (001), (010), (100), (011), (101), (110). The class 'Low project realization' (grade $y_2^3 = 2$) includes the worst estimates (111). The class 'Results will be used completely' (grade $y_3^1 = 0$) of the composite criteria L_3 includes the estimates (00). The class 'Results will be used partially' (grade $y_3^2 = 1$) includes the estimates (01), (10), (11), (02), (20). The class 'Results will be used poorly' (grade $y_3^3 = 2$) includes the estimates (12), (21), (22). Hereby and below commas between tuple components are omitted for the simplicity.

Combining analogously the attributes of the composite criteria L_1, L_2, L_3 with the tuple stratification technique, one may obtain the following results. The class D_1 'Superior efficiency of project' (grade z^1) consists of the best estimates (000). The class D_2 'High efficiency of project' (grade z^2) consists of the estimates (100), (010), (001), (002), (101), (011), (200), (110), (020). The class D_3 'Middle efficiency of project' (grade z^3) consists of the estimates (102), (012), (201), (111), (021), (210), (120). The class D_4 'Low efficiency of project' (grade z^4) consists of the estimates (202), (112), (022), (211), (121), (220), (212), (122), (221). The class D_5 'Unsatisfactory efficiency of project' (grade z^4) consists of the worst estimates (222). Thus, the real projects estimated by initial criteria are assigned directly to the generated decision classes. Note that this procedure requires essentially fewer efforts than other methods of multicriteria ordinal classification.

The model database consisted of expert assessments of results of the goal-oriented R&D projects, which have been completed in 2007 on the following areas: Mathematics, Mechanics and Computer Science (total 48 projects), Chemistry (total 54 projects), Information and telecommunication resources (total 21 projects) [16]. Two experts evaluated each project by eight basic criteria K_1–K_8. Examples of the integral indexes of project efficiency are shown in Fig. 2. The

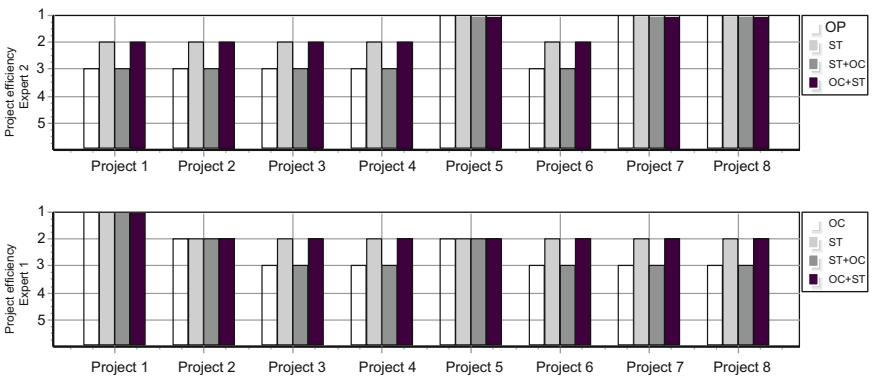


Fig. 2 Integral indexes of project efficiency obtained with different techniques

grades of efficiency coincide in 74 and 48 % cases (the area 01); in 72 and 24 % cases (the area 03); in 76 and 62 % cases (the area 07). The first number is related to projects estimated by the first expert, the second number—by the second expert.

It was also studied two various ways for aggregating the index of top level D "Project efficiency". There were forming the criteria hierarchies, which combined, firstly, the initial indicators K_1–K_3, K_5–K_7, and K_4, K_8 into three intermediate composite criteria L_1, L_2, L_3 (option A), and, secondly, the initial indicators K_1–K_4 and K_1–K_4 into two composite intermediate criteria L_1, L_2 (option B). So, for instance, in the area 03, 6 and 16 projects had the superior efficiency, 40 and 75 projects had the high efficiency, 59 and 13 projects had the middle efficiency, 1 and 2 projects had the low efficiency, 2 and 2 projects had the unsatisfactory efficiency. The first number is related to the option A, the second—to the option B. In general, the integral indexes of project efficiency coincided in 41 cases out of 108. In other cases, the integral indexes differed by no more than one gradation. These data confirm the high stability of the suggested approach to build the composite criteria on all hierarchy levels and constructing the integral indexes of project efficiency.

The best projects were selected with the ARAMIS method that allow us to order multi-attribute objects, which are described by many repeated quantitative and/or qualitative attributes, without constructing the individual rankings of objects [10, 13]. The multi-attribute objects are considered as points of the Petrovsky metric space of multisets [9]. The best and worst objects (which can be hypothetical) have, respectively, the highest and lowest scores by all criteria in the judgments of all experts. The objects are ordered, for example, by the value of an index $l_+(A_i) = d_+/(d_+ + d_-)$ of the relative proximity to the best object, where $d_+ = d(A_+, A_i)$ is the distance from the object A_i to the best object A_+ and $d_- = d(A_-, A_i)$ is the distance to the worst object A_- in the metric space. So, in the final ranking projects on Mathematics, Mechanics and Computer Science, there were 23 projects of rank 1 with the superior efficiency, 1 project of rank 2 with the efficiency between the superior and high, 24 projects of rank 3 with the high efficiency.

4 Conclusions

The new PAKS-M technology for solving the tasks of multi-criteria choice of a large dimension was proposed. In the technology, the procedures for the reduction of the attribute space dimension are used together with various methods of decision making and/or their combinations. An important feature of the PAKS-M technology is the opportunity to create hierarchical systems with various degrees of criteria aggregation, solve an initial task of choice using several methods, and give a clear explanation of the results. The analysis of the results obtained for different hierarchical systems of criteria aggregation allows us to choose the most convenient system of criteria and apply the constructed systems of criteria for increasing the validity of the choice of the preferred options.

While aggregating criteria, the establishment of semantic relationships between initial indicators and composite criteria plays an important role. An expert/DM is often faced with the inconsistency and contradictory of the results in the process of achieving a satisfactory solution. Such situations can be caused both by an unsuccessful combination of composite criteria and by unsuccessful forming the scale grades of composite criteria and an integral index.

The PAKS-M technology has a certain universality, because it allows us to operate both with symbolical (qualitative) and numerical (quantitative) information. An attractive feature of PAKS-M is also the ability to use various methods of decision making and technologies of processing information.

Acknowledgments This work was supported by the Research Program of the Russian Academy of Sciences "Intelligent Information Technologies and Systems"; the Russian Foundation for Basic Research (projects 14-07-00916, 14-29-05,025, 15-07-04760, 16-07-01125).

References

1. Doumpos, M., Zopounidis, C.: Multicriteria Decision Aid Classification Methods. Kluwer Academic Publishers, Dordrecht (2002)
2. Hwang, C.L., Lin, M.J.: Group Decision Making under Multiple Criteria. Springer, Berlin (1987)
3. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making—Methods and Applications: A State of the Art Survey. Springer, New York (1981)
4. Keeney, R.L., Raiffa, H.: Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge University Press, Cambridge (1993)
5. Larichev, O.I.: Verbal Decision Analysis. Nauka, Moscow (2006) (in Russian)
6. Larichev, O.I., Olson, D.L.: Multiple Criteria Analysis in Strategic Siting Problems. Kluwer Academic Publishers, Boston (2001)
7. Nogin, V.D.: Decision-Making in Multicriteria Environment: Quantitative Approach. Fizmatlit, Moscow (2005) (in Russian)
8. Petrovsky, A.B.: Decision Making Theory. Akademiya, Moscow (2009) (in Russian)
9. Petrovsky, A.B.: Spaces of Sets and Multisets. Editorial URSS, Moscow (2003). in Russian
10. Petrovsky, A.B.: Group Verbal Decision Analysis. In: Adam, F., Humphreys, P. (eds.) Encyclopedia of Decision Making and Decision Support Technologies, pp. 418–425. IGI Global, Hershey (2008)
11. Petrovsky, A.B.: Methods for the Group Classification of Multi-Attribute Objects. Art. Intel. and Decis. Mak. (Part 1) 3, 3–14 (2009), (Part 2) 4, 3–14 (2009). Translation: Sc. Tech. Inf. Proc. 37, 5, (Part 1) 346–356, (Part 2) 357–368 (2010) (in Russian)
12. Petrovsky, A.B.: Method Maska for Group Expert Classification of Multi-Attribute Objects. Dokl. Math. **81**(2), 317–321 (2010)
13. Petrovsky, A.B.: Method for Group Ordering Multi-Attribute Objects. In: Engemann, K.J., Lasker, G.E. (eds.) Advances in Decision Technology and Intelligent Information Systems, XI, pp. 27–31. The International Institute for Advanced Studies in Systems Research and Cybernetics, Tecumseh (2010)
14. Petrovsky, A.B., Lobanov, V.N.: Selection of Complex System in the Reduced Multiple Criteria Space. World Appl. Sci. J. **29**(10), 1315–1319 (2014)

15. Petrovsky, A.B., Lobanov, V.N.: Multi-Criteria choice in the attribute space of large dimension: multi-method technology PAKS-M. Art. Intel. Decis. Mak. **3**, 92–104 (2014) in Russian, translation: Sc. and Tech. Inf. Proc. **42**(6), 470–480 (2015)
16. Petrovsky, A.B., Royzenson, G.V., Tikhonov, I.P., Balyshev, A.V.: Group Ranking R&D Projects by Discordant Multiple Criteria Estimates. In: Engemann, K.J., Lasker, G.E. (eds.) Advances in Decision Technology and Intelligent Information Systems, XI, pp. 32–36. The International Institute for Advanced Studies in Systems Research and Cybernetics, Tecumseh (2010)
17. Roy, B., Bouyssou, D.: Aide Multicritère à la Décision: Méthodes et Cas. Economica, Paris (1993)
18. Saaty, T.: Multicriteria Decision Making: The Analytic Hierarchy Process. RWS Publications, Pittsburgh (1990)
19. Steuer, R.: Multiple Criteria Optimization: Theory, Computation, and Application. Wiley, New York (1986)
20. Vincke, P.: Multicriteria Decision Aid. Wiley, Chichester (1992)

Using Fuzzy Models by Systems Engineers at the Stages of System Lifecycle

V.K. Batovrin and A.S. Korolev

Abstract Authors consider the appliance of fuzzy models for decision making tasks at the stages of system lifecycle, especially at concept development stage, where such sort of tasks are more actual. Article demonstrates sample models and based on them decision-support tools that can be used by system engineers in the technical systems development projects. Also the example is provided, that shows a calculation during decision making process at the concept development stage in the artificial system lifecycle.

Keywords Systems engineering · System lifecycle · Decision making · Fuzzy inference

1 Introduction

Systems Engineering as a discipline and Lifecycle Approach as an instrument belong to the systems engineering methodology, that offers a methodological basis for successful artificial systems, the main criterion for the success of which is to achieve a balance between interests of all stakeholders [1, 2].

There are more than twenty of system engineering processes, each of which contains a number of repetitive actions needed to build a system. The iterative appliance of system engineering process at each stage of the system lifecycle allows you to reduce the cost of the complete lifecycle, reduce the time of system creation, as well as achieve other competitive advantages [3, 4].

V.K. Batovrin
Radiotechnics and Electronics, Moscow State University of Information Technologies, Moscow, Russia
e-mail: batovrin@mirea.ru

A.S. Korolev (✉)
National Research Nuclear University MEPhI, Moscow, Russia
e-mail: ASkorolev@mephi.ru; popesku@inbox.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_21

System engineer driving system lifecycle makes decisions in the lifecycle checkpoints. These decisions may be of a different nature, be both more engineering and management, affect both the design features of a particular module, and the strategic decision to continue the project [5]. To support decision-making, as well as other elements of its activities, such as requirements management, system engineer often uses so-called enabling systems. In this case, decision support systems (DSS) are applied as enabling systems. There are different types of DSS designed both for commercial purposes, and freely distributed [6].

Depending on the objectives, DSS can be both general as a shell for making decisions in different tasks, and specific, aimed at the decision of a concrete problem.

DSS typically uses a specific mathematical apparatus for implementing a decision-making procedure embedded in it. Below, we consider a decision-making procedure in the multi-criteria analysis of alternatives using fuzzy inference model [7]. The use of fuzzy models in the tasks of decision-making is important for the activity of system engineers due to the fact that at such stages of the system lifecycle, as Concept Development stage, much of the factors influencing the decision are qualitative, and these factors are expressed linguistically using relative scales.

2 Decision-Making Tasks at the Stages of Systems Lifecycle

Systems engineering, applied as methodology for development and maintaining of artificial systems, specifies the process comprising the basic steps for successful systems development activity based on a balance of interests of all stakeholders.

Found wide application lifecycle approach—one of the main approaches of Systems Engineering—helps to form the basis of the systems engineering application, being a kind of activity model for systems development [8, 9]. A. Kosjakov, W. Sweet, S. Seymour, S. Biemer describe a typical lifecycle of the system consisted of three main stages: Concept Development, Engineering Development and Postdevelopment stage [3].

Lifecycle approach supposes that you need to select the appropriate lifecycle model, to formalize the process at every stage of lifecycle model and manage these processes. In addition, each stage of the lifecycle can be considered as a separate project of system development with an output result of a target system in a certain physical terms (for example, the documentation of technical analysis, design documentation, layout, prototype, product, etc.). Systems engineering process is used for the implementation of each such project. Thus, the system engineering process is repeated within the lifecycle stages with a new specification of the input and output data for each iteration.

Let's consider the system engineering process at the stage of Concept Definition, which is a sub-stage of Concept Development Stage. This process consists here of the following four steps: Performance Requirement Analysis, Functional Analysis and Formulation, Physical definition, Concept Validation [3].

Decisions, being made during the process of Concept Definition, generally relate to selecting a particular system configuration, and defining the functions performed by the system. These solutions are becoming essential for the establishment of the final performance indicators, value and utility of new system. To support such decisions, the process of trade-offs analysis is applied, which is used also at the next phases of the system development. At the stage of the Concept Definition the process of trade-offs analysis affects the following aspects of the system project:

- Objectives of the project.
- Variants of the system architecture.
- High-level functional requirements.
- Performance indicators.
- Interfaces.
- Alternative concepts.
- Anticipated risks.

Well-formed system concept should have the following characteristics [10]:

- Can satisfy the project requirements.
- Clearly define the resource requirements of the project lifecycle.
- Identifies potential environmental effects.
- Defines the criteria for assessing the cost and schedule of the project.
- Determines the appropriate technological solutions.
- Identify, measure and indicates the most likely risks.
- Defines the criteria for success, and evaluation criteria for the next phase.
- Provides traceability of hierarchical requirements and planned work breakdown structure in the physical architecture at the next stages of lifecycle.

During the trade-offs analyzing at different stages of the system lifecycle it is advisable to use a special mathematical tool for finding quantitative weight or acceptability of alternative. At the same time, due to the fact that at the Concept Development stage a significant portion of the decisions are taken on the basis of fuzzy parameters, the use of fuzzy models in the decision-making process takes here the most urgent feature than in later stages of lifecycle.

Below we consider the decision-making example for system concept selection process, where each input evaluation version of the concept is carried out with the help of linguistic terms and a logic minimum operation is used for the fuzzy inference rules convolution.

3 The Use of Fuzzy Inference to Determine the System Concept

Suppose there are several alternative system concepts. It is needed to choose one for the final implementation of the project.

The discussion between the project participants gave the following results of the linguistic evaluation of alternatives taking into account the balance of the stakeholders interests:

d_1: If the concept is attractive to potential customers, allows to open up significant market share and profit, then it is satisfactory (meets the basic requirements);
d_2: If, in addition, the concept can be realized by a team of organizations-performers with experience in the implementation of projects like this, then it is more than satisfactory;
d_3: If, in addition to the condition d2, the concept has a high technical and economic level, including project scope, originality, priority areas, the degree of elaboration, legal protection and ecological level, then it is flawless;
d_4: If the concept has all agreed to d3, except the team of experienced performers, then it is a very satisfactory;
d_5: If the concept is very attractive to consumers, there is a team of experienced performers for its implementation, but the level of the market competition and the technical and economic level of the concept do not allow it to take the expected market share and there is a risk of the lack of the profit, then the concept would still be satisfactory;
d_6: If the concept is not attractive to consumers or cannot be implemented by existing executive team, then it is unsatisfactory.

The analysis of these fragments gives the five criteria for decision-making:

X_1: consumer attractiveness
X_2: market share
X_3: profit
X_4: the presence of a team of experienced performers
X_5: the technical and economic level of the concept

We perform the descriptions of the fragments above in the linguistic rules.

d_1: F X_1 = ATTRACTIVE AND X_2 = SIGNIFICANT AND X_3 = ACCEPTABLE, THEN Y = SATISFACTORY
d_2: IF X_1 = ATTRACTIVE AND X_2 = SIGNIFICANT AND X_3 = ACCEPTABLE AND X_4 = AVAILABLE, THEN Y = MORE THAN SATISFACTORY
d_3: IF X_1 = ATTRACTIVE AND X_2 = SIGNIFICANT AND X_3 = ACCEPTABLE AND X_4 = AVAILABLE AND X_5 = HIGH, THEN Y = PERFECT
d_4: IF X_1 = ATTRACTIVE AND X_2 = SIGNIFICANT AND X_3 = ACCEPTABLE AND X_5 = HIGH, THEN Y = VERY SATISFACTORY

d_5: IF X_1 = ATTRACTIVE AND X_2 = NOT SIGNIFICANT AND X_3 = NOT ACCEPTABLE AND X_4 = AVAILABLE, THEN Y = SATISFACTORY
d_6: IF X_1 = NOT ATTRACTIVE AND X_4 = NOT AVAILABLE, THEN Y = UNSATISFACTORY

The variable Y defined on the set $J = \{0; 0,1; 0,2 \cdots 1\}$
Define the following membership functions for linguistic variables.
SATISFACTORY (S):

$$\mu_S(x) = x, x \in J \tag{1}$$

MORE THAN SATISFACTORY (MS):

$$\mu_{MS}(x) = \sqrt{x^3}, x \in J \tag{2}$$

PERFECT (P):

$$\mu_P(x) = \begin{cases} 1, & if \quad x = 1 \\ 0, & if \quad x \neq 1 \end{cases} \tag{3}$$

VERY SATISFACTORY (VS):

$$\mu_{VS}(x) = x^2, x \in J \tag{4}$$

UNSATISFACTORY (US):

$$\mu_{US}(x) = 1 - x, x \in J \tag{5}$$

The choice is made of the three concepts $K = \{k_1, k_2, k_3\}$
There are next concepts valuations.
"Attractive concept for consumers" $A = \{0,8/k_1, 0,6/k_2, 0,3/k_3\}$
"The concept is able to take a large share of the market" $B = \{0,7/k_1, 0,7/k_2, 0,5/k_3\}$
"The concept that clearly profitable" $C = \{0,6/k_1, 0,4/k_2, 0,9/k_3\}$
"The concept, which can be fully implemented existing executive team in a short time" $D = \{0/k_1, 0,3/k_2, 0,8/k_3\}$
"The concept has a high technical and economic level" $E = \{0,3/k_1, 0,4/k_2, 0,7/k_3\}$
Thereafter, the fragments of knowledge take the following form:

d_1: IF X = A, AND B, AND C, THEN Y = S
d_2: IF X = A, AND B, AND C, AND D, THEN Y = MS
d_3: IF X = A, AND B, AND C, AND D, AND E,
THEN Y = P
d_4: IF X = A, AND B, AND C, AND E, THEN Y = VS

d_5: IF X = VERY A, AND NOT B, AND NOT C, AND D, THEN Y = S
d_6: IF X = NOT A, AND NOT D, THEN Y = US

We translate these operations according to the following rule intersection of fuzzy sets:

$$\mu_{Ai}(v) = \min_{v \in V}(\mu_{Ai1}(u_1), \mu_{Ai2}(u_2), \ldots, \mu_{Aip}(u_p)), \tag{6}$$

where, $V = U_1 \times U_2 \times \ldots \times U_p$, $v = (u_1, u_2, \ldots u_p)$, $\mu_{Aij}(u_j)$—degree of membership of u_i in fuzzy set A_{ij}.

The following results will be achieved.

For d_1:

$$\mu_{M1}(k) = \min(\mu_A(k), \mu_B(k), \mu_C(k))$$
$$M_1 = \{0, 6/k_1, 0, 4/k_2, 0, 3/k_3\}$$

For d_2:

$$\mu_{M2}(k) = \min(\mu_A(k), \mu_B(k), \mu_C(k), \mu_D(k))$$
$$M_2 = \{0/k_1, 0, 3/k_2, 0, 3/k_3\}$$

For d_3:

$$\mu_{M3}(k) = \min(\mu_A(k), \mu_B(k), \mu_C(k), \mu_D(k), \mu_E(k))$$
$$M_3 = \{0/k_1, 0, 3/k_2, 0, 3/k_3\}$$

For d_4:

$$\mu_{M4}(k) = \min(\mu_A(k), \mu_B(k), \mu_C(k), \mu_E(k))$$
$$M_4 = \{0, 3/k_1, 0, 4/k_2, 0, 3/k_3\}$$

For d_5:

$$\mu_{M5}(k) = \min(\mu_{A^2}(k), 1 - \mu_B(k), 1 - \mu_C(k), \mu_D(k))$$
$$M_5 = \{0/k_1, 0, 3/k_2, 0, 1/k_3\}$$

For d_6:

$$\mu_{M6}(k) = \min(1 - \mu_A(k), 1 - \mu_D(k))$$
$$M_6 = \{0, 2/k_1, 0, 4/k_2, 0, 2/k_3\}$$

Thus:

d_1: IF X = M_1, THEN Y = S
d_2: IF X = M_2, THEN Y = MS
d_3: IF X = M_3, THEN Y = P
d_4: IF X = M_4, THEN Y = VS
d_5: IF X = M_5, THEN Y = S
d_6: IF X = M_6, THEN Y = US

We use the following rule:

$$\mu_D(w,i) = \min_{w \in W}(\mu_{Hj}(w,i)), j = \overline{1,q} \tag{7}$$

for conversion of implication « IF X = M, THEN Y = Q » in the expression $\mu_D(k,i) = \min(1, 1 - \mu_M(k) + \mu_Q(k))$ for each pair $(k,i) \in U \times J$.

$\mu_Q(k)$ is calculated using the corresponding function (1)–(5).

We get the following fuzzy subsets of $U \times J$.

D_1:

	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
k_1	0,4	0,5	0,6	0,7	0,8	0,9	1	1	1	1	1
k_2	0,6	0,7	0,8	0,9	1	1	1	1	1	1	1
k_3	0,7	0,8	0,9	1	1	1	1	1	1	1	1

D_2:

	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
k_1	1	1	1	1	1	1	1	1	1	1	1
k_2	0,7	0,73	0,79	0,86	0,95	1	1	1	1	1	1
k_3	0,7	0,73	0,79	0,86	0,95	1	1	1	1	1	1

D_3:

	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
k_1	1	1	1	1	1	1	1	1	1	1	1
k_2	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	1
k_3	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	1

D_4:

	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
k_1	0,7	0,71	0,74	0,79	0,86	0,95	1	1	1	1	1
k_2	0,6	0,61	0,64	0,69	0,76	0,85	0,96	1	1	1	1
k_3	0,7	0,71	0,74	0,79	0,86	0,95	1	1	1	1	1

D_5:

	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
k_1	1	1	1	1	1	1	1	1	1	1	1
k_2	0,7	0,8	0,9	1	1	1	1	1	1	1	1
k_3	0,9	1	1	1	1	1	1	1	1	1	1

D_6:

	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
k_1	1	1	1	1	1	1	1	1	1	0,9	0,8
k_2	1	1	1	1	1	1	1	0,9	0,8	0,7	0,6
k_3	1	1	1	1	1	1	1	1	1	0,9	0,8

For general functional solution perform intersection operation: $D = D_1 \cap D_2 \cap D_3 \cap D_4 \cap D_5 \cap D_6$, or

D:

	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
k_1	0,4	0,5	0,6	0,7	0,8	0,9	1	1	1	0,9	0,8
k_2	0,6	0,61	0,64	0,69	0,7	0,7	0,7	0,7	0,7	0,7	0,6
k_3	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,8

To calculate the sufficiency of each alternative the composite output rule in fuzzy environment will be applied:$E_h = G_h \circ D$, where E_h—the degree of alternative h satisfaction, G_h—view of alternative h as a fuzzy subsets K, D—functional solution.

Thus $\mu_{Eh}(i) = \max\limits_{k \in K}(\min(\mu_{Gh}(k), \mu_{Dh}(k, i)))$. In this cause $\mu_{Gh}(k) = 0$, with $k \neq k_h$, $\mu_{Gh}(k) = 1$, with $k = k_h$. So $\mu_{Eh}(i) = \mu_D(k_h, i)$. With the other words, E_h is the row h in a matrix D.

Now we apply the comparison procedure to the fuzzy subsets E_1, E_2, E_3 in the unit interval for the best solutions.

For the first alternative

$$E_1 = \{0,4/0;0,5/0,1;0,6/0,2;0,7/0,3;0,8/0,4;0,9/0,5;1/0,6;1/0,7;1/0,8;0,9/0,9;0,8/1\}$$

Now we should evaluate the tier sets $E_{j\alpha}$. Their capacity $M(E_{j\alpha})$ is calculated from the formula:

$$M(E_{j\alpha}) = \sum_{i=1}^{n} \frac{x_i}{n}$$

where

$$0 \le \alpha \le 0,4; d\alpha = 0,4$$
$$E_{1\alpha} = \{0;0,1;0,2;0,3;0,4;0,5;0,6;0,7;0,8;0,9;1\},\}$$
$$M(E_{j\alpha}) = 0,5$$
$$0,4 < \alpha \le 0,5; d\alpha = 0,1$$
$$E_{1\alpha} = \{0,1;0,2;0,3;0,4;0,5;0,6;0,7;0,8;0,9;1\},$$
$$M(E_{j\alpha}) = 0,55$$
$$0,5 < \alpha \le 0,6; d\alpha = 0,1$$
$$E_{1\alpha} = \{0,2;0,3;0,4;0,5;0,6;0,7;0,8;0,9;1\},$$
$$M(E_{j\alpha}) = 0,6$$
$$0,6 < \alpha \le 0,7; d\alpha = 0,1$$
$$E_{1\alpha} = \{0,3;0,4;0,5;0,6;0,7;0,8;0,9;1\},$$
$$M(E_{j\alpha}) = 0,65$$
$$0,7 < \alpha \le 0,8; d\alpha = 0,1$$
$$E_{1\alpha} = \{0,4;0,5;0,6;0,7;0,8;0,9;1\},$$
$$M(E_{j\alpha}) = 0,7$$
$$0,8 < \alpha \le 0,9; d\alpha = 0,1$$
$$E_{1\alpha} = \{0,5;0,6;0,7;0,8;0,9\},$$
$$M(E_{j\alpha}) = 0,7$$
$$0,9 < \alpha \le 1; d\alpha = 0,1$$
$$E_{1\alpha} = \{0,6;0,7;0,8\}, M(E_{j\alpha}) = 0,7$$

Calculate a point estimation of E_1 from the formula:

$$F(E_1) = \frac{1}{\alpha_{\max}} \int_0^{\alpha\max} M(E_{1\alpha}) d\alpha = \frac{1}{1} \int_0^1 M(E_{1\alpha}) d\alpha$$
$$F(E_1) = 1/1(0,5 * 0,4 + 0,55 * 0,1 + 0,6 * 0, + 0,65 * 0,1 + 0,7 * 0,1 + 0,7 * 0,1 + 0,7 * 0,1)$$
$$F(E_1) = 0.59$$

For the second alternative:

$$0 \le \alpha \le 0,6; d\alpha = 0,6$$
$$E_{2\alpha} = \{0;0,1;0,2;0,3;0,4;0,5;0,6;0,7;0,8;0,9;1\},$$
$$M(E_{j\alpha}) = 0,5$$
$$0,6 < \alpha \le 0,61; d\alpha = 0,01$$
$$E_{2\alpha} = \{0,1;0,2;0,3;0,4;0,5;0,6;0,7;0,8;0,9\},$$
$$M(E_{j\alpha}) = 0,45$$
$$0,61 < \alpha \le 0,64; d\alpha = 0,03$$
$$E_{2\alpha} = \{0,2;0,3;0,4;0,5;0,6;0,7;0,8;0,9;1\},$$
$$M(E_{j\alpha}) = 0,6$$
$$0,64 < \alpha \le 0,69; d\alpha = 0,05$$
$$E_{2\alpha} = \{0,3;0,4;0,5;0,6;0,7;0,8;0,9;1\},$$
$$M(E_{j\alpha}) = 0,65$$
$$0,69 < \alpha \le 0,7; d\alpha = 0,01$$
$$E_{2\alpha} = \{0,4;0,5;0,6;0,7;0,8;0,9;1\},$$
$$M(E_{j\alpha}) = 0,7$$

Point estimation:

$$F(E_2) = 1/0,7(0,5 * 0,6 + 0,45 * 0,01 + 0,6 * 0,03 + 0,65 * 0,05 + 0,7 * 0,01) = 0,52$$

For the third alternative:

$$0 \le \alpha \le 0,7; d\alpha = 0,7$$
$$E_{3\alpha} = \{0;0,1;0,2;0,3;0,4;0,5;0,6;0,7;0,8;0,9;1\},$$
$$M(E_{j\alpha}) = 0,5$$
$$0,7 < \alpha \le 0,8; d\alpha = 0,1$$
$$E_{3\alpha} = \{1\},$$
$$M(E_{j\alpha}) = 1$$

Point estimation:

$$F(E_3) = 1/0,8(0,5 * 0,7 + 1 * 0,1) = 0.56$$

According to the calculations the first concept is considered to be the best with point estimation of 0.59.

4 The Results of the Method Application for the Future Project Strategy Determination

Domestic company engaged in the development of a unique portable device - gamma-ray detector for cancer diagnosing. It is assumed that the detector has better technical features and has a lower price in comparison with foreign devices. It should be marked that the detector prototype - wire device, based on silicon photomultiplier, is already developed and can be considered as a previous system. There are several options for the further project implementation of the device development:

1. Finalize the existing version of the device based on silicon photomultiplier to the production version.
2. Develop a wire locator based on semiconductor crystal CZT, which has better performance than its predecessor.
3. Develop a wireless locator based on semiconductor crystal CZT.

We performed an analysis of the above trade-offs applying the method of fuzzy inference and can offer the most efficient version of the project.

Based on the criteria's analysis: attractiveness for consumers, market share, profit, the availability of a team of experienced performers, technical and economic level of the concept, there were got the following values of the point estimations: $F(E_1) = 0.63$, $F(E_2) = 0.47$, $F(E_3) = 0.75$. Note that in this case, despite the fact that the cost of the latest version implementation is greatest, which is reflected in a lower level of the project feasibility, a maximum score of alternative is achieved by a higher market share.

According to the calculations the third concept with the point estimation of 0.75 should be considered as the most suitable. Thus, it is more profitable to start the implementation of the wireless device based on a semiconductor crystal CZT for the company.

5 Conclusions

The article shows the location of decision-making problems in the system engineering process at the stages of system lifecycle. We tried to demonstrate the opportunities of fuzzy inference model application in the trade-offs analysis at the Concept Development stage of artificial system lifecycle. It is also demonstrated how such models and decision-support tools based on them can be used by a system engineer in R&D projects of technical systems in the hi tech sectors of the economy. In particular, as an example of the R&D project a portable gamma-ray detector for cancer diagnosis was taken, and with the help of the trade-offs analysis it was carried out the selection of the most appropriate device concept that allowed to move to the next stage of lifecycle—Engineering Development.

References

1. INCOSE Systems Engineering Handbook v. 3.2.1/INCOSE–TP–2003–002—03.2.1/Jan 2011
2. Blanchard, B.S.: System Engineering Management, p. 560. Wiley (2008)
3. Kosjakov, A., Sweet, W., Seymour, S., Biemer, S.: Systems Engineering. Principles and Practice, 2nd edn., p. 528. Wiley (2011)
4. Korolev, A.S.: Providing iterative development throughout the complex technical systems life cycle. Control Systems and Information Technology, №1.1(55), pp. 160–164. The Scientific Book publishing house, Moscow-Voronezh (2014)
5. Kozevnikov, D.E., Korolev A.S., Sazonov, B.V.: Competency profile and role of systems engineer in lifecycle management process for nuclear object. The Materials of Forth Research and Practice Conference Actuals Problems of Systems and Program Engineering, pp. 76–84. Higher School of Economics, Moscow (2015)
6. Korolev A.S., Alexandrov V.S.: Instrumental support for complex technical systems life cycle management. Control Systems and Information Technology, №2.1(52), pp. 137–144. The Scientific Book publishing house, Moscow-Voronezh (2013)
7. Borisov, A.N., Krumberg, O.A., Fedorov, I.P.: Decision-making Based on Fuzzy Models: Examples of Application, p. 184. Riga, Znanie (1990)
8. ISO/IEC TR 24748: Systems and software engineering—Guide to Life Cycle Management
9. ISO/IEC 15288: Systems and software engineering—System life cycle processes
10. Chapman, W.L., Bahill, A.T., Wymore, W.A.: Engineering Modeling and Design. CRC Press Inc., Boca Raton, FL (1992). Chapters 5 and 6

The Features of Generations of Solutions by Intellectual Information Systems

Stanislav Belyakov, Marina Belyakova, Alexander Bozhenyuk and Igor Rozenberg

Abstract The paper analyzes the characteristics of the informational support of decision-making by geographic information systems. The problem of the accumulation of experience and the use of decision-making in the previously observed situations is analyzed. The situations are spatiotemporal and they can be described by maps. The main objective of the research is development of the data model that provides the upgrade of reliability of decision-making on the basis of experience. The peculiarity of the model of the experience proposed by the authors is its description by a set of transformations. The concept of the image of the situation which has a center and a neighborhood is introduced. The allowed transformations of situations and solutions are determining in the description of decisions and the conditions of their making. The coordinates in the feature space are not determining. With such an approach traditionally used precedent analysis gets a peculiarity associated with the logic of determining the similarity of situations. The information model of precedents' image and the problem of actualization of the image in the process of searching for solutions are described in the paper. The example of figurative representation of the experience for the implementation of the logistics project is given in the paper.

Keywords Geographic information systems · Case analysis · Figurative representation of experience · Decision making

S. Belyakov (✉) · M. Belyakova · A. Bozhenyuk
Southern Federal University, Nekrasovsky 44, 347922 Taganrog, Russia
e-mail: beliacov@yandex.ru

I. Rozenberg
Public Corporation "Research and Development Institute of Railway Engineers", 109029 Moscow, Russia
e-mail: I.rozenberg@gismps.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_22

1 Introduction

Geographic information systems (GIS) are widely used in the systems of decision-making as facilities of time-space data processing [1–4]. An important advantage of GIS is the rendering engine, which gives a special opportunity of real-world situations' analysis. Using geo-information services, specialists of different application areas get spatial data and cartographic images of data. It stimulates their professional intellectual activity.

The map databases are extensive knowledge of events and phenomena of the real world. At the same time, this knowledge is not enough, when decisions must be made in specific situations. The dynamism of the real world and the incompleteness of knowledge does not allow to construct reliable solutions of problem situations. Despite the rich arsenal of software tools of spatial and statistical analysis [1], it is not possible to make compensation for the uncertainty of the environment's state characterization. For this reason, the usage of the experience of decision-making is more important.

An important feature of the presentation and usage of knowledge in the GIS is the ability of their visualization. Visualization allows you to simulate creative thinking [5]. Knowledge representation by iconic image-mapping models and the use of knowledge by their specific comparison can be considered as the implementation of the conceptual semantics principles [5]. The features of the procedure for constructing intelligent GIS solutions using figurative representation of decision-making experience are analyzed in this paper.

2 The Conceptual Model of Decision-Making Experience

The concept of decision-making experience representation can be selected in different ways, based on the mode of the intellectual system's usage [2, 3]. The systems operate in mode of precedent analysis in many circumstances [4]. Case-analysis [6] is based on superficial knowledge, limited description of the characteristic features of objects or phenomena. The process of finding the solution is finding previously observed similar precedent in the knowledge base and the adaptation of decision to the concerned precedent. Using the analogy, as known, gives probable but no reliable conclusions [7]. In an effort to improve the accuracy, we can intellectualize as follows.

Supposing that $D(s)$ there is a dependency that is used for generation solutions for a given set of environmental parameters s. If you have experience in decision-making $d_0 = D(s_0)$ with the parameter values s_0, the spreading of this experience to the problematic situation s_p with the parameters means obtaining the decision $d_p = D(s_p)$. The condition of the positive effect of usage of the solution d_p in a problem situation is the inequality:

$$W(d_p) \geq W(d_0) \tag{1}$$

where W is the criterion of solutions' quality. It is obvious that the applicability of the solution d_p—only a hypothesis, the accuracy of which is higher, s_p and s_0 closer. This follows from the continuity property of the real world. If

$$|s_p - s_0| \to 0,$$

inequality (1) becomes valid equality. It is observed in the surroundings s_p where

$$D'(s_p) = D'(s_0) = 0 \tag{2}$$

as it ensures no solutions' quality loss.

As the dependence $D(s)$ is not known, the Eq. (2) can be regarded as a conceptual basis for knowledge representation, which is the accumulation of knowledge about deviations (transformations) of situations' parameters and solutions. The knowledge of variations is an opportunity to construct a fair solution $D(s)$ in a neighborhood Δs of place s with a probability of

$$\sum_{s_k \in \Delta s} P(D(s_k)) > 0,$$

where as the absence of such knowledge is

$$\forall s_k \in \Delta s : P(D(s_k)) = 0.$$

The condition (2) can be regarded as a formal basis for the concept of knowledge representation in the form of the transformation of situations. Reflection of the happened events and analysis of the decisions taken are the experience of mental activity. This experience has a high value.

Let us call the character of precedent information structure that reflects the experience of mental activity that occurs during the analysis of the essence of precedent. Conceptual model of the precedent I_p includes two components [8]

$$I_p = <I_s, I_d>,$$

the first of which (I_s) is the image of the situation, the second (I_d)—the image of solutions.

The image of the situation I_s is a set of admissible transformations of the situation, the transformation does not change the essence of the situation and the decision-making. The image describes a class of situations that are identical in meaning with the observed singular situation. Possible conversions of the specific situation always contain a generalization. It is arguable that the fragment of "picture of the world" is laid in the image of the precedent. The "picture of the world" is very essential for getting a reliable solution.

The image of the solution is a set of admissible transformations of solution I_d, it preserves its essence. The image of solutions sets a class of solutions, each of which is applicable to essentially identical situations I_s. Any decision of the class is a certain "reasonable" reliable solution.

Images of situations and solutions can be visualized in the GIS. This feature plays an important part in the described approach. Cartographic representation of transformation of point locations, trajectories, zones of placement and zones of influence are point, line and area objects of the map. Reflecting the transformation of the specific situation, the expert transfers the knowledge to the map. The representation of knowledge gets metric properties, which allow you to estimate the position, the shape and the size of transformations.

3 The Internal Structure of the Images and the Modeling of Visual Thinking

Using the ideas of conceptual semantics [5], let us consider the process of generation of solutions as the comparison of images. This operation simulates visual thinking. Technical implementation of the operation requires the definition of the metric [9]. The metric of the distance between the images $N(I_1, I_2)$ should be built to take into account the subjectivity of experts' visual thinking. Subjectivism appears, on the one hand, in the individual interpretation of the present situation, on the other hand, it appears in the estimation of its modifications in the future. For this we represent the substructure of the image in the form of

$$I_s = <c, H(c)>,$$

where c is the center of the image $H(c) = \{h_1(c), h_2(c), \ldots, h_M(c)\}$ is the set of its transformations. The center is the real situation that served as the basis for the inception of the image. Let us explain this through the example. Figure 1 shows a precedent of moving a cargo from the point A to the point B. This precedent creates the image of the matter of moving goods from the one area of the village to the other (Fig. 2). The areas reflect the possible locations of points A and B, the points does not affect the selected path and possibly the method of transportation.

The center of the image is a pair of points A and B with the description of the goods and the method of its transportation.

Figure 3 illustrates the image of solution. It includes a center—the trajectory AB—and possible conversions of trajectory. The transformations are shown by dotted lines.

There are the factors that influence on the form of metric $N(I_1, I_2)$:

1. The relative position of boundaries and centers of conversions;
2. The degree of generality of conversions' fields.

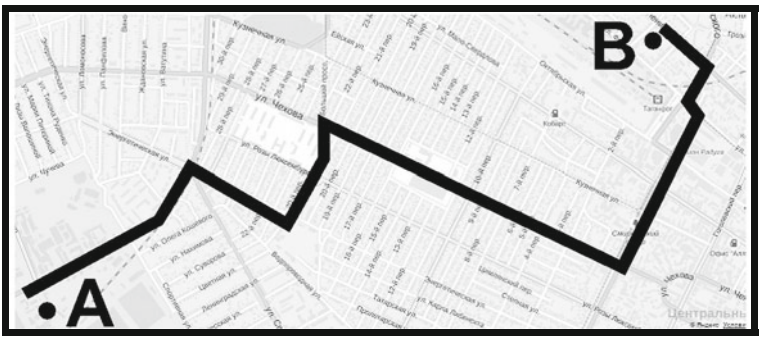


Fig. 1 The example of the precedent of moving goods

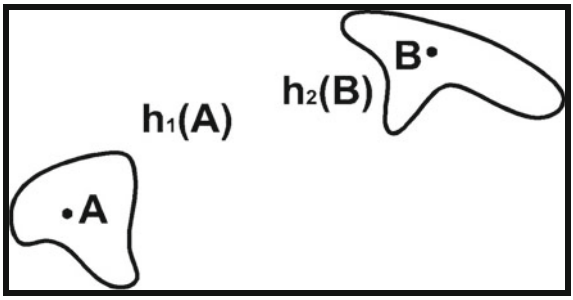


Fig. 2 The example of the precedent of moving goods

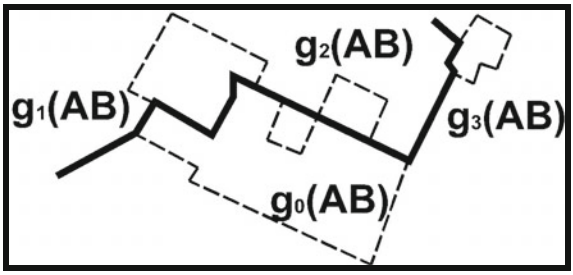


Fig. 3 Shows the image of decision of the precedent

The first factor is considered by using the proposed classification. Figure 4 shows the diagrams of the mutual arrangement of the pair of images. Ovals mean the areas of transformation, points inside ovals are the images' centers. The classes of topological relations arising during the comparison are labeled by $N_i(i = \overline{0, 5})$. The analysis of practical cases showed that the location of the centers of the images towards the intersection of transformations' areas is greatly influences on the subjective conclusion about the proximity of the situations. The distance between the centers is not important. The region of the intersection in which these centers are located is important. Class N_0 corresponds to the comparison of images without

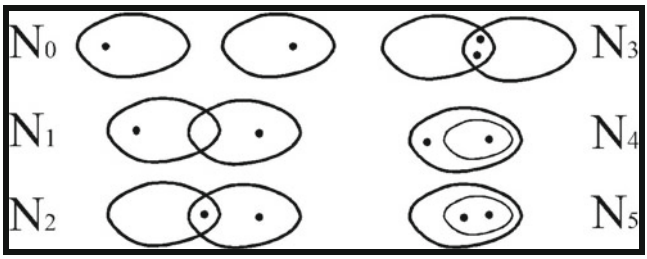


Fig. 4 The classes of the mutual position of the images

common variants of conversions. Class N_1—the presence of common changes that are not confirmed by practice so any of the centers is not included in them. The class N_5 includes a situation in which the conversion of one of the images is included in the conversion of another, and this is confirmed by the experience: the centers of the images are placed in the area of the intersection. It concluded that preferences when choosing the next image to the stated one would be described by the expression

$$N_0 \prec N_1 \prec N_2 \prec N_3 \prec N_4 \prec N_5$$

Thus, the proposed classification of topological relationships allows realizing a "reasonable" procedure of the assessment of vicinity.

The presence of the intersection regions in the analysis of changes suggests that there is some similarity of possible changes in the situation. The characterization of the degree of similarity of the transformation h_i of two images I_1 and I_2 is expressed by the formula

$$\alpha = \frac{2S(h_i^{(I_1)} \cap h_i^{(I_1)})}{S(h_i^{(I_1)}) + S(h_i^{(I_2)})},$$

where $S(x)$ is the area of the field x. The value $\alpha = 1$ holds in case of the coincidence of possible transformation, $\alpha = 0$ otherwise.

Comparative examples of situations for the determined problem is shown in Fig. 5. The proximity of the image I_1 is compared for a pair of points (A, B) with the images I_2 for a pair of points (C, D) and I_3 for a pair of points (E, F). The image I_1 is closer to the image I_3 although the centers A and C are located closer and the degree of generality $\alpha_{12} > \alpha_{13}$, because the relative location I_1 and I_3 is classified as a more preferable one.

Comparison of images of situations results in getting an image of solution. The center of image is practically tested solution. The example is shown in Fig. 3. The specific solution can be generated in several ways. In the simplest case, the solution is any transformation of the center of image. The result is reliable, as it is inferred from the generalized knowledge of possible solutions. However, the level of

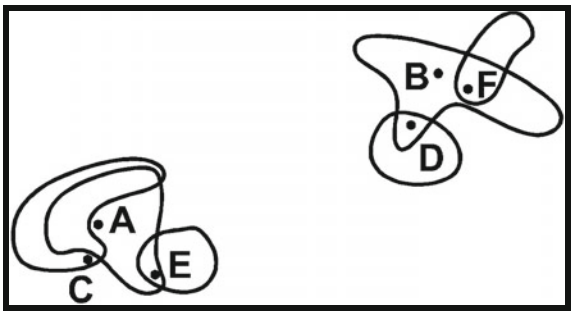


Fig. 5 The example of comparison of the situations

confidence of solution can be increased because the mapping basis cannot represent the real world at the time of analysis. The experience of reflection of expert about the situation cannot give an accurate forecast of the state of the environment. The reliability of the solution can be enhanced if the rapid mapping of the area of the problem situation is done. A comparison of the solutions with the state of the real world and the assessment of the feasibility of the solution is an independent task of intellectual GIS.

4 Conclusion

The analyzed graphic model of solutions' generation uses a special form of knowledge representation in the form of admissible transformations of situations and solutions. The purpose of the transition to this form of knowledge is the desire for a higher level of confidence of solutions generated by the intelligent system. The reliability is provided by the different quality of knowledge. Instead of sets of parameters of the observed situation the experience of reflecting on the situation is fixed. At the same time the real fact in not ignored; it becomes a key element of the image emerged. Using the key element has its own specifics, it reflects the psychology of creative thinking. The relative position of the centers and borders of transformation underlies the logic of matching images.

Analyzing the process of case analysis in relation to the processing of images, it can be concluded that the conservation of solutions generated in the process of case analysis is not appropriate. It is known that the accumulation of precedents is the implementation of training via examples. This mechanism does not involve such training. The main thing is the sense of situations and solutions expressed in admissible transformations.

Acknowledgments This work was supported by the Russian Foundation for Basic Research, Projects № 15-01-00149a, № 15-07-00185a.

References

1. Shashi, S., Hui, X.: Encyclopedia of GIS. Springer, LLC, New York (2008)
2. Zhi-Hua, H., Zhao-Han, S.: A decision support system for public logistics information service management and optimization. Decis. Support Syst. **59**, 219–229 (2014)
3. Fernandes, S., Captivo, E., Clímaco, J.: A DSS for bicriteria location problems. Decis. Support Syst. **57**, 224–244 (2014)
4. Haghighi, P.D., Burstein, F., Zaslavsky, A., Arbon, P.: Development and evaluation of ontology for intelligent decision support in medical emergency management for mass gatherings. Decis. Support Syst. **54**(2), 1192–1204 (2013)
5. Lakoff, J., Women, F., Dangerous, T.: What Categories Reveal About the Mind. University of Chicago Press (1987)
6. Eremeev, A., Varshavskiy, P.: Case-based reasoning method for real-time expert diagnostics systems. Int. J. Inf. Theor. Appl. **15**, 119–124 (2008)
7. Vagin, V.N., Yeremeyev, A.P.: Modelling human reasoning in intelligent decision support systems. In: Proceedings of the 9th International Conference on Enterprise Information Systems, Madeira, pp. 277–282 (2007)
8. Belyakov, S., Savelyeva, M., Rozenberg, I.: The construction of fuzzy based on case-based reasoning. In: 19th International Conference on Soft Computing MENDEL, Brno, pp. 273–276 (2013)
9. Belyakov, S., Belyakova, M., Rozenberg, I.: Approach to real-time mapping, using a fuzzy information function. In: Geo-Informatics in Resource Management and Sustainable Ecosystem, Wuhan, pp. 510–521 (2013)

Decision Assessment in Automated Design Intelligent Systems

Georgy Burdo and Boris Paliukh

Abstract The paper discusses the principles behind forming criteria systems for selecting intermediate decisions in automated design and control intelligent systems built on the design process decomposition principle. The decision assessment at the upper levels of design process decomposition is known to be complicated because of the implicit connection of the decision structure and parameters with the final level decision parameters. The article, on the example of the automated systems of machine-building technological processes design, shows the application of the method of assessing generalized (conceptual) decisions at the upper levels of the design process hierarchy by means of quantitative interval criteria. The selection of candidate decisions at intermediate levels allows dramatic possible decision reduction at the final level and thus automated design systems efficiency advancement.

Keywords Automated design systems · Artificial intelligence · Mechanical engineering · Criterion decision assessment · Design process decomposition

1 Introduction

In intellectual machine-building systems of the automated design (both of the design and technological orientation) and control it is expedient, for the universalization purpose, to use the design process decomposition principle. The essence of the principle consists in the hierarchical design process creation to ensure gradual specification of the synthesized decision structure and parameters. However, such approach realization is complicated by the decision variance at the decomposition intermediate levels, and, as a result, by the need to assess a considerable number of the synthesized decisions at the lowest level with a parametrical objective. Given the decision variance at every design process decomposition level, it is required, at

G. Burdo (✉) · B. Paliukh
Tver State Technical University, A. Nikitin emb., 22, Tver, Russia
e-mail: gbtms@yandex.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_23

its final level, to analyze a large number of final variants using parametric criteria. To limit unpromising decisions at the upper levels of design process decomposition in the automated mode is very difficult due to the lack of obvious connections between the structure and the parameters of the technical decision of the upper levels of the design process hierarchy (conceptual solutions) and the parameters of the objective. This fact leads to longer cycles of production design and technological preparation and to its quality decrease, as well as postpones the terms of new high technology engineering product launch. Therefore, in our opinion, this work is essential for mechanical engineering.

2 The Criteria of Decision Assessment at the Design Process Decomposition Levels

2.1 The Principles of Criteria Systems Formation

The criteria system at the intermediate design process (PP) decomposition levels in the CAM and CAD systems with artificial intelligence (II) elements is a necessary condition for selecting intermediate appropriate options [1, 2]. The criteria precision degree when proceeding to the next levels should increase, which allows object-oriented approaching the optimal overall decision. To improve the efficiency of the decision search procedures, several variants are chosen at each level as there is no mathematical proof of the fact that an optimal decision at the upper levels ensures the global optimum in general. The necessity to choose several decisions is also caused by criteria roughness. Therefore, the intervals of criteria values are chosen so that, on the one hand, not to increase the number of the options detailed further on, and on the other hand, not to lose the best decision, or a decision close to it.

The principles defining criteria formation were formulated:

1. Complexity: taking into account the major factors as well as these factors and the general task decision results interference.
2. The criteria must have the numerical appearance to allow revealing the connection of the common decision parameters and the decomposition level criterion value. The criterion must be sensitive towards the designed project (PO) change and must be proportional, that is, the increase in criteria intervals has to entail the increase in the object parameter intervals.
3. The criteria must reflect all PP levels, the main structural, parametrical and system PO properties at the levels and be modified in case of changes in the requirement specification.
4. The ultimate goal achievement and hierarchy principle: the subordination of intermediate criteria to the global criterion, the objective, that is, the consideration of level criteria as a single hierarchical system.

5. The criteria semantic content must correspond to the semantic contents of the decision of the level, which simplifies the designing control.
6. The criteria at the levels must provide decisions in gradually narrowing search spaces.

2.2 The Criteria of Intermediate Decision Assessment in Technological Process CAD

The enlarged scheme level. Most often at the first level of technological processe (TPr) design, during the enlarged scheme synthesis, the following criteria are used: the minima of production cost, $C_t \rightarrow min$; of total costs, $C_p \rightarrow$ min; execution time, $t_{op} \rightarrow$ min; the maximum of cutting technological productivity, $\mathrm{K}_t \rightarrow$ max. The criteria take into account the knowledge of parameters of the transitions synthesized at the final level. The use of crude data reduces accuracy. The main disadvantages are: no connection of the criteria and the actual condition in equipment loading technological divisions (TP); implicit connection of the criteria with the decision parameters: surfaces and details processing routes, etc. Taking into consideration that a bar stock choice in multiproduct production conditions is often limited by various profiled rollings, and processing stages allow only partial shift, the main criterion control parameter is surface processing routes (MOP) depending on a bar stock type and the reflecting peculiarities of stages structure (heat treatment type and place, procuring stage availability, etc.). Therefore MOP parts can be sufficiently objective assessment of the decision level.

Let us consider some characteristics—the power of the surfaces processing generalized route $M_O = \bigcup MOP_i = \bigcup \{PER_i\}_j$, where i is the processed surface number, j is the transition serial number; and the power of processing routes crossing is $M_p = \bigcap MOP_i = \bigcap \{PER_i\}$. Let us estimate the M_O and M_p sets according to the system transition parameter, that is, to a type of processing (a method of process impact). The number of elements in the sets is N_O/N_p, let us find their ratio $K_p = N_O/N_p$ and analyze it.

The N_O/N_p increase corresponds to the use of a larger number of different methods, and determines the TPr operation differentiation, the increase of the equipment setups and readjustments number, the charging of a larger number of machines. N_O/N_p reduction promotes the processing concentration due to the reduction of its various techniques, the reduction of the operation number, etc., as well as reflects the degree of technology standardization at an enterprise.

Let us estimate the M_O and M_p sets according to the system characteristic, that is, to the type of a tool used (the number of elements in the N_o^l and N_{p}^l sets). The $K_{\mathrm{i}} = N_{\mathrm{o}}^l \Big/ N_{\mathrm{p}}^l$ ratio directly characterizes the tendency towards the increase (decrease) of the number of tools used and their cost, and indirectly characterizes the

aspects mentioned above for TPr. The ratio characterizes the structural TPr components and corresponds to its level.

Let us consider the value, having designated it as K_{tp}, the coefficient of a TP condition:

$$K_{\text{tp}} = (\text{Z}_\text{T}/0,85)^{m_1},$$

where Z_T is a given load according to the types of equipment applied for the M_O transition set realization, $\text{Z}_\text{T} = \sum^{n} \text{Z}_{\text{p}i}$, $\text{Z}_{\text{p}i}$ is a given load of the ith machine type, n is their quantity; 0,85 is the standard load recommended for the enterprises of the mentioned type of scheduled works production; m is an exponent. A load is taken into account only for the limiting types of equipment. To take into account the impact of each type processing amount $\text{Z}_{\text{p}i}$ is calculated according to the formula

$$\text{Z}_{\Pi\, i} = (\text{Z}_i p_i)/p,$$

where Z_i is the actual load of the ith machine type at the present moment; p_i is the number of transitions carried out at this machine type; p is their total number by detail.

K_{tp} takes into account the equipment load by volume, characterizes the TPr readiness to work; the greater is its value, the more differentiated TPr is to be (in order to ensure uniform equipment loading and bottleneck prevention).

Let us consider the K_p and $(1/K_{\text{tp}})$ product, having designated it as K_1^1

$$K_1^1 = (\cup \text{MOP}_i/\text{MOP}_i) \times (1/K_{\text{T}\Pi}) = \left(N_\text{o}/N_\text{p}\right) \times (1/(\text{Z}_\text{T}/0,85)^{m_1}).$$

K_1^1 reflects TPr structure by processing methods, takes into account the expected distribution of these methods by machines, the change of the equipment load, which allows using it as a criterion. Let us consider $K_2^1 = K_\text{i} \times (1/K_{\text{tp}}) = \left(N_\text{o}^1/N_\text{p}^1\right) \times (1/(\text{Z}_\text{T}/0,85)^{m_1})$. K_1^1 describes the TPr structure in terms of the tool and the impact on the load by many foreseen transitions, i.e. it has a complex character, but is more clearly connected to costs than K_1^1 Therefore, the first criterion, $K_1^1 = E_1^1$, should be used, if a TPr objective is the productivity maximum, $Q \rightarrow$ max, and the second one, $K_2^1 = E_2^1$, should be used when a TPr objective is a prime cost minimum, $C \rightarrow$ min:

$$K_{1\min} \leq E_1^1 \leq K_{1\max},$$
$$K_{2\min} \leq E_2^1 \leq K_{2\max}.$$

If the criterion is the minimum cycle of detail manufacturing, $T_z \rightarrow min$, it suggests TPr differentiation into operations and parallel-sequential detail processing at working places. In this case the actual TP load is not taken into attention ($K_{\text{tp}} = 1$), and K_1^1 is shifted towards larger values; therefore the criterion is $E_3^1 = K_\text{p}$ and $K_{3\min} \leq E_3^1 \leq K_{3\max}$.

The values of dimensionless criteria $K_{1\max}$, $K_{2\max}$, $K_{1\min}$, $K_{2\min}$, $K_{3\max}$, $K_{3\min}$ are determined on the basis of expert evaluations, and refined during the working process. Their values are not given because of the limited amount of work.

The routing technology level. At the routing technology level, the following criteria are applied: operation performance maximum, $Q_{op} \rightarrow \max$, reduced costs minimum, $C_p \rightarrow \min$, technology prime cost minimum, $C_p \rightarrow \min$, and labour content minimum (the time calculated by detail), $t \rightarrow \min$. The criteria are approximate, as there are no exact parameters that allow calculating them, and there is no obvious correlation between these parameters and TPr structure and the criteria.

Let us consider the $K_o = k/q$ ratio, where k is the number of operations in a route, q is number of machining stages. The K_o increase leads to TPr differentiation, the K_o reduction, to concentration, that is, it characterizes the main structural TPr components, and its productivity.

Let us consider the formula $K_{\mathrm{d}} = \left(L_p/L_{\mathrm{d}}\right)^{m_2}$, where L_{d} is the optimum production run quantity defining TPr differentiation; L_p is the actual production run quantity. With the increase in the L_p lot it is advisable to increase TPr differentiation, and vice versa. At this level the machines group within a type are already specified, therefore let us introduce the value $K_{\mathrm{tp}}^2 = \left(\mathrm{Z_g}/0,85\right)^{m_1}$, where $\mathrm{Z_g}$ is the given load by equipment groups within a type, $\mathrm{Z_g} = \Sigma 3_{\mathrm{gp}i}$, 3_{gpi} are the given load of the ith group, $\mathrm{Z}_{gpi} = \mathrm{Z}_i p_i/p$, Z_i is the actual load of the ith machine group.

Let us consider $K_1^2 = K_0 \times \left(1/K_{\mathrm{tp}}^2\right)$. The K_1^2 value reflects the designed technology structure and the TPr condition. Therefore $K_1^2 = \frac{k}{q} \times \left(1 \Big/ \left(\mathrm{Z_g}/0,85\right)^{m1}\right)$ can be used as criterion for E_1^2, if a TPr objective $Q \rightarrow \max$ (the index above means the level): $K_{1\min}^2 < E_1^2 < K_{1\max}^2$.

Let us analyze $K_2^2 = E_2^2 = K_{\mathrm{o}}\left(1 \div \left(K_{\mathrm{tp}}^2 K_{\mathrm{d}}\right)\right)$. Kd takes into account the economic aspect of the technology.

Then, with $K_{1\min}^2 = K_{2\min}^2$, and $L_p < L_d$ ($\mathrm{K_d} < 1$), maintaining of the $K_{2\min}^2 K_{2\min}^2 < E_2^2 < K_{2\max}^2$ relations leads to K_{o} reduction, that is, to the processing concentration and vice versa. The criterion is effective when a TPr objective $C \rightarrow \min$. In accordance with the said above, with a TPr objective $T_z \rightarrow \min$, $K_{3\min}^2 < E_3^2 < K_{3\max}^2$, and the criterion $K_{3\min}^2 < E_3^2 < K_{3\max}^2$, but the boundaries are shifted toward higher values. The boundary values are defined by the same reasons as the first level criteria, and take into account the specificity of numerical program control (CNC) equipment. A choice of the compromise criteria (in the form of a product) scheme reflecting the principle of a fair relative concession and saying: "fair is considered the compromise at which the total relative level of the quality decline of one or several criteria does not surpass the total relative level of quality improvement according to other criteria", is caused by the following reasons. Criteria are based on the analysis and comparison of the local criteria change values which, taking into consideration a compromise, is inevitable. The criteria

formation according to the fair relative concession principle provides local criteria level smoothing. It is also important that the method is not sensitive towards the criteria change scale.

The process technology level. At the level of process technology design it is usually recommended to use the criteria [3–6] of incomplete piece t'_{hT}, incomplete auxiliary $\sum t'_{\text{vp}}$ time, the prime costs of processing the whole of transitions $\sum C_{\text{p}\,i}$. The wish to find correspondence between the parameters of the global decision estimated by the objective of Q or C is obvious. However, the absence of the data on the cutting modes, on the tool used (these are defined later) complicates calculation, or leads to iterative procedures; criteria do not reflect the TPr condition either.

Let us consider the formula for the incomplete piece time that takes into account only changeable operation performance timetable for each structure option:

$t'_{\text{hT}} = \left(\sum \left(l_{\text{p.x}\,iv}/S_i\right) + \sum \left(l_{\text{x.x}\,jv}/V_{\text{x.x}j}\right) + t_{\text{c.i}\,k}\right)_V$, where $l_{\text{p.x}\,i}$, $l_{\text{x.x}j}$ are the lengths of cutting strokes and non-cutting strokes respectively; S_i, $V_{\text{x.x}j}$ are the values of cutting strokes movement and the non-cutting strokes speed respectively; $t_{\text{c.i}\,k}$ is the time for the k-th tool change; v is the a variant number, i is a transition number. The first member in the formula characterizes the main timetable, two others, the auxiliary one. Let us consider the connection of an equipment type with the t'_{hT} criterion use possibility. For hand-operated equipment the cutting stroke length and tool change sequence are defined with the operation structure synthesis, and non-cutting strokes are randomly defined by a worker, therefore the total length of cutting strokes will be the only objective criterion. CNC equipment should be considered according to types. For CNC turning equipment, the cutting strokes costs assessment is more important since the shortest non-cutting strokes are chosen, their speeds are an order of magnitude bigger than the speed of cutting strokes movement. Cutting strokes define the tool choice and the time of its change, but, having in mind its smallness in comparison with processing, the time of its change can be ignored. For mill-drill-bore CNC machine the situation is different. Volumetric milling processing is based on the basis of standard contours or with the software use, and the processing time costs for a given detail are constant. The main time of surface processing by drilling, boring, reaming, screw cutting, etc. is defined by the surface length, therefore, these time costs for a given detail are constant. For this equipment type the structure optimality is defined by the time of transition from one surface processing to another surface processing.

To account for a TP condition, and having in mind that the equipment model is defined at this level, we will introduce the $K_{\text{Z}q} = \left(\text{Z}_q/\text{Z}_{\text{TO}}\right)^{m_3}$ criterion, where Z_q is the loading of the q-machine from a chosen group; $\text{Z}_{\text{TO}} = \sum \text{Z}_q/K$ is average loading of K machines of a group. From here, we get the following E^3 criteria. For multiwork machines with the objective $Q \rightarrow \max$, designating

$\forall v,\ q\left[\sum^{m} l_{\text{p.x}\,iv} \times K_{\text{Z}_q}\right]_{\min} = E^3_{1\min}$ and $\left[\sum^{m} l_{\text{p.x}\,iv} \times K_{\text{Z}_q}\right] = E^3_1$, we have $K^3_1 \times E^3_{1\min} \geq E^3_1 \geq E^3_{1\min}$, where v is an option number; K^3_1 is a coefficient, recommended at the first stage to be equal 1,3 … 1,4.

If $C \rightarrow \min$ is used as an objective, then assuming the equipment and tool costs to be proportional to the cutting stroke lengths, we will receive the total outlay value:

$3_{ql} = \left(\sum^{m} l_{\text{p.x}\,iv}\right) C_{\text{c}q} + \sum^{n}\left(\left(\sum^{p_i} l_{p.xj}\right)_l\right) C_{\text{i}l}$, where C_{cq} is the cost of machine-hour of the q-th machine without tool cost; C_{il} is the cost on the lth tool ($l = 1, n$), related to the 1st hour of cutting time; $l_{\text{p.x}jl}$ is the jth cutting stroke of the lth tool, $j = 1, p_i$; $\sum^{n}(\sum^{p_i} l_{\text{p.x}j})_l = \sum^{m} l_{p.xiv}$; $n \times p_i = m$ is the total cutting stroke number. Designating $\forall q,\ l(Z_{ql})_{\min} = E^3_{2\,\min}$, we get the E^3_2 criterion: $K^3_2 \times E^3_{2\min} \geq E^3_2 \geq E^3_{2\min}$. It is possible to accept $K^3_2 = K^3_1$. If $T_z \rightarrow \min$ is taken as an objective then equipment loading is ignored. Designating $\sum^{m} l_{\text{p.x}\,iv} = E^3_3$ and $\forall v\left(\sum^{m} l_{\text{p.x}\,iv}\right)_{\min} = E^3_{3\min}$, we have $K^3_3 \times E^3_{3\min} \geq E^3_3 \geq E^3_{3\min}$. K^3_3 can be taken equal to $K^3_1 = K^3_2$. For CNC turning equipment, the criteria coincide with those of hand-operated machines,

$$E^3_{4T} = E^3_1(Q \rightarrow \max), E^3_{5T} = E^3_2(C \rightarrow \min), E^3_{6T} = E^3_3(T_z \rightarrow \min).$$

It is known that for expensive equipment (mill-drill-bore CNC machines) $Q \rightarrow \max$ and the $C \rightarrow \min$ objective define close operation structure and parameters, therefore in these cases the $E^3_7 = E^3_8 = t'_{vp} \times K_q$ criterion is expedient Designating $\forall v, q(t'_{vp}K_q)_{\min} = E_{7,8\min}$, we receive $K^3_3 \times E^3_{7,8\min} \geq E^3_{7,8} \geq E^3_{7,8\min}$. If an objective is $T_{\text{я}} \rightarrow \min$, then $E^3_{9\min} = t'_{vp\,\min}$, and $E^3_{9\min} = t'_{vp\,\min}$.

The transition parameter analysis level. At the level of manufacturing transition parameter finding the criteria, except for $T_z \rightarrow \min$, are investigated rather thoroughly [4, 5, 7–10], including those of multicriteria optimization. We are given the functional relation of Q and C and the parameters of the execution of operation and transition (cutting modes, time losses of different types, losses of cash, etc.). For the right choice shift organization, tool sharpening tool at a given enterprise are important. In general terms the K^4_1 criterion can be presented as follows:

$$K^4_1 = Q = f\left(\left\{l_{\text{p.x}\,i}\right\}, \left\{\left\{\text{PR}_j\right\}_i\right\}, \left\{C_j\right\}, t_w\right) \times K_{3q} \rightarrow \max,$$

where $\left\{\left\{\text{PR}_j\right\}_i\right\}$ is a set of cutting modes parameter at the i-th cutting strokes; $\left\{l_{\text{p.x}\,i}\right\}$ is a set of cutting stroke lengths; $\left\{C_j\right\}$ means time losses depending on $\left\{\left\{\text{PR}_j\right\}_i\right\}$; t_w is extracyclic losses not depending on $\left\{\left\{\text{PR}_j\right\}_i\right\}$. The K^4_2 criterion is then $K^4_2 = C = f_2\left(\left\{l_{\text{p.x}\,i}\right\}, \left\{\left\{\text{PR}_j\right\}_i\right\}, \left\{C_{\text{c}q}\right\}, C_{\text{i}l}\right) \times K_{3q} \rightarrow \min$.

If the objective is $T_z \rightarrow \min$, then we have to pay attention to the criterion of the maximum process productivity (cutting productivity), K_{pi}, at the ith cutting stroke. Let us note that the criterion use suggests centralized tool sharpening and adjustment, without a worker being distracted. For each transition or cutting stroke $K_3^4 = K_p$, then $K_3^4 = \sum \left((n_i \times s_i)/l_{\text{p.x}\,i}\right) \rightarrow \max$, where n_i, s_i are a spindle turn number and feed per revolution respectively. If such an event is not provided, then $K_3^4 = Q \rightarrow \max$ is chosen as a criterion.

2.3 The Procedures of Criteria Application Experience Accumulation and Generalization

In accordance with the requirements imposed on intelligent systems, criteria have to be specified on the basis of designing experience. Having in mind that the criteria presented are functionally dependant on the TPr parameters of the decomposition level being analyzed, as well as on the process divisions parameters, this requirement implementation does not cause difficulties. Let us concretize the process mechanism. The criteria have to be specified on the basis of simulation modeling (IM) or assessment of the results received during the process. With IM the following aims are pursued: specifying the criteria boundaries at each level; defining the number of the options left at each level; establishing each criterion value change influence on the technical-and-economic indexes value change. The IM procedure function is

$F : \left\{E_i^j\right\} \times \{CX_k\} \times \{TE_l\} \rightarrow \left\{\Delta \mathrm{E}_i^j\right\}, \{TE_l\} \leftrightarrow \left\{\Delta \mathrm{E}_i^j\right\}$, where $\left\{E_i^j\right\}$ is the sets of the criteria set; $\{\mathrm{CX}_k\}$ is the TPr system characteristics set, $\{\mathrm{TE}_l\}$ is the TPr technical-and-economic indexes set (C, Q, T_z); $\left\{\Delta \mathrm{E}_i^j\right\}$ is the criteria change corresponding to $\{\mathrm{TE}_l\}$. The TPr performance result assessment in a production system pursues the same aims. Because receiving a number of the actual TE is complicated, the analysis should be made on the basis of expert TPr quality evaluation, comparing that with the criteria value at the design process decomposition levels [6, 11].

The essence of the algorithm is as follows. 1. Details are grouped together on the basis of their structural and technological features. 2. For specific details, their technological decisions and detail parameters determining them (their information models) are stored. 3. Decisions are evaluated on the basis of their realization in production. 4. The detail parameter value intervals, defining identical decisions, are clarified and estimated. 5. Decision and parameter intervals are knowledgebased. Further on, an engineering decision of any level is chosen by comparing the parameter values of a new detail to the parameter value intervals in the knowledgebase. CAM system performs parameter intervals accumulation and determination, while their assessment and generalization are carried out with the participation of experts.

3 Conclusions

This approach, reflecting the features of the automated design-engineering production preparation in multiproduct mechanical engineering of TPr process control and management designing has been realized when developing TPr CAD systems for enterprises of the town of Tver. According to the results of the system trial operation we can note the sufficient quality of the designed TPr, their compliance to the objective of production method and conditions.

Acknowledgments The work is done with funding from RFBR project 14-10-00324.

References

1. Burdo, G., Paliuh, B.: Optimizacija procedur poiska technologicheskih reshenij v kompleksnoj SAPR TP—ASUTP. Vestnik Izhevskogo gosudarstvennogo universiteta. **3**(47), 109–112 (2010)
2. Burdo, G.: Issledovanie procedur proektirovanija technologij v sistemah avtomatizirovannogo proektirovanija I upravlenija technologicheskimi processami. Vestnik Izhevskogo gosudarstvennogo universiteta. **4**(48), 109–1123 (2010)
3. Kapustin, N.: Avtomatizirovannaja sistema proektirovanija technologicheskih processov. Mashinostroenije, Moscow (1979)
4. Tsvetkov, V.: Sistemno-structurnoje modelirovanije i avtomatizacija proektirovanija technologicheskih processov. Nauka i technika, Minsk (1979)
5. Yevgenev, G.: Intellektualnye sistemy proektirovanija. Izdatelstvo MGTU im. Baumana, Moscow (2009)
6. Domigo, M., Maniezzo, V., Colorni, A.: The Ant System. Optimization by a colony of cooperating objects. IEEE Trans. Syst. Man, and Cybern.—Part B **26**(1), 29–41 (1996)
7. Watson, G.H.: Strategic Benchmarking: How to Rate your Company's Performance Against The World's Best. Wiley, 270 pp. (2003)
8. Allan, J.J., (ed.) A Syrvey of Commercial Turnkey CAD/CAM Systems, 2nd (edn.). Productivity International, Inc., Dallas, Tex (1990)
9. Groover M.P.: Automation, Production Systems and Computer-Aided Manufacturing, Ch. 10. Co Prentice-Hall, Inc., Englewood Cliffs, NJ (1985)
10. Amirouche, M.L.: Computer-Aided Design and Manufacturing. Co Prentice-Hall, Englewood Cliffs (1993)
11. Holsheimer, M., Siebes, A.: Data mining: the Search for Knowledge in Databases. Technical report CS-R9406, CWI, 78 pp. (1994)

Intelligent Decision Support Systems in the Design of Mobile Micro Hydropower Plants and Their Engineering Protection

Denis V. Kasharin

Abstract The article tackles issues related to the creation of an intelligent decision making system for the design of mobile micro hydroelectric plants and their engineering protection for a rough terrain and decentralized water and power facilities of low power consumption. This system is based on multi-criteria optimization and provides the most optimal option to solve the given issue when the decision maker (DM) cannot choose from a range of project solution options.

Keywords Intelligent decision-making system · Multi-criteria optimization · Mobile micro hydroelectric plant · Composite materials · Facilities engineering protection · Decision maker

1 Using Mobile Micro Hydroelectric Plants and Facilities Engineering Protection Made of Composite Materials

The territory of the Russian Federation which is currently deprived of centralized water and power supply is estimated at 70 %. Half of the railways are not sufficiently powered and rely on reserve sources for power supply and station water supply. This is particularly a pressing matter for the mountainous conditions of North Caucasus region, where landslides damage power lines, water pipes, etc.

That is why, it is necessary to provide reserve decentralized water and power supply. The most promising option is to use water resources of small mountain rivers. However, building derivational micro-hydroelectric plants on mountains involves considerable difficulties associated with dangerous geological and meteorological phenomena, along with the lack of road infrastructure around the construction sites of water intake facilities and pipelines. We have developed a new solution for small mountain river conditions, namely the mobile micro-hydroelectric plant (MMHEP) for seasonal activities and engineering protection facilities from

D.V. Kasharin (✉)
OOO «Impulse», 136428 Novocherkassk, Rostov Region, Russian Federation
e-mail: dendvk1@mail.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_24

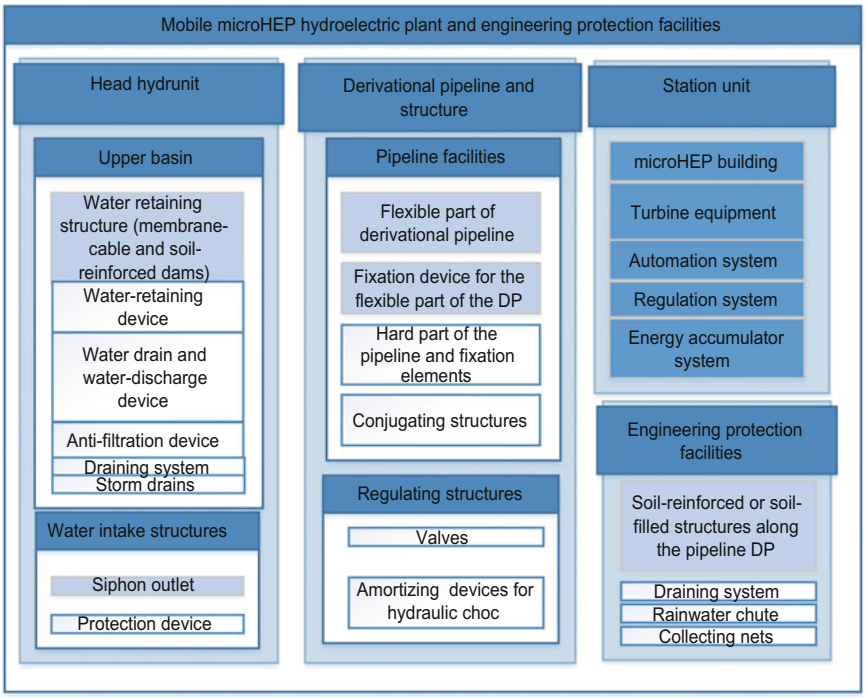


Fig. 1 Components of the MMHEP (Non-typical elements have been highlighted)

composite materials. Its advantages over permanent derivational hydropower plants include: construction without the use of heavy construction equipment, and using local building materials; reduction in construction costs and improvement in its reliability through the use of shell structures made of composite materials; repeated use; less negative impacts on the environment [1–3].

The structural schematics of the MMHEP and engineering protection facilities specifying non-typical elements and their variations are presented in Fig. 1.

Non-typical elements include:

- Water retaining structures for the Main Hydrounit (MHu) of the MMHEP (membrane-cabled and soil reinforced dam)) [3];
- Flexible land for derivational pipelines (DP) reinforced with a cable system [2];
- Engineering protection facilities (reinforced soil retaining walls; closed or open ground-filled structures) [1].

The main elements of these buildings are closed and unclosed shells (membrane) of composite materials, as well as cabling systems [3].

A membrane-cable dam is used for pressures up to 5 m when local building materials are absent and consists of three main elements: an open water-retaining shell; a base; a cable system. Simulation modelling (SM) of the membrane-cable dam is only possible through iteration approximation, analytic calculation and numerical simulation of the membrane-cable dam.

For heavy duty and for pressures of more than 5 m, ground-reinforced dams are sued, which are characterized by greater stability and reliability as compared to membrane-cable dams.

One of the main elements of the MMHEP is an integral derivation pipeline (DP). Its application in complex terrain provides a greater range of pressures ranging from 10 to 500 m and a capacity of 5–100 kW. The upper portion of the DP for pressures up to 50 … 60 m is made of flexible membrane shells on a water or soil-filled base. Standard rubber sleeves and pipes made from polymers are used for the lower portion. [2].

The upper parts of the pipeline (derivational pipeline) DP include a flexible water-filled membrane (hereinafter water-filled shell) made of composite materials, including water-filled or ground-filled bases for pressures of 50 … 60 m, according to the analysis of the calculated provisions considered in the first phase of the report.

The main difference between flexible water-filled membranes DP and hard ones is the relationship of their form with the internal and external loads and other boundary conditions. Therefore, closed water retaining shells are calculated considering changes in their shape, depending on: the impact of hydrodynamic flux on the water-filled shell DP; fixation conditions and the characteristics of the shell base. To reduce the twisting effect of flexible DP with a single-shell structure under transverse circulation and to provide sufficient stability and reliability of operation, a multiple-shell design can be used. The fixation of the flexible portion of the DP with a cabling system is provided when installing the DP on landslide areas, as well as for slopes greater than 45° along the pipeline. The cable system provides conduit consolidation in areas of high-power alluvial deposits and landslides [2–4].

The MMHEP design by a team of insufficiently qualified people considering the difficulty of the calculations with non-typical elements for large displacements under mountain terrain conditions and dangerous geographical and meteorological phenomena may lead to critical errors.

Also the decision maker (DM) should choose an optimal option among a wide range of options taking into account the criteria of suitability, optimality and adaptivity. As a result of a decision from the DM on the basis of an informal method based on insufficient knowledge and experience can lead to a suboptimal choice and inappropriate options for the MMHEP [1, 5].

Hence, it is necessary to develop a system to make decisions based on a multi-criterial mathematical model to optimize the MMHEP and intellectual system of expert evaluation.

2 Developing an Intelligent Decision Making System to Design a MMHEP from Composite Materials

2.1 Multicriterial Mathematical Model to Optimize MMHEP from Composite Materials

The mathematical model (MM) is designed to provide an optimal layout and parameters for the elements of the MMHEP, which provide the best technical and economic performance and reliable operation during planning.

When creating a MMHEP we use a MM with a systematic approach. A systematic approach will integrate simulation (SM) results, non-standard elements of MMHEP numerical methods, in proportion to their importance, into a single optimization model [2]. In this approach non-typical MMHEP elements are composed of individual subsystems, which are then integrated into their general structure, taking into account the choice of standard elements. For example, water retaining structures are elements of the HHu MMHEP and include a sub-system of elements: shell, cable systems, base (drooping), apron, etc. Each subsystem has its own objective functions, selection criteria and range of options. This approach means that each system is an integrated whole, even when it is disconnected from individual subsystems.

A generalized structure of the MM of the MMHEP optimization is presented in Fig. 2.

Based on a systematic approach, the optimization structure of the MM of the MMHEP suggests that the creation of a MM has a common goal—G, which generates a series of initial requirements—R and selection criteria—SC. On the basis of the initial requirements, a separate subsystem—S is created, from which a group of elements—E is formed. The final MM is a set of interrelated elements, obtained on the basis of selection—V on pre-formed criteria. Baseline—B is determined at the stage of initial requirements and additionally at the formation of model elements, thus creating a model made from general to specific.

To create a multicriterial optimization MM for the MMHEP we select a multicriterial optimization method for MMHEP parameters.

Justifying the selection of the multicriterial optimization method for the parameters of the MMHEP.

In order to select the multi-criterial method to optimize the parameters of the MMHEP, we shall analyze the existing methods which can be used depending on the supplementary information about the preferences of the decision maker (DM): probing the parameter space; excluded preferences of the DM; priori methods; posteriori methods; Interactive (adaptive) methods [5–7].

The method of parameter space probing is based on the systematic examination of multidimensional areas at uniform probing points, for example, the LPT sequence, in the variable space. The precision of the method is directly proportional to the number probing points [6]. Its disadvantage is the use of linear dependencies, while target functions of the elements of the MMHEP mainly correspond to the nonlinear dependencies.

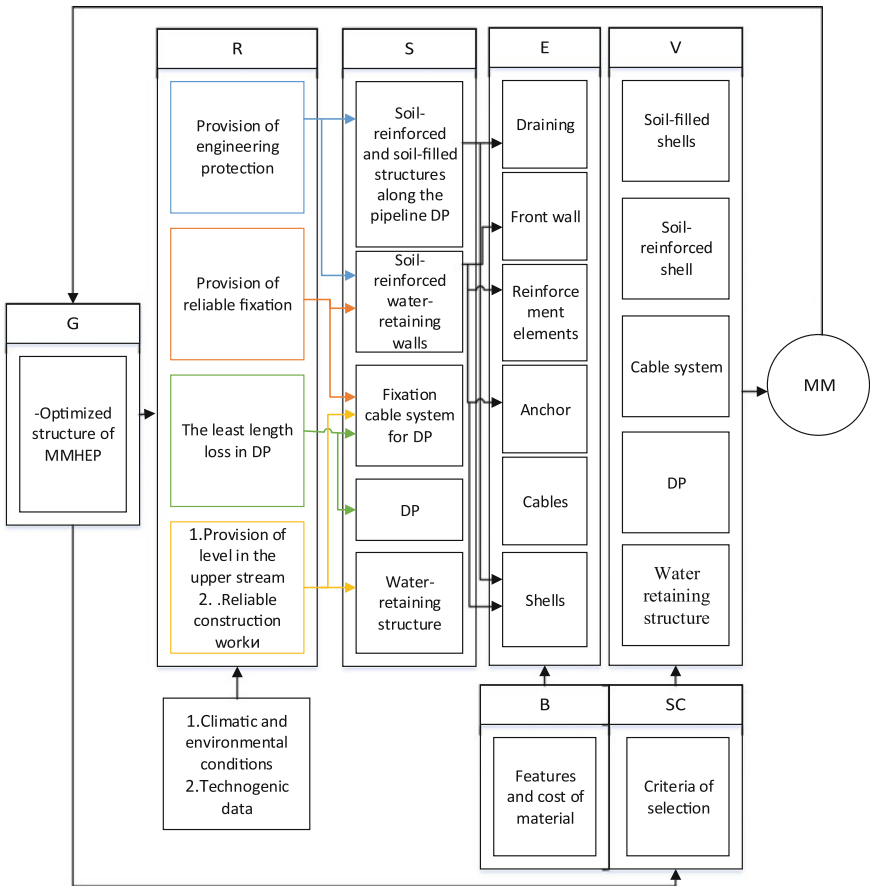


Fig. 2 Generalized structure of optimization MM for the MMHEP

Methods which exclude the preferences of the DM, the global criteria and the neutral compromise solution are meant to find a compromise solution [7]. This method cannot be used, since for a complex system, the probability of getting a clear (formal) solution is very low.

Posteriori methods include the participation of the DM in the multicriterial optimization (MCO) system which is to provide his preferences in the information system after receiving a set of non-dominant solutions. In these methods, an approximated Pareto set is built first, which are effective in optimizing with a large number of criteria [7]. The main disadvantage of posteriori methods is the need for large computation expenses in the uniform approximation of a set and/or Pareto front, which are not computable by a personal computer (PC), available to the MMHEP designers. Also, to improve the accuracy of approximation it is necessary to increase the number of non-dominant solutions, which is more time consuming

for the DM. Also it is difficult to visualize the Pareto front for tasks with a large number of criteria of more than two.

Priori methods do not require building a whole set of solutions [7]. Prior to the decision the DM provides information about his preferences, after which he formalizes them, transforming them from a multi-criterial into a mono-criterial one. These methods include the scalar convolution techniques, limitations and lexicographical ordering, and targeted programming. The main disadvantage is that it is difficult for the DM to articulate his preferences before starting the design of the MMHEP.

Interactive methods consist of a set of iterations, each of which includes an analysis step performed by the DM, and a step of calculation performed by the MCO-system [7]. Depending on the nature of the information obtained from the MCO-system the DM isolates classes of interactive methods in which the DM directly assigns weights coefficients to partial optimization criteria in the analysis phase; limits the values of the partial optimization criteria; evaluates alternatives proposed by the MCO-system.

It was decided to use a priori method such as scanning with a variable pitch and random directions. Since the DM may not always pre-identify his preferences for a technical decision for the MMHEP, even with a known range of parameters, in the future we plan the remote use of an intelligent expert system. Optimization using this method is the least demanding for PC resources. As a basis for the modification of the scalar convolution we use the so-called Method of weighted sums. The method consists of converging multi-criterial tasks to mono-criterial ones by replacing the criterion of vector optimality, consisting of several partial criteria, with one generic criterion, called the objective function. In this case, if there are multiple partial target functions represented in the form:

$$\mathrm{CF}_1 \rightarrow \min; \mathrm{CF}_2 \rightarrow \min \ldots \mathrm{CF}_n \rightarrow \min, \tag{1}$$

they converge to find the minimum of the generalized scalar target function in a general form:

$$\mathrm{CF}_{\sum} = \alpha_1 \frac{\mathrm{CF}_1}{\mathrm{CF}_{1.st}} + \alpha_2 \frac{\mathrm{CF}_2}{\mathrm{CF}_{2.st}} + \cdots + \alpha_n \frac{\mathrm{CF}_n}{\mathrm{CF}_{n.st}} + \mathrm{Shtraf} \rightarrow \min \tag{2}$$

where $\alpha_1, \alpha_2 \ldots \alpha_1, \alpha_2, \ldots, \alpha_n$—are the weight coefficients for partial target functions; $\mathrm{CF}_{1.st}, \mathrm{CF}_{2.st} \ldots \mathrm{CF}_{n.st}$—are the values of the partial target functions for the reference state of the MMHEP (these values are required to use the scalar convolution for dimensionless values); penalty function, which takes a large value some restriction on the parameters of the MMHEP is ignored and zero when all restrictions are executed:

$$\mathrm{Shtraf} = \begin{cases} 10^{10} & \text{upon not execution of the restriction;} \\ 0 & \text{upon execution of the restriction.} \end{cases}$$

For example, in the invariant form, the MM membrane-cable dam as a managing system of target functions, as per Eq. (1) can be expressed as follows:

$$\begin{cases} h_{up}/L = f(\varphi_{os}; \varphi_b) \rightarrow \min; \\ C_c = f(N; L; B; R) \rightarrow \min; \\ N_C/B = f(R_C; f; N_C) \rightarrow \min. \end{cases} \quad (3)$$

where h_{up}/L—is the relative depth from upstream to the shell perimeter L; N_C/B—is the ratio of the tension per unit length to overlap span; C_c—the relative cost of membrane cable dam as compared to a concrete dam; R_C—the radius of the cabling system of the membrane-cable dam; φ_{os}, φ_b—the attachment angle and base for the open shell of the membrane-cable dam.

The following invariant was obtained for the soil-reinforced dam:

$$\begin{cases} 1 - \left[H_{hsm}/\left(\sum l_{lr} + L_{efw}\right)\right] = f\left(h_{up}; B; C_p\right) \rightarrow \min; \\ C_C = f\left(N; \sum l_{lr} + L_{efw}\right) \rightarrow \min; \\ N/N_{\text{cr}} = f\left(l_{lfa}; H_{hsm}\right) \rightarrow \min, \end{cases} \quad (4)$$

where H_{hsm}—the height of the soil mound of the water retaining structures; l_{lr}—length of the reinforced tape; L_{efw}—length of front wall; l_{efa}—length of the reinforced tape sealing in the first approximation; C_p—the relative cost solid-reinforced dam.

In accordance with the invariant forms (3, 4), we make up a sequence of calculation for an optimization model of the water retaining shell membrane cable dam and its optimal parameters.

Creating a program for the multi-criteria optimization of MMHEP parameters.

To create a program for the multi-criterial optimization of MMHEP parameters, involving various MM in combination with man-machine selection procedures for the best solutions based on the assumption of the existence of a generalized criterion for the problem of multi-criterial optimization. This approach is best when it is difficult for the decision-maker (DM), to assess the contribution of partial criteria in an integrated indicator, as well as in the event of a decision in accordance with existing database on the computer and controlled by the DM [8, 9].

When solving the problem of choosing the optimal parameters of MMHEP several options of alternative solutions for the layout of elements, parameters made up of closed and open shells from composite materials, characteristics of different aggregates, are compared with traditional elements.

The structure of the multi-step optimization of the MMHEP is built upon an aggregate principle and consists of five levels [9]:

- 1st level—analysis and optimization of source data, selection of criteria and limits for elements of the MMHEP and engineering safety facilities;
- 2nd level—creation of MM elements of the MMHEP;

- 3rd level—mono-criterial optimization of MMHEP parameters for single target functions of MMHEP elements;
- 4th level—multi-criterial optimization of MMHEP parameters on multi-target functions of MMHEP elements;
- 5th level—final calculations of the selected facility (structure) and its elements using the developed calculation methods.

During the first stage hydrological, hydro-rheological and morphological characteristics of the water body are defined; its location. The basic criteria and constraints used to select the layout and parameters of the facility (structure) are established.

The second step is the choice of the pipeline DP and its layout thus produced in two ways. The first method is based on rapid—calculation using analytical functions to select the most optimal facilities. The second method involves engaging intelligent expert systems to select elements, it uses the database formed on the computer and given by the SM which is constantly updated upon decision.

The final choice of the parameter values of a certain element of construction is possible only after the parametric optimization on the 3rd and 4th levels, because the data obtained during the optimization of the parameters may be significantly different from the preliminary express calculation results.

To select a final design decision at the 4th level of optimization, justified in the previous steps, the parameters of construction are defined with the use of software modules based on SM using Ansys 15, Simulation Floworks (SolidWorks app), Plaxis. At this level, the final choice is made for the technical solutions of the non-standard elements of MMHEP, selection of model and final calculation parameters of the MMHEP [1–4, 8, 9].

At the fifth level a final calculation is made for the selected facilities (structure) based on the penalty function and its output parameters for specific installation conditions.

It was created with Delphi 7 programming language program “Optimization modeling parameters of MMHEP” not requiring computing resources of the PC and can be used in the design of MMHEP. However, at the stage of formation of the DM’s preferences, with insufficient design experience, blunders can happen. In this connection it is necessary to develop an intelligent expert system (ES), which can be used at the stage of preference at the initial stage of decision making.

2.2 *Justification of the Requirements for an Intelligent Expert Evaluation*

The basis of intelligent system ES should be a knowledge base, developed from knowledge acquired from the subsystem, controlled by an engineer and corrected by an expert, thus creating solution selection rules.

Structuring in knowledge bases occurs in the form of models and is subdivided into universal and specific problems [10]. The structure of each model also contains simulations and empirical data obtained from design and operation, in the form of computed procedures.

The model of the knowledge base can be represented as follows:

$$\mathrm{M}_b = f(\mathrm{B}_{IM}; B_{em}; B_{pr}; B_{CC}; B_I; B_{GK}), \tag{5}$$

where B_{IM}—is the SM base of non-typical elements; B_{em}—empirical database obtained upon analysis of the work of field facilities; B_{pr}—production rules database, formed for non-typical elements; B_{CC}—climatic conditions database; B_I—geoinformation system database of users location, sources of water and power supply, linearly extended facilities; B_{GK}—general knowledge and reference database on the typical elements of the MMHEP.

To develop the knowledge database it is necessary to provide "open systems" of knowledge database with three access levels: expert—database editing and updating of production rules databases; Research Engineer—update of the SM and empirical database of general knowledge and background data on the typical elements of MMHEP; the designer—possibility to edit and refine the databases of climatic conditions and geoinfo systems.

As the intelligent ES knowledge database forms, it allows the DM to formulate a preliminary requirement to justify the weight and importance of the criteria of choice and range parameters of the MMHEP according to production rules that would reduce the likelihood of errors upon design.

It is also important to develop an explanatory component of the intellectual ES, which will increase the efficiency of the DM, and reduce the time of decision-making when designing MMHEP.

3 Conclusion

1. The necessity to apply multi-criteria optimization of the MMHEP in the decision making process and to create a generalized optimization MM structure has been justified.
2. On the basis of a critical analysis of multicriteria methods, priori scanning methods with a variable step and random directions were chosen.
3. The application of intelligent expert systems, allowing to reduce the risk of making poor decisions when projecting a MMHEP has been justified.

The work was performed by the authors in accordance with the agreement № 14.579.21.0029 about the granting of subsidiaries dating from 05.06.2014 г. On the topic: "The development of technical solutions and technological construction of mobile micro hydroelectric power plants of derivative type for seasonal water and power supply" as a task from the Ministry of education and science of the Russian

Federation FCP “Research and development in priority fields of the scientific-technological complex of Russia for the years 2014–2020”. Unique identifier for the applied scientific research (Project) RFMEFI57914X0029.

References

1. Kasharin, D.V.: Protective Engineering Structures Made of Composite Materials in the Water Construction. South-Russian State Technical University (NPI), pp. 51–119 (2012)
2. Kasharin, D.V., Kasharina, T.P., Godin, P.A., Godin, M.A.: Use of pipelines fabricated from composite materials for mobile diversion hydroelectric power plants. Power Technol. Eng. T **48**(6), 448–452 (2015)
3. Kasharin, D.V., Kasharina, T.P., Godin, M.A.: Mobile Derivational Micro-HPP for Reserve Water Supply and Standby Power Service of Recreation Facilities and Harbour Installations of Russky Island, Nase More, vol. 62, № 4, pp. 272–277. Springer (2015)
4. Kasharin, D.V., Kasharina, T.P., Godin, M.A.: Numerical simulation of the structural elements of a mobile micro-hydroelectric power plant of derivative type. In: Kasharin, D.V. (ed.) Advances in Intelligent Systems and Computing. Proceedings of the 1st European-Middle Asian conference on Computer Modelling 2015, EMACOM 2015, pp. 51–63. Springer (2015)
5. Semenov, S.S., Petrovsky, A.V., Maklakov, V.V.: Analysis methods of decision-making in the development of complex technical systems All-Russian conference on management?, pp. 8101–8123. VSPU, Moscow (2014)
6. Antonova, G.M.: Grid Methods of Sensing Even for Research and Optimization of Dynamic Stochastic Systems, p. 240. Fizmatlit, Moscow (2007)
7. Schwartz, D.T.: Interactive methods of solving the problem of multi-criteria optimization. In: Overview, Science and Education, vol. 4, pp. 245–264. MGTU, Moscow. http://technomag.edu.ru/doc/547747.html (2013)
8. Kasharin, D.V.: Optimization of parameters selection and substantiation of mobile structures for engineering protection, made of composite materials, at water basin. In: Izvestiya B.E. Vedeneev VNIIG. 2010, vol. 260, pp. 50–60. Saint-Petersburg (2010)
9. Kasharin, D.V.: Multicriteria Optimization of Transportable Retaining Water Structure of Engineering Protection of Composite. Series: Construction and Architecture, vol. 23, pp. 145–155. Bulletin of Volgograd State University of Architecture and Civil Engineering, Volgograd (2011)
10. Gelovany, V.A., Bashlykov, A.A., Britkov, V.B., Vyzilov, E.D.: Intelligent Decision Support System in Emergency Situations Using Information About the State of the Environment, p. 304. Moscow, Editorial URSS (2001)

Hybrid Bioinspired Search for Schematic Design

Vladimir Kureichik, Vladimir Kureichik Jr. and Daria Zaruba

Abstract The paper deals with the one of the most important problem for schematic design of electronic computing equipment—parametric optimization. Due to the high complexity of this problem, the authors suggest a hybrid bioinspired search based on algorithms inspired by natural systems. The paper contains description and formulation of the parametric optimization problem. The suggested architecture is based on the multi-population genetic algorithm (GA). This approach differs from other search methods because search process is divided into two levels and at each level there are used different algorithms. So, it allows the authors to parallelize search process and obtain optimal and quasi-optimal solutions in polynomial time. Computational experiments were carried out on the basis of developed software. As a consequence of tests the authors was convinced that hybrid bioinspired search is a promising method for parametric optimization problems solution. Time complexity of developed algorithms is represented as (nlogn) in the best case and $O(n^3)$—in the worst case.

Keywords Hybrid bioinspired search · Schematic design · Parametric optimization · Bioinspired algorithm · Genetic algorithm

1 Introduction

Development of modern science and technology and Information and Communication Technologies (ICT) is closely connected with each other. Growing importance of ICT has great influence on design and manufacture fields [1]. Design of complex devices,

V. Kureichik (✉) · V. Kureichik Jr. · D. Zaruba
Southern Federal University, Rostov-on-Don, Russia
e-mail: vkur@sfedu.ru; vkur@tgn.sfedu.ru

V. Kureichik Jr.
e-mail: kureichik@yandex.ru

D. Zaruba
e-mail: daria.zaruba@gmail.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_25

such as electronic computing equipment (ECE), requires an integrated computer-aided maintenance.

Element base of electronic devices contains system-on-a-chips (SoC), very large scale integration circuits and very high speed integrated circuit (VLSI and VHSIC) [2].

One of the most important stage in the design cycle is schematic design in which there are solved several problems—structural synthesis, parametric synthesis and parametric optimization. Among common tasks at the schematic design stage the parametric optimization problem is the most difficult.

Rapid progress in terms of VLSI design causes development of new tools for automated schematic design [2, 3]. Under conditions of modern development of informational technologies existing automated design algorithms can not find optimal solutions or require too much CPU time [4]. Consequently, due to the high complexity and dimension of schematic design problems and new tendencies in the field of VLSI manufacturing there is need to develop new approaches, methods and algorithms for solving this class of problems. One of such approach is development of hybrid algorithms inspired by natural systems [5–7].

2 Description of the Problem

At the schematic design stage typical project procedures involves synthesis, optimization and analysis [2]. Synthesis is usually considered as solution of two tasks—selection of structural circuit (structural synthesis) and calculation of parameters (parametric synthesis) providing required characteristics [3]. The result of synthesis is a schematic solution which is one of the possible configuration of device.

Analysis is a determination of features of the schematic solution. This procedure allows to estimate the degree of satisfaction of this schematic solution to specified requirements and its suitability.

But, estimation of the schematic solution is quite problematic during to the structuring synthesis. This connected with the fact that obtained solutions can have non-optimal values of input parameters, which lead to inadequate estimation [3]. Hence, after the obtaining of alternative configuration it is necessary to conduct parametric optimization.

Parametric optimization of schematic solution includes finding of such input parameters that output parameters would have specified characteristics, but circuit elements and its connections would remain constant.

Parametric optimization has four stages [3]:

- generation and search of new values of input parameters of the schematic solution;
- estimation of obtained circuit configuration;
- making decision on suitability of the circuit configuration;
- making decision on continuation of search.

Let us formulate the parametric optimization problem. The quality of the schematic solution is estimated on the basis of the following objective function (OF) [4].

$$F(X) = F(x_1, x_2, \ldots, x_n). \tag{1}$$

In the general form, the considering problem with optimization of OF F(X) can be formalized as follows:

$$extrF(X), X \in D_x, \tag{2}$$

$$D_x = \{X | \varphi(X) > 0, \psi(X) = 0\}, \tag{3}$$

where F(X) is the objective function, X is a vector of control parameters, $\varphi(X)$ and $\psi(X)$ are constraint functions, D_x is an accessible area in the space of control parameters.

3 Search Architecture

To effective solution of the parametric optimization problem the authors suggest bioinspired algorithms in which the search process is a sequential transformation of one finite set of alternative solutions to another with the use of principles of genetic and evolution in nature [8–10].

The architecture of hybrid search based on the multi-population genetic algorithm and modified ant colony optimization (MACO) [11] algorithm is shown in Fig. 1.

Here GA1, GA2, GA3 and GAn are genetic algorithms, N is a number of populations in multi-population GA, a migration operator (MO) exchanges the best solutions between populations.

Hybrid search (HS) is formalized as a following tuple:

$$HS = \langle MACO, GA, MO, migration\ criterion, stop\ criterion \rangle. \tag{4}$$

Let us consider the hybrid search architecture in more detail.

MACO algorithm is based on simulation of a modified probabilistic ant colony optimization (ACO) algorithm [11].

The key idea of ACO algorithm consists in the following. During the motion ants leave marks on the ground (pheromone trail) which can be recognized by other ants. Initially an ant moves in a random way. When he meets a pheromone trail there is high chance that he will change his direction according this trail. Probability of selection this direction (follow this trail or not) is proportionally with value of pheromone in the founded trail. Consequently, the more ants pass along this trail, the more attractive this way become. So, pheromone trails, leaved by ants, send information to other ants and used as knowledge storage means in terms of solution of a routing problem. At the same time collective behavior is a form of auto-catalytic behavior [11].

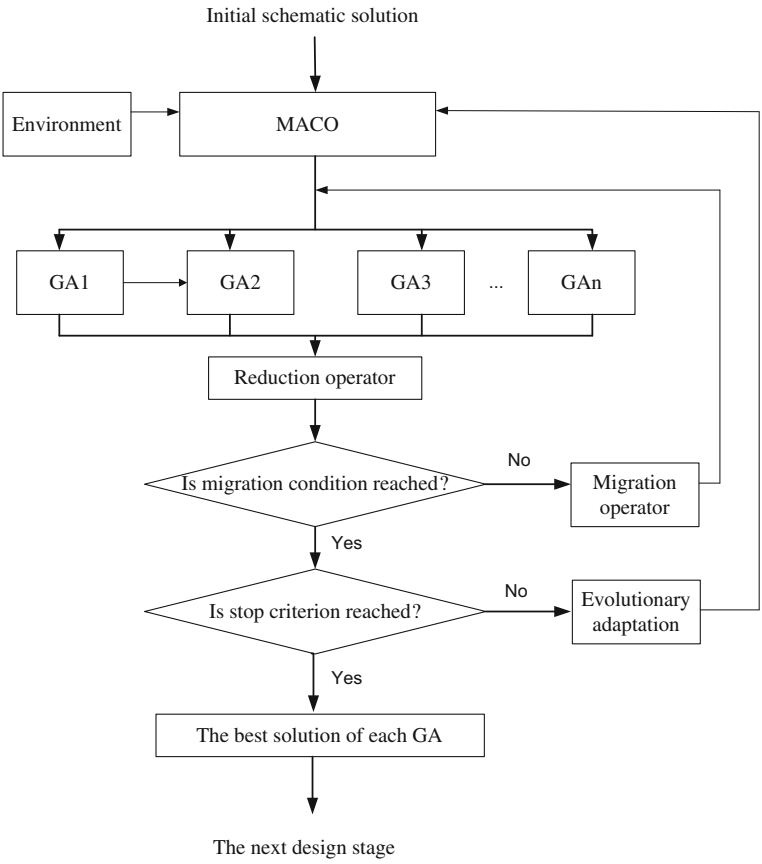


Fig. 1 The architecture of hybrid search

So, finding the shortest way to a source of food is characterized by a positive feedback, where probability of trail selection increases with growth of pheromone quantity, which was left by other ants at this trail.

The authors should be note that:

- artificial ants (agents) have memory;
- ants are in the space where time is discrete;
- from each point in the space an ant can move to any other point in the space at on transition.

The suggested algorithm has an importance feature: unpromising solutions can be considered once more. Hence, the algorithm can adapt to changing in environment, improvement or degradation of direction index. In the other words, the algorithm allow to find the best way if unpromising direction become promising, and vice versa.

The MACO algorithm is used to analyze the population development in the multi-population genetic algorithm. It reduces execution time of the hybrid algorithm. The MACO algorithm has the following assumptions:

- each vertex in a graph model correspond to a population in multi-population GA;
- quantity of agents is equal to quantity of populations;
- agents assign to each vertices sequentially;
- transition between vertices in a graph model is based on probabilistic coefficients.

The migration operator provides a interconnection between population by emigration of these solutions in the worst population. There are several migration mechanisms [9, 10]. In this case migration mechanism works as follows. Initially all populations is ranked in ascending order of OF value. After that, in each population, except the best one, the worst agents $q \times r$ are replaced by the best ones (q is percentage of exceptions chromosomes, r is a number of chromosome in population). The best chromosomes migrate from the best population.

Chromosomes are selected for migration with a following probability

$$p_i = \frac{OF(h_i)}{\sum_{j=1}^{r-q\times r} OF(h_j)}, \tag{5}$$

where $OF(h_i)$ and $OF(h_j)$ are objective functions value of alternative solutions i and j correspondingly.

Migration mechanism is implemented when a migration criterion is reached, i.e. when predetermined number of iteration is reached. In other words, after each iteration a time counter of migration t_m is incremented and if predetermined number of iteration is reached (t_m = max t_m) then migration mechanism is implemented. After that t_m is set to zero and the algorithm continues until stop criterion have been reached.

Next there is performed a genetic search which based on the modified genetic algorithm dealing with the parametric optimization problem. This algorithm can be formalized as a following tuple

$$GA = \langle P, EM, S, OF, RO, GO, h_i, N\_gen \rangle, \tag{6}$$

where P is a population; EM is a method of alternative solutions encoding; S is a selection; RO is a reduction operator; GO are genetic operators; $h_i \in P$ is a chromosome. Also genetic search can be parallelized depending on computational resources [9].

As a stop criterion we use a predetermined number of iteration.

A fundamental distinction of the suggested hybrid approach from other methods of parametric optimization of schematic solutions are dynamical adaptive distribution of population development in time. This is allow to analyze development of each population and, if unpromising solution is found, to redistribute search time of these populations. So, there is reduce search time of quazi-optimal solutions.

4 Experiments

To solve the parametric optimization problem we developed a software with the use of the Visual C++.

To determine efficiency of the suggested hybrid approach we investigated execution time and solutions quality for different sets of benchmarks. As tests examples we chose a following schematic solutions [12]:

1. audio power amplifier (C1);
2. audio power amplifier (C2);
3. audio power amplifier (C3);

Table 1 The experimental results

Test case	GA		MGA		GBA	
	Time (min)	Objective function (m)	Time (min)	Objective function (m)	Time (min)	Objective function (m)
C1	28.3	0.627	183.7	0.134	153.5	0.13
C2	28.3	0.636	175	0.144	141.8	0.218
C3	24	0.526	159.7	0.277	141.9	0.129
C4	27.4	0.628	202.6	0.109	165.4	0.123
C5	33.8	0.468	214.8	0.127	183.2	0.063

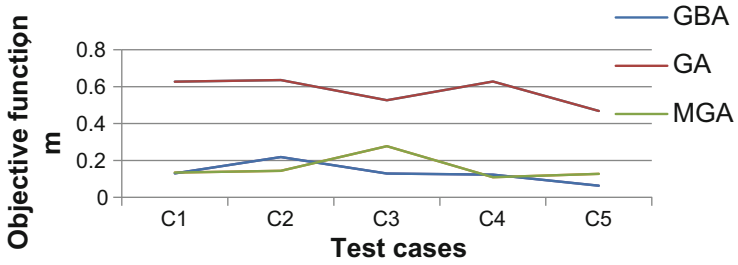


Fig. 2 The results of the algorithms for test cases

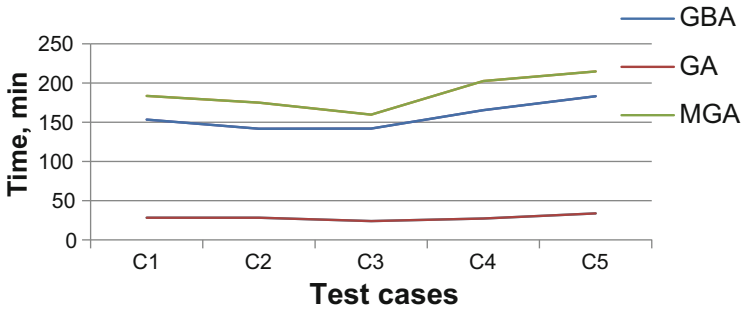


Fig. 3 The average time of the algorithms for test cases

4. bipolar circuit with active load (C4);
5. bipolar circuit with resistive load (C5).

The result of experiments is shown in Table 1 and on Figs. 2 and 3.

5 Conclusion

In the paper we considered parametric optimization of schematic solutions. The developed new architecture of hybrid search is based on a multilevel approach which enables us to generate quasi-optimal alternative solutions with probable migration between solutions. On the basis of developed software we carried out experiments. As a result, the quality of schematic solutions, obtained by the hybrid algorithm, is greater than solutions, obtained by other algorithms, by 5.64 %. This fact confirms of efficiency of suggested approach. So, time complexity of the algorithm in the best case is represented as ≈O(nlogn), in the worst case—$O(n^3)$.

Acknowledgments This research is supported by grants of the Ministry of Education and Science of the Russian Federation, the project # 8.823.2014.

References

1. Kureichik, V.V., Kureichik, V.M., Malioukov, S.P., Malioukov, A.S.: Algorithms for Applied CAD Problems, 487 pp. Springer, Berlin, Heidelberg (2009)
2. Sherwani, N.A.: Algorithms for VLSI Physical Design Automation, 3rd edn. Kluwer Academic Publisher, USA (2013)
3. Kacprzyk, J., Kureichik, V.M., Malioukov, S.P., Kureichik, V.V., Malioukov, A.S.: General questions of automated design and engineering. Stud. Comput. Intell. **212**, 1–22 (2009)
4. Alpert, C.J., Dinesh, P.M., Sachin, S.S.: Handbook of Algorithms for Physical Design Automation. Auerbach Publications Taylor & Francis Group, USA (2009)
5. Zaporozhets, D.U., Zaruba, D.V., Kureichik, V.V.: Hybrid bionic algorithms for solving problems of parametric optimization. World Appl. Sci. J. **23**(8), 1032–1036 (2013)
6. Zaporozhets, D.Y., Zaruba, D.V., Kureichik, V.V.: Hierarchical approach for VLSI components placement. Adv. Intell. Syst. Comput. **347**, 79–87 (2015)
7. Gladkov, L.A., Gladkova, N.V., Leiba, S.N.: Electronic computing equipment schemes elements placement based on hybrid intelligence approach. Adv. Intell. Syst. Comput. **348**, 35–44 (2015)
8. Gladkov, L.A., Kureichik, V.V., Kravchenko, Y.A.: Evolutionary algorithm for extremal subsets comprehension in graphs. World Appl. Sci. J. **27**(9), 1212–1217 (2013)
9. Kureichik, V.V., Kureichik, V.M.: Genetic search-based control. Autom. Remote Control **62** (10), 1698–1710 (2001)
10. Bova, V.V., Lezhebokov, A.A., Gladkov, L.A.: Problem-oriented algorithms of solutions search based on the methods of swarm intelligence. World Appl. Sci. J. **27**, 1201–1205 (2013)
11. Markov, I.L., Hu, J., Kim, M.-C.: Progress and challenges in VLSI placement research. Proc. IEEE **103**, 1985–2003 (2015)
12. Swetha, R.R., Devi, S.K.A., Yousef, S.: Hybrid partitioning algorithm for area minimization in circuits. In: International Conference on Computer, Communication and Convergence (ICCC 2015), vol. 48, pp. 692–698 (2015)

Figuratively Semantic Support in Interactions of a Designer with a Statement of a Project Task

P. Sosnin, M. Galochkin and A. Luneckas

Abstract The paper presents a system of means that allows improving the interaction of a designer with generated and used textual descriptions by automated constructing the semantic images extracted from these texts. Such constructing involves both hemispheres of the designer's brain where the logical and figurative processes are combined in the most efficient wholeness that facilitates to understanding and positively influences onto creative actions of the designer. To translate the definite part of the investigated text (for example, the statement of the project task) in a corresponding semantic image, the designer can use a specialized linguistic processor, converter and graphical editor that supports creating the schemes with programmable interpretations concerned their semantic content. The process of such investigation is implemented in the toolkit OwnWIQA.

Keywords Conceptual experimenting · Human-computer interruption · Multitasking · A queue of project tasks · Semantic memory · Question-answering · The flow of tasks

1 Introduction

For designers, the use of semantics begins with first steps of their interactions with the initial description of the future work, and it is especially important at steps of conceptual designing when they try to find initial solutions of project tasks some of

P. Sosnin (✉) · M. Galochkin · A. Luneckas
Ulyanovsk State Technical University, Severny Venetc Str. 32, 432027
Ulyanovsk, Russia
e-mail: sosnin@ulstu.ru

M. Galochkin
e-mail: m.galochkin@ulstu.ru

A. Luneckas
e-mail: lunakorp@inbox.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_26

which unexpectedly arise. For any new task Z_j, its necessity is discovered by the designer in an own thought activity exciting left and right hemispheres of the brain when the designer is solving the concrete task Z_i. Processes in hemispheres are intertwined in the real time. For coordinating the intertwining, they should reflect in textual and graphical forms for creating them, first of all, in an initial statement of the corresponding task. All of these actions are fulfilled by the designer who interacts with owned naturally professional experience and models of an accessible experience.

The described situation requires interactions with the experience and its models from following subject areas:

- The human-computer interruption that helps to manage relations between tasks Z_i and Z_j in the fulfilled work;
- A subject area of a system or object being created;
- An instrumentally technological environment that supports a collaborative activity of designers;
- Other subject areas (for example, Logics and Linguistics) that can be useful or serve as sources of borrowing.

Usually, designers have problems with the experience in Logics and Linguistics. The lack of the experience in these subject areas should be compensated by its inclusion in appropriate toolkits. In this paper, we offer an approach and toolkit that support to create the statement for newly appeared tasks in the context of personal activity of the designer. The offered version of interactions is based on the use of reflections of the operational space of designing on the semantic memory of the toolkit WIQA (Working In Questions and Answers) that has been created for conceptual designing of Software Intensive Systems (SISs) [1].

We start the rest of this paper with highlighting the features of the approach in conditions of multitasking. The third section introduces related works used as a source of comparing and inheritances for the offered approach. The fourth section opens features of figuratively semantic support. The fifth section presents some rules of managing the work of the designer with statements of new tasks.

2 Preliminary Bases

Interactions of any person with the experience inevitably accompany any human activity. In designing, interactions can be caused by different reasons some of which are stimulated by interfaces of an instrumentally technological environment that support designer's activity. Unexpected interruptions of the current work of the designer are another source of reasons among which one can mark discovering by the designer a necessity to define and solve a new task, for example, the task Z_i. In

our paper, situations of such type occupy the central place. Furthermore, interests of our paper are restricted by interactions with the new task when the designer creates its statement.

Often enough, the new task (for example, the task Z_j) arises when the designer solves another task Z_i, and, it should be noted, that tasks Z_j and Z_j can be substantively related or not. If they are not related in content, then the designer can be forced to switch to another subject area for interacting with the corresponding experience and its models. Therefore, in the appearance of the new task Z_j in described conditions, the designer should define and solve the task Z_j in the context of managing the relations between tasks Z_i and Z_j. To take into consideration of these relations, the designer should use the experience of the (human-computer) interruption management, especially the experience of self-interruptions [2].

Indicated kinds of human-computer interruptions occur in behavioral actions of designers. At the life cycle of the task, these actions dominate at early stages when the designer creates the statement of the task and builds its conceptually algorithmic solution. Both of these activities require the work with semantic representations of textual units and conceptually-algorithmic constructions. Furthermore, interactions of the designer with models of these components are intertwined in the real time. It is caused by the necessity of achieving the sufficient understanding the task. It should be noted that an inclusion of the graphical support in indicated interactions facilitates achieving of understanding.

In the offered approach, these processes and results are implemented and attained in a semantic memory of the toolkit WIQA that supports following basic functionalities [1]:

- Reflecting the operational space of the work with task and groups of tasks onto the semantic memory cells of which are oriented on question-answer specifying the mapped objects;
- Pseudocode programming the behavioral actions with such objects and their compositions in the context of interactions of the designer with the natural experience and its models;
- Conceptual experimenting with declarative and algorithmic constructions reflected onto the semantic memory;
- Agile managing of implementing the flows of tasks and sets of tasks in multitasking mode, based on human-computer interruptions.

It should be noted that indicated functionalities are accessible in conditions that are shown in the scheme (Fig. 1) where one can see the Experience Base with embedded Ontology also mapped onto the semantic memory of the question-answer type (or, shortly, QA-memory).

The scheme indicates that the operational space S includes a conceptual space (C-space) as an activity area where designers fulfill the automated part of own conceptual thinking when they work with conceptual artifacts. It should be noted,

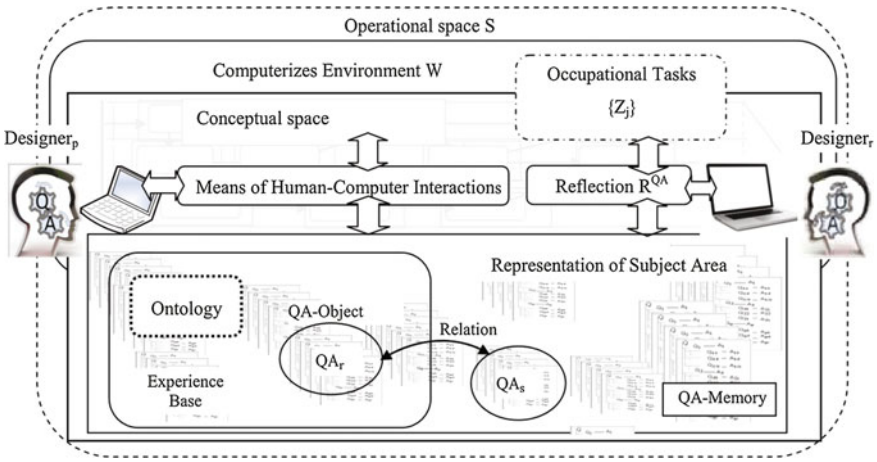


Fig. 1 Operational space

the C-space is a kind of the actuality, that inherits definite regularity from the operational space, and it has owned regularities expressed the behavioral nature of the human activity.

In solutions of project tasks $\{Z_j\}$, any result of the reflection (or shortly R^{QA}) is accessible for any designer as an interactive object (QA-object). Conceptually thinking, designers work in the C-space, for example, conducting the necessary experiments. In such actions, they interact with necessary concepts, units of the Experience Base and other components of the C-space. Thus, the C-space is the system of conceptual artifacts that are created and used by designers in processes of conceptual thinking when they conceptually solve the project tasks. In this kind of the activity, the C-space models entities and processes of designing, and such modeling is oriented on their reflections in the consciousness of designers. For example, reasoning, the designer combines corresponding dialog processes in consciousness with their models registered in cells of the QA-memory.

3 Related Works

The nearest group of related works concerns the subject area "Conceptual Development and Experimentation (CD&E)." The name CD&E often use for the corresponding technology and method. For example in [3] CD&E is defined as "A method which allows us to explore and predict, by way of experimentation, whether new concepts that may impact people, organization, process and systems will contribute to transformation objectives and will fit in a larger context."

The specificity of processes in this subject area is presented in the report [4] where the role and place of conceptual experimentation in military applications are

indicated in detail. The following publication [5] defines the version of the occupational maturity of the CD&E-process. The publication [6] demonstrates some specialized solutions that are focused on a behavioral side of experimenting. It should be noted that, in published works, analogies with thought experiments and reflections on the semantic memory are not used.

One more group of related tasks deals with the explicit and implicit programming of a human behavior caused by an interaction potential embedded in screen interfaces. Registering the human activity in program forms has been offered and specified constructively for Human Model Processor (MH-processor) in the paper [7]. The good example of implicit programming of the human behavior is described in [8] where multitasking and interruptions are estimated for a web-oriented activity. Both these examples indicate the usefulness of implementing the human behavior in program forms.

The next group combines the papers presents self-interruptions in the human-computer activity. In this group, we mark the paper [9] that develops a typology of self-interruptions based on the Multitasking integration of Flow Theory and Self-regulation Theory. The typology is based on negative (Frustration, exhaustion, and obstruction) and positive (stimulation, reorganization, and exploration) triggers. The very useful paper [2] also suggest a typology of self-interruptions including seven basic types ("adjustment," "break," "inquiry," "recollection," "trigger" and "wait") each of which generates for increasing the effectiveness of multitasking.

It should be noted, all papers indicated in this section were used as sources of requirements for developing the set of instrumental means provided the offered version of the interruption management. Any of these papers concerns only a part of the offered approach to interrupting in conditions of multitasking.

4 Features of Figuratively Semantic Support

Means of pseudocode programming can be applied as for the use of existed functionalities of the toolkit WIQA but for the development of its new components, and this possibility has been used for the implementation of the offered version of figuratively semantic support as an additional functionality. Such extension has required the development of a graphical editor that admits its programmable inclusion in actions aimed at achieving of understanding in the work with new tasks.

More details, the essence of the authors' suggestion consists in the following:

1. After discovering the necessity to include the new task Z_j in the executed work, the designer activates own thinking aimed at creating the initial model of the task in a symbolic form. This process includes two parts the first of which happens in brains of the designer while the second part provides transferring the results of thinking in their expression using the naturally occupational language.

In this form, initial results are mapped on the QA-memory. So, the second part of the process is doing that intertwined with indicated thinking.

2. In the offered approach, the graphical support adds thinking and doing of the designer in interactions with the creating task. Switching on a graphics explicitly initiates the process in the right hemisphere of the brain. Drawing helps to extract and visualize only essential semantic components (of the generating symbolic specification of the task) and the scheme of their relations (for example, in a form of a block-and-line scheme). Nonessential semantic components bring difficulties for understanding. The use of significant components allows reducing, on the one hand, the text description, and with another side, opens an opportunity for a deeper analysis of the text from the viewpoint of contradictions.
3. Reflecting the investigated text and built block-and-line scheme on the question-answer memory of the toolkit WIQA [1]. It helps to create the question-answer model of the text that is not only coordinated with the built block-and-line scheme, but it can be transformed into a pseudo code program simulating the process of drawing this scheme. Results of reflections open additional opportunities and in checks of the text on "integrity of the informational content". Besides it, as a technical figure with specifications prepares formation normative design schedules.
4. Coordinating the block-and-line scheme with the project ontology [1] that is also stored and evolved in the semantic memory of WIQA. Coordinating is a process of enriching the scheme by semantics units inherited by scheme components from corresponding ontology components.

It is necessary to note that called transformations are better to understand and implement as translations from one form of a description (symbolic or graphical) to another form, for example, from the symbolic form to another symbolic form or from the symbolic form of the corresponding figuratively semantic scheme (FS-scheme). An instrumental complex that provides such translation and other transformations of the FS-scheme is presented in Fig. 2.

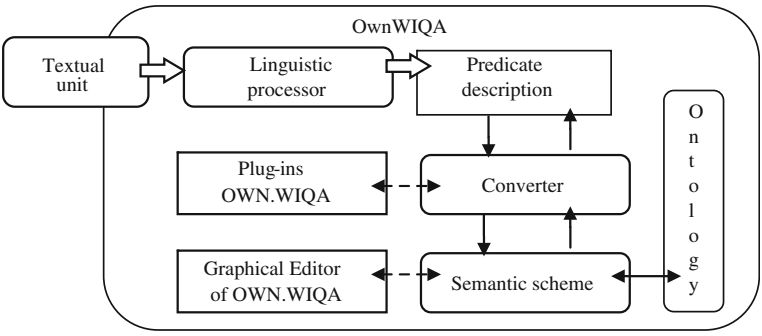


Fig. 2 The complex of means for transformations

The complex is implemented as an application in the environment of the toolkit OwnWIQA. This complex uses the OwnWIQA as the kernel that evolves by a number of means oriented on the figuratively semantic support. Only these means are obviously indicated in Fig. 1, and they include:

- Linguistic processor that provides automated translation of the investigated textual unit in its prolog-like description;
- Converter transforming such description in versions of the FS-scheme in iterative process of its development;
- The graphical editor that supports some transformations of the created FS-scheme from its initial state till the understandable version that is needed for the designer.

The first transformation (by linguistic processor) helps to translate the textual units T_i in the list of simple clause $T^*_i(\{C_j\})$ any of which has one of following structures:

$$\begin{aligned} C_j &= P_j(Ob_q),\, AC_j \\ C_k &= P_j(Ob_r,\, Ob_s),\, AC_k \end{aligned} \tag{1}$$

where P—predicate of the clause, Ob—name of the definite object, AC—associative component that is located out of the predicate form. This component is included in the sentence of the text for the necessary utilization of the described meaning.

The second transformation fulfilled by the designer with the help of the converter allows creating the initial state of the corresponding block-and-line FS-scheme. In this work, the description $T^*_i(\{C_j\})$ is processed as a program of the declarative type. Such understanding is inherited from the logical programming that uses prolog-like operators.

The third transformation has some varieties:

- Translation in the pseudo code program that helps to draw the FS-scheme;
- Corrections of the FS-scheme on the base of information from associative components;
- Enriching the components of the FS-scheme and relations among these components by information that is registered the current state of the project ontology.

This description will be stored in the semantic memory of the toolkit WIQA, and it will be visually accessible to the designer in interface environment where the designer has the possibility of correcting the description. After that, the Prolog-like description will be interpreted as a declarative program that is written in an extension of the pseudocode language L^{WIQA}. Operators of this extension have the syntax that corresponds to the expressions (1). Such transformations are iteratively implemented in an operational surrounding that is presented in Fig. 3.

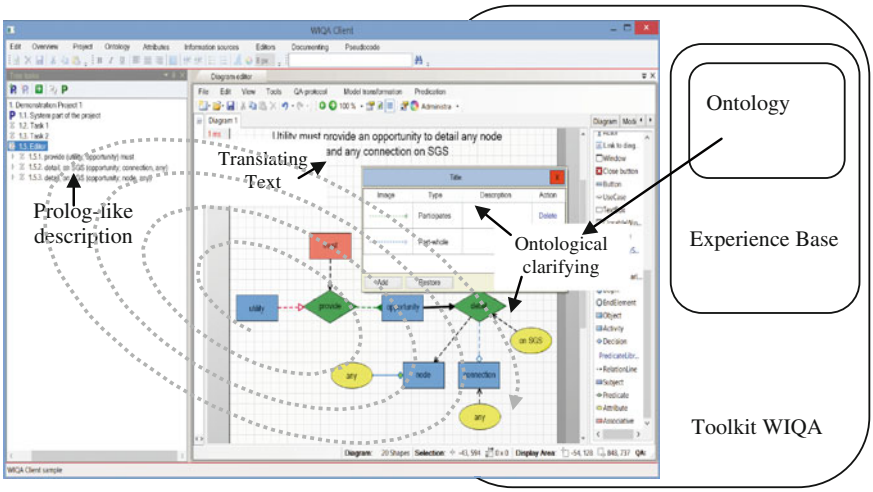


Fig. 3 Iterative coordinating the text and SG-scheme

This figure includes clarifying labels. The FS-scheme can be plugged to the ontology for the scheme specification and its informational enriching. Interactions with the ontology help to open the actual variants of the systematization registered in the ontology for nodes of the built scheme.

The scheme also reflects (by the spiral) the dynamics of iterative coordinating the text unit with its FS-scheme. The basis of such coordination is the feedback of FS-scheme with corresponding prolog-like description (based on forms presented in Fig. 2). Any correction at the scheme leads to the correction of the text equivalent. This mechanism is useful when translating complex sentences into simple clauses described at their prolog-like versions. Note, that the aggregation of the text T, its prolog-like version, and the coordinated graphical model express the understanding that the creator of these constructs embeds in the result of transformations.

5 Features of Management

The final aim of the offered support is to create the understandable version of the task statement for the task that unpredictable arises in the multitasking mode of the designer's activity. In the general case, any such task Z_j' begins its life-cycle when the designer self-interrupts the work with the task previously chosen from the definite queue of tasks being solved in pseudo-parallel mode. For the general case, such situation is schematically shown in Fig. 4.

The scheme demonstrates that at the definite interval of time, the designer (as a member of the team developing the definite project) usually works with a set of tasks. Moreover, in the current moment of time, the designer works only with the

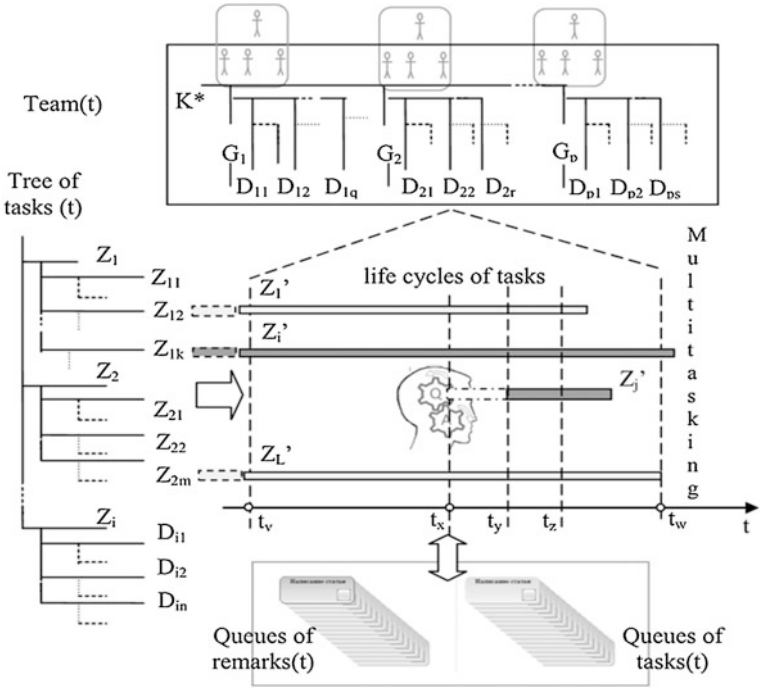


Fig. 4 Pseudo-parallel activity of the designer in a multitasking mode

one of these tasks and this active task Z_i' is chosen by the designer from the definite queue of tasks extracted from the tree of tasks in accordance with agile management rules.

The scheme also indicates that, at the moment of time t_x, the designer discovers the appearance of the new task Z_j' that, let us suppose, should be included in the process of designing. As told above, the new task starts an own life cycle in the mind of the designer or, more definitely, it explicitly starts with a composition of signs, for example, as a set of words (w_1, w_2, …, w_k) that appeared in the left hemisphere of designer's brains.

Beginning with the moment of time t_x till a point-in-time t_y, the designer must decide to continue the work with the task Z_i' or switch own attention to the new task Z_j'. At the interval (t_x, t_y), thinking of the designer estimates signals w_1, w_2, …, w_k in the context of the tasks Z_i' for resolving the following consequences:

1. If the work with task Z_i' will be continued without transferring and registering the signals (keys) w_1, w_2, …, w_k out of brains then these keys to the task Z_j' may be lost, probably without re-return.
2. If the attention of the designer will be fully paid to the task Z_j' then a temporary break from the work with the task Z_i' can have a negative impact on the solution

of this task. Moreover, the degree of negatives is usually growing when the interval of the break is increased [9].

The offered approach is based on a position that the statement $St(Z_j, t)$ of the task Z_j is a dynamic construction any state of which includes the textual part $T(Z_j, t)$ and the corresponding FS-scheme if it has already been built. Dynamics of the statement has following features:

1. After a point of self-interruption, when the first step of life-cycle, the state $St(Z_j, t_0)$ includes only a list of essential keywords that implicitly indicates on the initial state of the text $T(Z_j, t_1)$. The list of keywords is saved in a specialized queue of remarks.
2. At the second step, the designer creates the initial statement of the task $T(Z_j, t_1)$ iteratively using the figuratively semantic support that has led to the corresponding state of FS-scheme.
3. At the third step, the task Z_j is included in the multitasking set for processing in accordance with normative rules [4].
4. Any next step evolves the state $T(Z_j, t_i)$ by the stepwise refinement method with using the question-answer reasoning. Any result of evolving is qualified as the next increment $\Delta T(Z_j, t_{i+1})$ that is processed the initial version of the statement similarly. Such work is repeated for the state $St(Z_j, t_k)$ the informational content of which is sufficient for the work with the task Z_j at the next stages of designing.

Enumerated steps are implemented under control of predefined managerial actions with using means of agile and interruption management.

6 Conclusion

The following conclusions can be drawn from the present study of figuratively semantic support in creating the statements of new tasks that are discovered by designers in the process of designing in the real-time. For any of such tasks, its life cycle should begin with registering the list of keywords presented the task as the reason of the next self-interruption. It allows including the newly appeared task in the multitasking set. After which this task will be processed and managed by common normative rules.

For next steps of the life cycle, the designer reflects the task on textual and graphical models that are intertwined in the interactive process of their collaborative developing. The aim of this process is to attain the necessary level of understanding that finds its expression in the created statement of the task. In its turn, the achieved understanding fixates that the developed statement reflects a concrete wholeness of requirements and restrictions in conditions corresponding the described task. It should be noted that interactions with textual and graphical parts of the statement activate the processes in the left and right hemispheres of any member of the

designer team. The reuse of these processes also leads to the version of understanding which can be compared with the first version. So, understanding fulfills the controlled function.

References

1. Sosnin, P., Maklaev, V.: Question-answer reflections in a creation of an experience base for conceptual designing the family of software intensive systems, In: A. Kravetc et al. (eds.) Knowledge-Based Software Engineering. CCIS, vol. 466, pp. 658–672. Springer International Publishing (2014)
2. Darmoul, S., Ahmad, A., Ghaleb, M., Alkahtani, M.: Interruption management in human multitasking environment. In: Preprints of the 15th IFAC Symposium on Information Control Problems in Manufacturing, 11–13 May 2015. Ottawa, Canada. IFAC-PapersOnLine, vol. 48 (3), pp. 1179–1185 (2015)
3. MCM-0056-2010: NATO Concept Development and Experimentation (CD&E) Process. NATO HQ, Brussels (2010)
4. Sosnin, P.: Combining of kanban and scrum means with programmable queues in designing of software intensive systems. In: 14th International Conference on Intelligent Software Methodologies, Tools and Techniques (SoMeT-2015). Communication in Computer and Information Science (CCIS), vol. 532, pp. 367–377 (2015)
5. Wiel, W.M., Hasberg, M.P., Weima, I., Huiskamp, W.: Concept maturity levels bringing structure to the CD&E process. In: Proceedings I/ITSEC 2010, Interservice Industry Training, Simulation and Education Conference Orlando, pp. 2547–2555 (2010)
6. Jaros, V.: Concept development and experimentation in project structure. In: Cybernetic Letters. Expertia, o.p.s., Brnovol, vol. 2, pp. 36–41(2011)
7. Karray, F., Alemzadeh, M., Saleh, J.A., Arab, M.N.: Human-computer interaction: overview on state of the art. Smart Sens. Intell. Syst. **1**, 138–159 (2008)
8. Leiva, L.A.: MouseHints: easing task switching in parallel browsing. In: CHI ‘11 Extended Abstracts on Human Factors in Computing Systems, pp. 1957–1962. ACM, New York, NY, USA (2011)
9. Adler, R.F., Benbunan-Fich, R.: Self-interruptions in discretionary multitasking. Comput. Hum. Behav. **29**(4), 1441–1449 (2013)

Part V
Intelligent and Fuzzy Railway Systems

Cloud-Assisted Middleware for Intelligent Distributed and Mobile Objects

Andrey V. Chernov, Alexander N. Guda and Oleg O. Kartashov

Abstract This paper proposes a novel concept and technical architecture of supporting computing environment for groups of intelligent objects and devices. The main aim of proposed system is a providing an efficient reliable and scalable remote access to sensors and actuators of moving objects with using a local cloud-assisted architecture that performs intensive intelligent and distributed data processing. A key feature of our system is the presence of a communication middleware that enables a low-cost wireless network deployment around a group of mobile moving objects and connects it to the stationary control node. In the paper, we describe the low power, low range and low-cost design and implementation of proposed intelligent community middleware.

Keywords Intelligent system · Cognitive nodes · Mobile nodes · Cloud-assisted middleware

1 Introduction

Currently, there is significant progress in improving the performance of compact robots that move embedded computing facilities, the creation of new types of sensor systems and measurement systems with reducing their sizes and increasing their accuracy. These circumstances create the preconditions for a new generation of sophisticated technical facilities, capable of autonomous operation in a dynamic environment for the solution of their engineering problems. These objects include unmanned vehicles of various types (ground, air, underwater) and destination

A.V. Chernov (✉) · A.N. Guda · O.O. Kartashov
Rostov State Transport University, Rostov-on-Don, Russia
e-mail: a.v.chernov@hotmail.com; avcher@rgups.ru

A.N. Guda
e-mail: guda@rgups.ru

O.O. Kartashov
e-mail: okrstu@yandex.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_27

(military, civilian, commercial), a variety of robotic systems (controllers, control systems) and others.

Now intelligent distributed and mobile objects (*IDMO*) are widely used for the following tasks:

- Aerial photography;
- Exploration of the territory;
- Patrol (guarding) the territory and infrastructure;
- Search for objects of interest in the territory;
- moreover, etc.

Government organizations and agencies, private commercial companies, use the *IDMO* in their activities. However, despite the widespread *IDMO* to control it now, in most cases it requires human intervention—an operator. The operator in the remote control mode using a remote control, a workstation, or directly carries the control of an *IDMO* in real time. It substantially limits the scope of application of *IDMO* and reducing the economic impact of it. In other words, now it is a very urgent task of developing such complex *IDMO* management systems that would allow management entities performs their activities within the context of high-level human tasks in an entirely automatic mode.

Our paper proposes a cloud-assisted architecture to enable the communication among *IDMO* subsystems, organizing a local area low-cost, low-power and low-range wireless network. In Sect. 2 some related work is carried out. In Sect. 3 we propose the hardware architecture to our system. The system evaluation is described in Sect. 4.

2 Related Work

To date, we know some approaches to the construction of complex mechanical objects control systems, unmanned vehicles, and mobile robots. One of the most common is an approach based on functional decomposition when the control system is a set of modules, each designed to address a particular type of task. The number of such modules and principles binding them into a single structure, in general, are not regulated. However, in the late 90 s of the twentieth century, attempts were made certain standardization in this field. It may be noted *JAUS* standard [1] defines the protocols and formats for inter-module interactions. It was assumed that following this standard make it possible to create control systems by integrating modules from different vendors, as well as “quick-change” modules where appropriate. An example of a single-level, modular control system, is described in [2]. In the early 2000s, a the National Institute of Standards (*NIST*) model of the control system architecture was designed for unmanned vehicles, which are not based on functional

and modular decomposition. It is named *4D/RCS* [3]. It is important to note that the *4D/RCS* involves the ability to create unlimited nesting hierarchies of control nodes, i.e., in fact we are talking about a multi-level governance with a potentially unlimited number of levels. Note that the idea of layered architectures is not new in robotics, artificial intelligence, control theory, and related disciplines. However, usually, when it comes to layering, there are two of them, which we assume to be called deliberative and reactive levels. An important issue in this approach is the principle of organization of interaction between the two levels. The study of this issue is devoted to some works, for example [4].

It can be argued that nowadays the use of one- and two-tier architecture for building control systems of complex mechanical objects *IDMO* is impractical for the following two interrelated factors. First, the increase in the volume of data coming from sensors of the control object. Secondly, by analyzing and processing a particular input image, it becomes possible to solve a broad range of applications, combine in one deliberative level leads to excessive complexity modular architecture and redundancy modules, a concept designed to solve individual tasks, but using different techniques. As an example, the task of planning, at the same time, the decision to Advisory levels in different contexts and using different methods. Accordingly, it seems promising and justified further subdivision of the deliberative level to sub-levels: strategic, which is responsible for the formulation and selection of goals, forecasting, and high level of information processing and tactical, which is responsible for the recognition, terrain mapping, path planning, etc. This control system consists of three levels, is investigated in the paper. Now we will describe the current state of research in the field of problems arising at each level of control.

The architecture of control systems, including strategic management level, is called cognitive architectures. One of the most advanced used in the industry of cognitive architectures is *SOAR* architecture [5], which is used to develop agents that possess actual operation complex behavior in a dynamic environment. Knowledge in *SOAR* is stored in memory, consists of three main blocks: the block of long-term memory, short-term (working) memory and estimated memory. The description of the objects like frame structure in which is not set by default, no attached procedures, and in which there are no inheritance properties of the hierarchy tree. The transitions between states are stored in long-time memory, which is a standard rule base. A significant disadvantage of *SOAR* is that rules, once be set, do not change over time. There are no procedures for learning new rules. The *SOAR* architecture, despite its popularity, has no mechanisms of formation of new rules and coalitions of agents that sufficiently strongly limits its application outside the narrow class of problems.

In the Project *Icarus* [6] introduced the division of long-term memory to conceptual memory and the memory of action. Both the first and second class have the classical hierarchical organization of the primitives on the lower level. The strategic level in the *Icarus* architecture are well demonstrated in the problem of vehicle control in a transport network of the city, but mechanisms of learning in her poorly developed and do not include the mechanism for learning new concepts, what makes to form the conceptual long-term memory again on hand for a new task.

Much more complex mechanisms of learning, claimed in one of the last cognitive architectures, are *Clarion* [7]. This project mainly aims at developing the architecture of agents that can work together, and not used to control any object, but it serves as a good example of the implementation of the strategic level of management. It is assumed that the control system can act by trial and error, getting feedback on actions. In the case of an unwanted result, the network learns is by the method of backpropagation of errors. One of the differences *Clarion* from other cognitive architectures is the presence of a subsystem of motives and goals, which should be the internal stimulus of activity system governance. Actually, in this subsystem, new goals are formed, and the motifs are descriptions of desired situations, i.e. the same objectives, but which are updated periodically. *Clarion* were first proposed internal mechanisms of the formation of new concepts without an external teacher. However, the method of trial and error requires time and resources for a large number of unproductive actions that leads to poor management of real systems. It was the main reason that this cognitive architecture has not been successfully applied to real problems.

We reviewed cognitive architectures have various disadvantages that are in some way affected the quality of the high-level planning, which is responsible for considering the level of strategic control. In any system are not implemented effective mechanisms for the formation of new ideas, new concepts, and new actions. It entails the most important the lack of them—for the control system in advance and sets goals that will reach this system. However, in most cases, the formal description of the purpose language of knowledge representation that is used in the system, make up a large part of the problem.

One way of overcoming these difficulties is to develop such a knowledge representation in which the description of the concept. Its relationship with the basic levels of management and a description of possible actions with this concept, will be divided into various blocks and subsystems, and will represent a single unit of information that is produced and used during planning as a whole.

3 A Cloud-Assisted Architecture for *IDMO*

We propose a complete architecture for the multilevel intelligent control system (*MICS*) *JSC* "*Russian Railways*" in [8]. The general system concept of *MICS* is presented in Fig. 1.

The component multilevel architecture of *MICS* is based on the domain-oriented software platform. This platform provides *Manufacturing Execution System*-level components, domain-oriented platform, multi-agent software for sensor and actuators, domain ontology, domain specific language and other high-level software and user interfaces. Lets' consider the "*Multi-agent sensor components*" subsystem more detailed. Now, at the railway stations and railway sections, there are many different low-voltage sensors and high-powered actuators. They control of huge complex infrastructure of railways, ensure rail traffic and transportation safety. The sensors

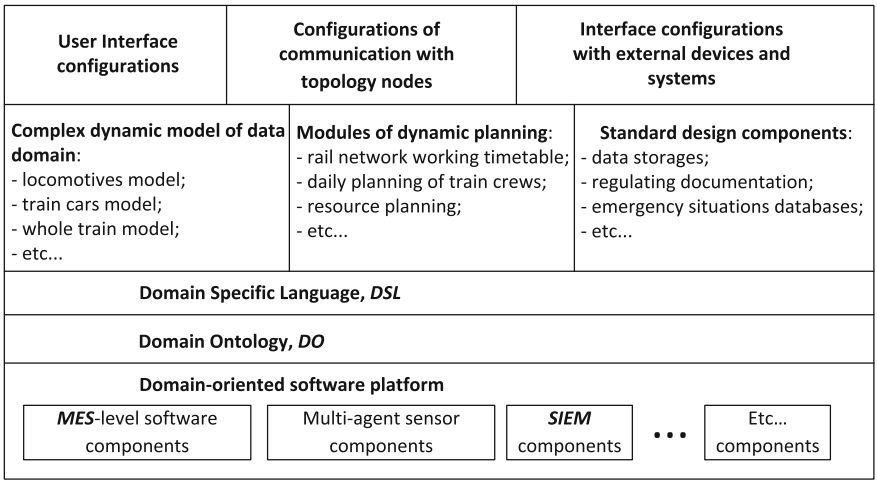


Fig. 1 System concept of *MICS JSC* "*Russian Railways*"

and actuators are used by various automated control systems, such as automatic block signaling, centralized traffic control, direct traffic control automatic train control systems and many others subsystems and objects in *MICS*. However, it is lack of united middleware for such sensor systems.

Next we propose a system cloud-assisted architecture for *IDMO* as part of aforementioned *MICS* in Fig. 2.

The main objective of management of complex technical object on a reaction level is to ensure the specified characteristics of the dynamic object with the help of control action by feedback. For this purpose, the process control signal from a tactical level receives the desired trajectory and the desired motion parameters (e.g. speed). Level can operate in two modes: management of real object and numerical modeling of process control.

At the tactical level, the control tasks navigate object in space, the main of which are: building, updating, refining the model maps, localization, binding of the control object to the current map and trajectory planning. The last task is divided into 3 phases: forecasting, scheduling and monitoring. In the forecasting, the module receives information from the level strategic management of the target space region and temporal constraints on the achievement of this field, the object followed by a preliminary calculation of the required motion parameters to perform the assigned tasks. The main difference of methods for predicting the trajectory from the actual planning methods is that the former can ignore the restrictions on the dynamics of object motion control and other restrictions. In this way, the trajectory prediction is performed with a minimum expenditure of computing resources. Calculated by the forecasting module the motion parameters are transmitted to the lower, reaction level of control that is subject to the restrictions on the dynamics of the control object. To form the geometric constraints on the shape of the trajectory, the compliance with

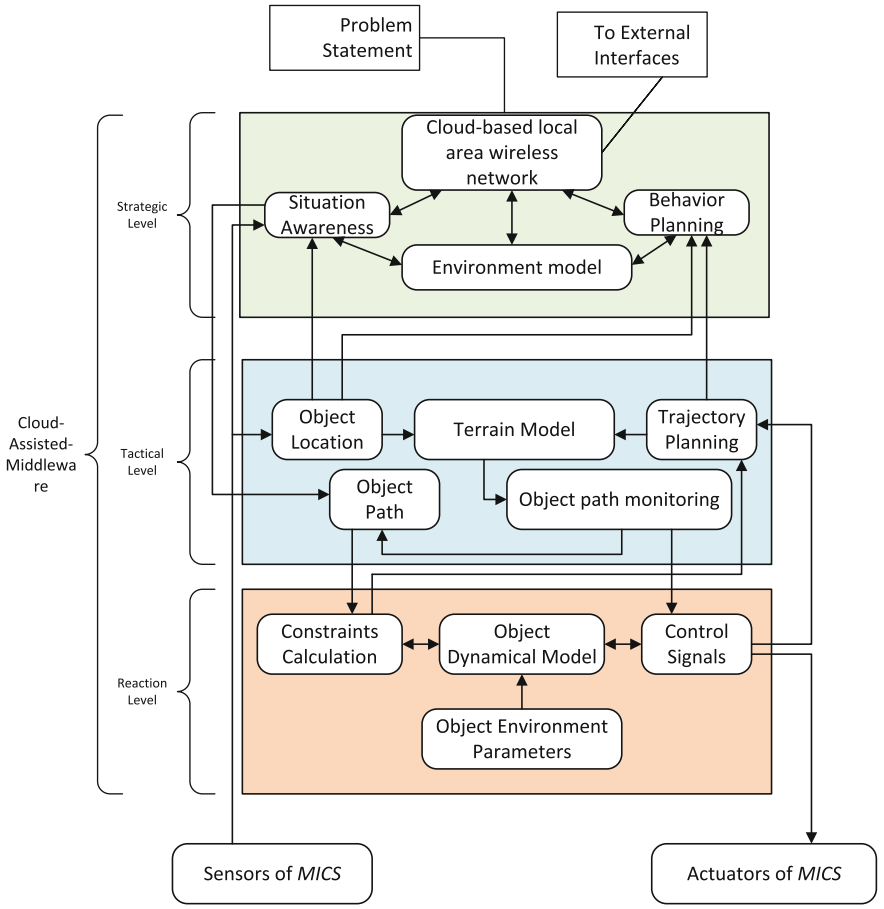


Fig. 2 Proposed cloud-assisted architecture for *IDMO* as part of *MICS*

which ensures the possibility of following it at fixed, previously calculated motion parameters. The planning module builds the trajectory taking into account these geometric constraints. The result is a built path or a signal that under the given constraints, the task of path planning impossible in the allocated time. In the latter case, at the strategic level is transmitted a request for rescheduling, i.e., a different target region in space and/or time constraints. Thus, path planning is an iterative process with feedback from both the upper and lower levels of government, which is a significant difference of the proposed architecture from modern analogs.

The main task of the management at the strategic level is to build a concerted plan between the members of the coalition and the behavior of each participant of joint activity. Each participant has its environment specific due to how the characteristic properties are restricting the set of available system control actions and components of the environment. The values of all elements of the environment are the same for

all members of the coalition by definition. Building behavior plan occurs through the exchange of messages with other members of the coalition. At each stage of the plan, the update has to be described the current situation, including involving the information coming from the sensors. From the description of the present situation is highlighted spatiotemporal information, which generates a job arriving at the tactical level and contains the spatial description of the target area and time constraints to achieve it. The rescheduling procedure takes a result on the possibility/impossibility assigned to perform the task of moving of object and if necessary, make agreed with the other members of the coalition adjustments to the overall plan.

4 Proposed Technical Concept

The goal of the proposed cloud-assisted middleware is to standardize the protocols, *APIs* and services between managed *IDMO* and *MICS*. *IDMO* is equipped with short range connectivity to local *Mobile Hub*, as it shown in the Fig. 3.

For this brief low-powered connectivity, we have designed general and technology independent communication stack of protocols and *APIs*. For it first hardware realization we have used *Bluetooth* 4.0 *LE Smart* (*Low Energy*/*Smart solutions*). *BLE* [9] is very promising technology, that being made available the growing amounts of mobile operating systems, such as *Android*, *iOS* etc., and supports approximately 2000 simultaneous connections. As a programmable system, our middleware implements the *Scalable Data Distribution Layer* (*SDDL*) [10] which connects stationary nodes in a wired "core" network with all the *IDMO*. *MR-UDP* [11] aims at providing reliable communication based on *UDP* from/to mobile nodes through extending a reliable *UPD* protocol.

A key issue is arising in the solution of problems of cloud-assisted management level is selecting a way of representing information about the state of the control

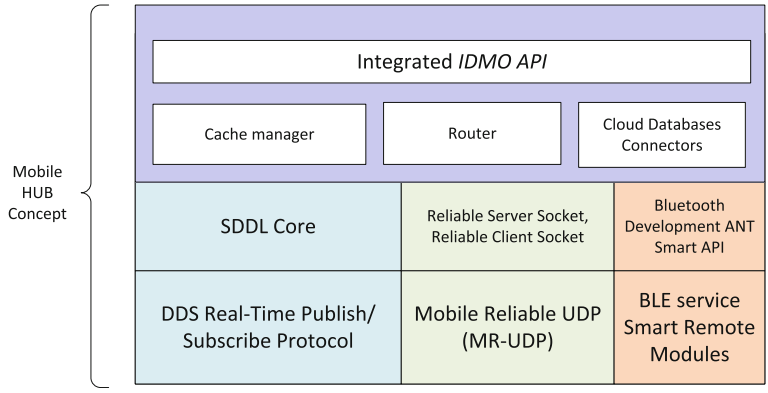


Fig. 3 Proposed concept of local *Mobile HUB* for *IDMO*

object and the state of the external environment, i.e. the management system needs to use a particular method of describing situations and transitions between them. Under the situation, in this case, refers to a set of environmental objects that make up the situation, relations between them and the processes that involve these objects. In this regard, in each control system, containing a strategic level should be solved the following partial tasks [12]:

- storing descriptions of objects and processes;
- storage of connections between objects, both temporary and permanent, the implementation of a mechanism for establishing relationships between objects based on their description;
- storage of actions that transform one situation to the other;
- storage situational goals, mid-term and long-term, the achievement of which will go towards management;
- the translation of ideas from lower levels of management in a situation and back translation of action into low-level control actions;
- planning situations;
- provide feedback, i.e. development of evaluation mechanisms of action by the current plan and the current target;
- the formation of coalitions and distribution of roles for the overall task.

IDMO may use a different cloud storage to expand memory. Having 1–32 GB memory on a mobile device, one can use the extra 1 TB from the cloud, without increasing the physical size of the device and thus cost. However, these services use full data synchronization that is not applicable to mobile devices with low memory. For efficient use of cloud drives in the *IDMO* is the effective caching of data, i.e. the choice of which data should reside on the device and which remain in the cloud. The system module that performs this function is called by the cache manager. It is in our middleware. In various technical solutions, the cache manager uses various data to determine the cached subset of data, in particular: caches all the data (synchronization); the user sets cached data; predict cached data, based on statistics of querying data from the cloud. As an intermediary for data transfer used a router that one can access from a mobile object or device is carried out wirelessly. The router is a "computing environment" (*PC*, cluster, distributed system), which contains three software modules: a storage module, the prediction module, and the delivery module. On a mobile device installed specialized software that allows to make requests to various network resources via the router and to receive the requested data. When the user makes a request, the request is sent to the router over a wireless network, where the query is stored in the module storage. Moreover, finally, the connector to cloud storage is a driver, adapted to a particular type of server's cloud storage.

We estimate some distance and energy parameters of proposed middleware in *MICS*, for a short-range warning and notification system for train conductors as it shown in Fig. 4.

Each *Mobile Hub* in passenger car connected with other through inter-train *Ethernet* network and they organize *Bluetooth LE Smart Pico Network* within a car.

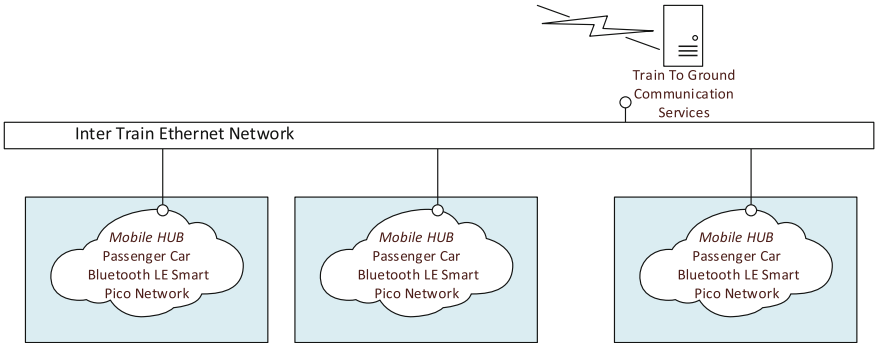


Fig. 4 A testbed for *Mobile HUBs*: warning and notification system for train conductors

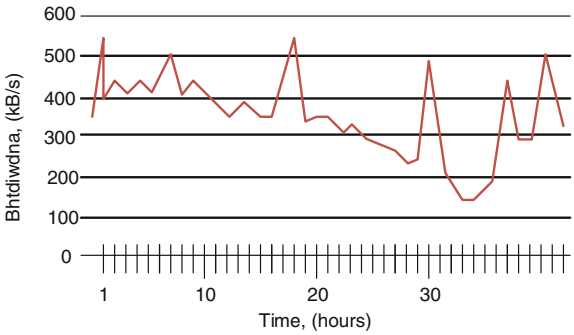


Fig. 5 An estimated bandwidth of active *BLE Smart* connection

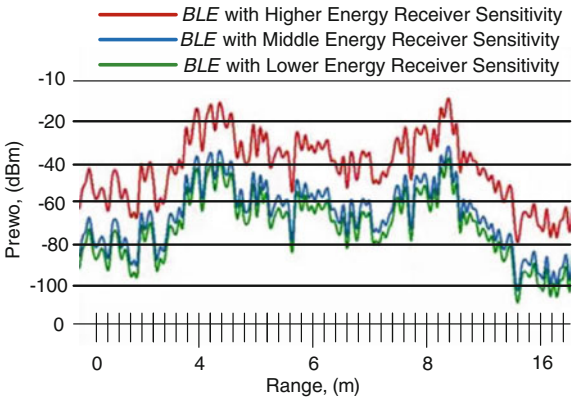


Fig. 6 Estimation of BLE received power with higher and lower energy receiver sensitivity

Train conductors use their mobile phones to connect other network through *Bluetooth* connection. We estimated the bandwidth and range of received power of active link, and the results are depicted in Figs. 5 and 6 respectively.

As it follows from Figs. 5 and 6 we received decent conditions of bandwidth and electricity/range performance.

5 Conclusions

In the paper, we propose a new cloud-assisted middleware for intelligent distributed and mobile objects as part of the multilevel intelligent control system, which is owned and has been developed by *JSC* "*Russian Railways*." We carried out a detailed related work that concerns the construction of complex technical moving objects control systems, which is functioning in the automated and autonomous modes. Next we highlighted the advantages and drawbacks the intelligent distributed and mobile objects control systems. We propose a new cloud-assisted architecture for intelligent distributed and mobile objects as part of the multilevel intelligent control system. Also, we propose technical architecture of system. As a testbed of proposed technical architecture, we received decent conditions of bandwidth and power/range performance of *Mobile HUB* concept for passenger train car as part of pico network with *Bluetooth LE Smart* technology.

Acknowledgments This work has been financially supported by the Russian Foundation for Basic Research (Grants No. 15-08-01886-a, No. 16-01-00597-a).

References

1. Joint Architecture for Unmanned Systems. https://en.wikipedia.org/wiki/JAUS
2. Jameson, S., Franke, J., Szczerba, R., Stockdale, S.: Collaborative Autonomy for Manned/Unmanned Teams. AHS International Forum 61, Grapevine, TX (2005)
3. Albus, J., et al.: 4D/Real-time Control System (4D/RCS): A Reference Model Architecture for Unmanned Vehicle Systems v2.0, NIST, NISTIR 6910 (2002)
4. Narayan, P., Wu, P., Campbell, D., Walker, R.: An intelligent control architecture for Unmanned Aerial Systems (UAS) in the National Airspace System (NAS). In: Proceedings of 22nd International Unmanned Air Vehicle Systems Conference (Melbourne, Australia), pp. 1–11 (2007)
5. Laird, J.: Extending the SOAR cognitive architecture. In.: Proceedings of the First AGI Conference, pp. 224–235 (2008)
6. Langley, P.: Cognitive architectures and general intelligent systems. Artif. Intell. Mag. **27**, 33–44 (2006)
7. Sun, R.: The Clarion cognitive architecture: extending cognitive modeling to social simulation, 434 pp. Cambridge University Press, New York (2006)
8. Chernov, A.V., Butakova, M.A., Karpenko, E.V.: Security incident detection technique for multilevel intelligent control systems on railway transport in Russia. In: Telecommunications Forum Telfor (TELFOR), 2015 23rd, pp. 1–4 (2015)
9. Bluetooth Smart (Low Energy) Technology. https://developer.bluetooth.org/TechnologyOverview/Pages/BLE.aspx
10. Scalable Data Distribution Layer (SDDL). http://lac-rio.com/sddl/
11. Mobile Reliable UDP (MR-UDP). http://lac-rio.com/mr-udp/
12. Yants, V.I.; Chernov, A.V.; Butakova, M.A.; Klimanskaya, E.V.: Multilevel data storage model of fuzzy semi-structured data. In: 2015 XVIII International Conference Soft Computing and Measurements (SCM), pp. 112–114. IEEE Press, New York (2015)

Intellectualization of Technological Control of Manufacturing Processes on Railway Transport Based on Immunological Models

Alexander Dolgiy, Sergey Kovalev, Vladimir Samsonov and Agop Khatlamadzhiyan

Abstract The article develops a new immunological approach to solving the tasks related to the information-and-technological control of manufacturing processes which are represented by the complex time-series. The article offers an innovative immunological model for detecting deviations in the time-series, which is based on simulation of interaction processes between antibodies and antigens within the biologically inspired immune system.

Keywords Complex time-series · Immunological model · Antibodies · Antigens · Temporal proposition · Interpretation of temporal formulas · Track maintenance trains · Intelligent computer vision system

1 Introduction

The article presents the possibility to apply the principles of the immune networks functioning for the real-time control of technological processes of servicing the railway cars using the track maintenance trains. The model of technological process is represented as a complex time-series which describes the dynamic of changes of one or several typical patterns of the controlled object, which are recorded by the intelligent computer vision system and primary information sensors. The task for

A. Dolgiy · S. Kovalev · V. Samsonov (✉) · A. Khatlamadzhiyan
Rostov State Transport University, Rostov-on-Don, Russian Federation
e-mail: samsonov.sem@gmail.com

A. Dolgiy
e-mail: adolgy@list.ru

S. Kovalev
e-mail: ksm@rfniias.ru

A. Khatlamadzhiyan
e-mail: agopx@yandex.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_28

the diagnostics system is to predict and detect any important deviations of a time-series with respect to the normal development of the controlled technological process which is presented by this time-series.

2 Problem Statement

Let us present the versatile model of the controlled technological process in the form of a complex time-series (CTS):

$$S = (\bar{s}_1, \bar{s}_2, \ldots, \bar{s}_n), \tag{1}$$

Whose elements are the vector values $\bar{s}_i = (x_i^1, x_i^2, \ldots, x_i^k)$, which define (distinguish) the certain states of the process in the discrete instants of time t_i.

The vector values $\bar{s}_i \in S$ are acquired through the physical measuring by the primary information sensors of the controlled process's parameters, which are important for making diagnostics decisions. For instance, when controlling the technological process of servicing the gondola cars and hopper-cars prior to supplying them to the track maintenance trains' premises or during their presence over there, the most important figures will be the exterior of railway cars, load footprint, collocation of certain units of loading cars with respect to the railway cars, railway cars speed, weight and other parameters recorded by the intelligent computer vision system and sensors.

The task is to generate, based on the analysis of the training set O, the system of classifying rules which allow to recognize as "normal" all samples from the set O and those related to them according to the criterion J, and to consider the realizations of the time-series, which significantly differ from the "normal" ones, as abnormal.

For the purpose of solving the present task, the article offers an approach based on the analysis of the CTS, which represent the controlled process, for detecting deviations present within it, which would be reasoned by some technological failures or fails, using the methods of the artificial systems theory.

3 Immune Approach to Technological Control

The biologically inspired immune system comprises several functionally different components, which provide for the multi-level control and protection for the body from the external viruses. The main role of the present system is to detect all the cells of the body and to classify them according to their types—"friend" or "foe". In the course of evolution, the immune system learns to recognize foreign cells (antigens) and the own body cells. The recognition of antigens is performed by specific types of cells which have the receptors on their surface called antibodies. There are two important biologically inspired processes that take part in functioning

of the immune system, and they are called affirmative and negative selection. The affirmative selection involves the generation of antibodies which recognize the own cells of the body, meanwhile the negative one involves the generation of antibodies which recognize only the foreign cells and those that do not react on the own molecules.

The immunological mechanism of negative selection has been widely spread among practical applications in the field of control and diagnostics [1]. For the first time the negative selection algorithm for detection of changes during the monitoring of processes, generated based on the "friend-or-foe" recognition principles, was introduced in [2] and has received a further development in papers [3–7].

The set of detectors is created in a random manner. In future, the detectors which recognize the "friend" cells are eliminated from this set. As a result, the remaining detectors can detect only not "friend" cells. These detectors are used as detectors for negative selection which form the base for the generation of classifying rules of diagnostics systems.

The immunological mechanism of the negative selection is the basis of the developed approach to the control of technological processes represented in the form of CTS. The negative selection algorithm with respect to the detection of deviations in CTS, is developed as follows.

1. The training set $O = \{KBP_1, KBP_2, ..., KBP_m\}$ is generated from the CTS samples which define the different possibilities of normal flow of the controlled process.
2. The formal description of the set "Me" (Me—description) is created based on the set O and expert information, in the form of the system of logic formulas representing the training set O.
3. Based on the set "Me", by generating the negations for the "Me—description" formulas, the set "Not Me" is created in the form of the system of formulas acting as detectors of the negative selection, based on which the diagnostics rules are generated.

4 Immunological Model's Main Components

The key issue during implementation of the immune recognition algorithms is the selection of models for representing the antibodies and the mechanism of their interaction with the antigens.

The model of antibody is presented in the temporal-proposition form, which consists of the sequence of the interval-and-temporal events (ITE).

Definition 1 The interval-and-temporal event is the tuple

$$\tau_i = (F_i, \Delta t_i), \tag{2}$$

where $F_i = F_i(x^1, x^2, \ldots, x^k)$—the logic formula describing the behavior of CTS at its certain time intervals; $\Delta t_i = [l_i^m, l_i^k]$—the variation range for duration of the temporal interval from the lower border l_i^m to the upper border l_i^k.

The model of antibody is presented in the form of the in-series temporal proposition (ISTP) [8].

Definition 2 The in-series temporal proposition is the sequence of consecutive ITE:

$$W = \tau_1 rts \tau_2 rts \ldots rts \tau_m \tag{3}$$

where *rts*—the time-bound relation of immediate sequence of the Allen's logic [9].

Consequently, each ISTP represents some sort of generalized temporal image describing an infinite set of particular realizations of CTS. Each realization of CTS, related to the ISTP as in (3), can be presented in the form of splitting of CTS into m segments of σ_i, which correspond to the m of ITE $\tau_i = < F_i, \Delta t_i >$ within the ISTP, in such a manner that each one of the segments will have a duration which would satisfy the restriction on duration of the corresponding ITE, and will contain the vector values which would satisfy the corresponding formula F_i of the ITE.

Formally, the compliance of the CTS with the in-series temporal proposition is established through interpretation of W over S.

Definition 3 The temporal interpretation of the ISTP W, which is represented by the proposition as in (3), over CTS S, which is represented by the proposition as in (1), is a partitioning (splitting) $\sum = (\delta_1, \delta_2, \ldots, \delta_m)$ of the S sequence into m consecutive segments, which correspond to the ITE within ISTP, whereby the length of each segment satisfies the restriction on duration of the corresponding ITE within ISTP, that is

$$\forall \delta_i \in \sum \quad l_i^m \leq L(\delta_i) \leq l_i^k \tag{4}$$

where $L(\delta_i)$—the duration of δ_i segment, l_i^m—the lower border of the duration of the ITE τ_i, l_i^k—the upper border of the duration of the ITE τ_i.

In order to have the CTS fully complied with the description presented in the form of the ISTP, it is necessary to have such temporal interpretation of the ISTP W over S, where the logic correspondence between all segments δ_i within the CTS S and the related ITE in W will be observed. The fact of compliance of a time-series with the description presented in the form of ISTP is established through the concept of validity (trueness) of the temporal proposition formula (3) with respect to the CTS (1).

Definition 4 The ISTP W is true with respect to the CTS S, in case there is the temporal interpretation $\sum$ for the ISTP W over S, which establishes the logic correspondence between all segments δ_i within CTS S and ITE within ISTP, that is

$$\begin{gathered} J(W|S) = 1 \\ \Leftrightarrow \\ \exists \Sigma = (\delta_1, \delta_2, \ldots, \delta_m) \quad \forall \delta_i \in \Sigma \quad (l_i^n \le L(\delta_i) \le l_i^k) \& (F_i(\delta_i) = 1) \end{gathered} \tag{5}$$

where $J(W|S)$—the true value of ISTP W with respect to the CTS S; $F_i(\delta_i)$—the true value of the formula F_i on the fragment δ_i of the CTS S.

The algorithm controlling the deviations in CTS is based on the negative selection algorithm. In order to realize it, it is necessary to generate a set of negative selection detectors which were acquired from negation of formula of trueness (validity) (5). Based on Definition 4, by the means of negation of formula (5), we can formulate the definition of falseness of ISTP W with respect to the CTS S.

Definition 5 The ISTP W is false with respect to the CTS S, in case there is no interpretation $\sum$ for ISTP W over S, whereby there is a logic correspondence between all δ_i segments within CTS S and ITE within ISTP, that is

$$\begin{gathered} J(W|S) = 0 \\ \Leftrightarrow \\ \forall \Sigma = (\delta_1, \delta_2, \ldots, \delta_m) \quad \exists \delta_i \in \Sigma \quad (L(\delta_i) < l_i^n) \vee (L(\delta_i) > l_i^k) \vee (F_i(\delta_i) = 0) \end{gathered} \tag{6}$$

5 Immunological Algorithm for Detecting Deviations

The knowledge base for immunological algorithm of detection of deviations in CTS, which is based on the negative selection principle, consists of negative selection detectors. The last ones are generated from the negations of formulas F_i, included into description of ITE of temporal propositions (3). In order to establish a fact of non-compliance of CTS S with the description presented in the form of ISTP W, it is necessary to consider the whole set of possible temporal interpretations W over S and for each one to establish a fact of "triggering" at least of one of the negative selection detectors, that is the falseness of the formula F_i. This procedure is the basis for the simulation mechanism for interaction between antibodies and antigens with the purpose to detect foreign ("foe") cells.

Review of the whole set of temporal interpretations W over S is the combinatorial task, however, in order to solve it, it is offered to use an effective approach having a polynomial estimation of complexity, based on principles of composite dynamic programming [10]. With that end in view, let us introduce in reviewing two auxiliary graph models.

Definition 6 Let $W = \tau_1 rts\, \tau_2 rts \ldots rts \tau_m$—ISTP, which is given using proposition as in (3), and S—CTS, is given using proposition as in (1). The for ITE $\tau_i = (F_i, \Delta t_i)$ *the temporal-pattern graph* will be the graph $R_{\tau_i} = (C, \Gamma_{\tau_i})$, which has

been defined at the vertex set $C = \{\vec{s}_i \in S\}$, identified with the elements $\vec{s}_i$ of the CTS S, the connections among which are defined by the Γ_{τ_i} relation as follows:

$$\Gamma_{\tau_i}(s_r, s_p) = 1 \quad \Leftrightarrow \quad (L(\delta_i) < l_i^n) \vee (L(\delta_i) > l_i^k) \vee (\neg F_i(\delta_i) = 1) \tag{7}$$

where $\delta_i = [\vec{s}_r, \vec{s}_p]$—the fragment of the CTS S.

Definition 7 For ISTP W and CTS S the compositional graph will be

$$R_{WS} = R_{t_1} \circ R_{\tau_2} \circ \cdots \circ R_{\tau_m} \tag{8}$$

which has been acquired trough the mini-max composition of the temporal-and-pattern graphs R_{τ_i}, taken in accordance with how they enter into W.

Based on the results of the paperwork [11] the following statement can be proved.

Theorem 1 *The ISTP W is false with respect to CTS S if and only if in the compositional graph $R_{WS} = R_{t_1} \circ R_{\tau_2} \circ \cdots \circ R_{\tau_m}$ there is*

$$\forall s_r, s_p \in C \quad R_{WS}(s_r, s_p) = 1 \tag{9}$$

Based on the mentioned theorem, the algorithm for detection of deviations in CTS has been implemented. The knowledge base for algorithm is the system of ISTP W_i, which describe all possibilities for "normal" flow of the controlled process, generated by the experts for the purpose of a particular task. The set of negative selection detectors is generated based on the ISTP, in the form of inverses of the logic formulas F_i, which are incorporated into ITE of temporal propositions. The algorithm for detection of deviations comes to the performance of the following steps.

1. Receive the subject-to-control CTS S at the model's input. For each W_i, incorporated into the DB of the immune model, the clauses 2–3 to be performed.
2. For ISTP W_i and CTS S the temporal-and-pattern graphs are created based on the Definition 6 and the compositional graph $R_{WS} = R_{t_1} \circ R_{\tau_2} \circ \ldots \circ R_{\tau_m}$ is generated based on the Definition 7.
3. In case that in R_{WS} there will be even one couple of vertexes $(s_r, s_p) \in C \times C$, for which $R_{WS}(s_r, s_p) = 0$, then make transfer to the clause 5, otherwise to the clause 2.
4. The clauses 2–3 are to be performed for all W_i, included into the DB of the immune model. In case that at each iteration of algorithm when performing the clause 3 there was no transfer made to the clause 5 then the deviation of CTS with respect to the normal flow of process is recorded. Stop.
5. The controlled process has no deviations. Stop.

6 Example of Immunological Model Creation

Let us consider the realization of the described above approach using the example of creation of immunological model for the control of the technological process of servicing the railway cars using the track maintenance trains [12].

The track maintenance trains (TMT) are designed to perform the reparation of the track and the ground bed, preventative and predictive track alignment by machine units, maintenance and reparation of the maintenance vehicles. The TMT fleet counts on mobile specific-purpose machines in order to perform above stated works. The diagram of the controlled technological process of loading the rolling units B_1—B_k with empty gondola cars and hopper-cars from stations « Station 1 »—« Station N » toward the TMT's premises is shown in the Fig. 1.

In the control area 1, the technical parameters regarding the important elements of the gondola cars' and hopper-cars' trucks are recorded in the real-time mode with further diagnostics of possible failures [13] using the T video sensors of the railway cars diagnostics system called *Technovizor* [14]. Then, the already loaded gondola cars and hopper-cars in the area 2 are checked by the T video sensor with respect to the allowed values for height of the ballast "hat", after that the loaded car leaves the territory of TMT's premises.

The process shown in the Fig. 1 shows how the empty B_k railway car passes the whole technological chain of loading the ballast and it is presented in the form of a graph, which is shown in the Fig. 2.

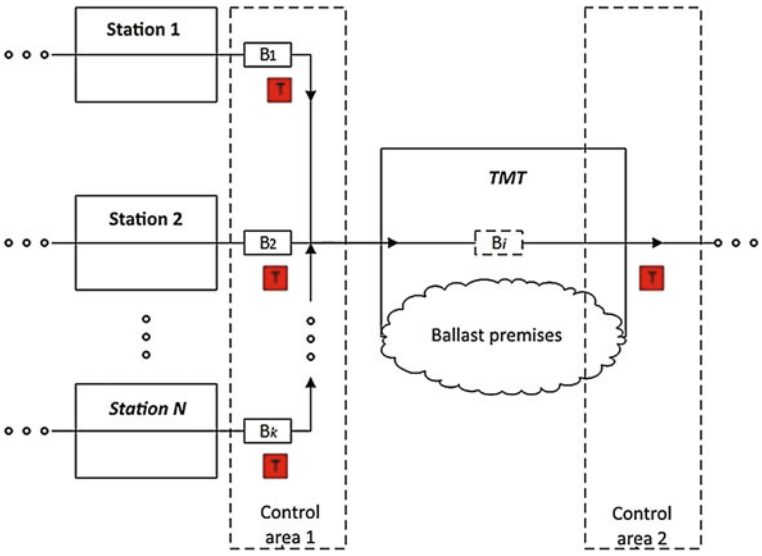


Fig. 1 Technological chain for delivering the gondola cars and hopper-cars to the place of loading the ballast on the territory of TMT's premises

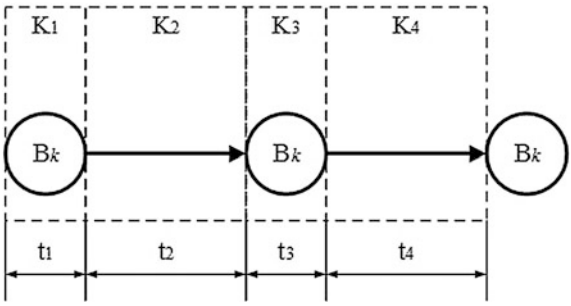


Fig. 2 The graph of transfers of the Bk railway unit from one type of work to the other type of work or state

The technological diagram includes 4 main fragments: K_1- automated "visual" inspection of empty railway units by the computer vision system; K_2—transfer of railway units from the place of inspection toward the place of loading the ballast; K_3—loading of ballast and control of specified parameters for loaded cars (for example, dimensions considering the ballast hat and the weight of loaded car); K_4—transfer from the place of loading of ballast and beyond the borders of the TMT's premises.

The formal presentation of technological process of loading the cars is CTS $S = (\bar{s}_1, \bar{s}_2, \ldots, \bar{s}_n)$, the elements of which are the vector values $\bar{s}_i = (v_i, m_i, \beta_i)$ which are recorded at each one of the instants of time t_i of the patterns: v_i—the current velocity at instant of the time t_i; m_i—the current weight, β_i—recorded-by-video-camera pattern of presence of the allowed height of the "hat" at the moment t_i. α_i—the recorded-by-computer-vision-system fact proving the readiness of the car for further operations (use).

Let us present the model of the normal flow of the technical process based on the diagram shown in the Fig. 2 in the form of the ISTP W which includes 4 ITE:

$$W = (F_1\Delta t_1)rts(F_2\Delta t_2)rts(F_3\Delta t_3)rts(F_4\Delta t_4) \tag{10}$$

where F_i—is the logic formula of description of the ith ITE, $\Delta t_i = [l_i^n, l_i^k]$—the variation range for duration of the ith ITE.

The expressions for F_i have been acquired based on the analysis of semantics of scripts of development of related fragments K_i of technological process and are as follows:

$$F_1 = (\exists t_i \in \Delta t_1)\ (\alpha_i = 1)$$
$$F_2 = (\forall t_i \in \Delta t_2)\ (v_i > 0),$$
$$F_3 = ((\forall t_i \in \Delta t_3)\ (v_i = 0))\&((\exists t_i \in \Delta t_3)\ (\beta_i = 1)\&(m_i \in \Delta\theta))$$
$$F_4 = ((\forall t_i \in \Delta t_4)\ (v_i > 0))\&(\beta_i = 1)\&(m_i \in \Delta\theta)$$

Here, besides the legend introduced earlier, the allowed range of variation of weight of the fully loaded car is identified using $\Delta\theta$.

The system of negative selection detectors is generated based on the inversion of the logic formulas of ITE:

$$\begin{cases} \neg F_1 = (\forall t_i \in \Delta t_1)\ (\alpha_i = 0), \\ \neg F_2 = (\exists t_i \in \Delta t_2)\ (v_i = 0), \\ \neg F_3 = ((\exists t_i \in \Delta t_3)\ (v_i > 0)) \vee ((\forall t_i \in \Delta t_3)\ (\beta_i = 0) \vee (m_i \notin \Delta\theta)), \\ \neg F_4 = ((\exists t_i \in \Delta t_4)\ (v_i = 0)) \vee (\beta_i = 0) \vee (m_i \notin \Delta\theta). \end{cases} \tag{11}$$

When the controlled CTS S based on (11) is received at the input of immunological model, the temporal-and-patter graphs for all ITE, included into ISTP W, are generated:

$$\begin{gathered} R_{\tau_i} = (C, \Gamma_{\tau_i}), \\ \Gamma_{\tau_i}(s_r, s_p) = 1 \\ \Leftrightarrow \\ (L(\delta_i) < l_i^m) \vee (L(\delta_i) > l_i^k) \vee (\neg F_i(\delta_i) = 1) \quad (i = 1, 2, 3, 4). \end{gathered}$$

Based on the temporal-and-pattern graphs, we build the compositing graphs

$$R_{WS} = R_{t_1} \circ R_{\tau_2} \circ R_{\tau_3} \circ R_{\tau_4}.$$

The decision rule of immunological model is based on the analysis of all elements within the R_{WS} compositing graph matrix which has been generated for the given CTS S. In case that all elements of the matrix equal to 1, then the fact of the technological process deviation from the normal behavior is recorded.

7 Conclusion

The article presents the possibility for application of the principles of the immune networks functioning in order to solve the tasks related to the information-and-technological control of the manufacturing processes presented by the complex time-series. The developed immunological algorithm for detecting the deviations in the time-series has a polynomial estimation of algorithmic complexity, which allows to use it in the real-time mode diagnostics systems. The algorithm has been approved for solving a range of tasks related to the on-line control of technological processes of servicing the railway cars using the track maintenance trains.

The offered approach can be used for automatic monitoring of the railway infrastructure plants, trackside equipment and technological processes with an increased hazard level.

Acknowledgments Work is done with support of RFBR grants №16-07-00994 a.

References

1. Forrest, S., Perelson, A.S., Allen, L., Cherukuri, R.: Self-nonself discrimination in a computer. In: Proceedings of IEEE Symposium on Research in Security and Privacy. Oakland, CA (1994)
2. D'haeseleer, P., Forrest, S., Helman, P. An immunological approach to change detection: algorithms, analysis, and implications. In: Proceedings of IEEE Symposium on Research in Security and Privacy. Oakland, CA (1996)
3. Ishida, Y.: Fully distributed diagnosis by PDP learning algorithm: towards immune network PDP model. In: Neural Networks, pp. 777–782. San Diego, CA (1990)
4. Ishida, Y.: An immune network model and its applications to process diagnosis. Syst. Comput. Jpn. **24**(6), 38–46 (1993)
5. Jerne, N.K.: The immune system. Sci. Am. **229**(1), 52–60 (1973)
6. Jerne, N.K.: Towards a network theory of the immune system. Annales de l'Institut Pasteur/Immunologie. **125C**, 373–389 (1974)
7. Kovalev, S.M., Ternovoy V.P.: Immunnyj podhod k vyyavleniyu anomalij vo vremennyh ryadah. In: OPiM. **18**(4) (2011). (The immune approach to the detection of abnormalities in the time series)
8. Kovalev, S.M.: Modeli analiza slabo formalizovannyh dinamicheskih processov na osnove nechetko-temporal'nyh sistem. Izvestiya vuzov. Sev.-Kav. region. Estestvennye nauki. **2**, 10–13 (2002). (Analysis models for slightly formalized dynamic processes based on the fuzzy temporal systems)
9. Allen, J.F.: Towards a general theory of action and time. Artif. Intell. **23**(2) (1984)
10. Vinzuyk, T.K.: Analiz, raspoznavanie i interpretaciya rechevyh signalov. Naukova dumka, Kiev (1987). (Analysis, recognition and interpretation of voice signals)
11. Kovalev, S.M.: Interpretiruyushchie modeli dlya nechetko-temporal'nyh skhem vyvoda v intellektual'nyh sistemah dinamicheskogo tipa. In: Proceedings of Devyataya nacional'naya konferenciya po iskusstvennomu intellektu s mezhdunarodnym uchastiem KII2004, vol. 1, pp. 378–385. Fizmatlit, Moscow (2004). (The interpreting models for fuzzy temporal output circuits in the intelligent systems of dynamic type)
12. Instrukciya po tekhnicheskomu osnashcheniyu putevyh mashinnyh stancij: JSC "VNIIZHT", Moscow (2012). (Instruction on equipment of the track maintenance trains)
13. Dolgiy, A.I., Kovalyev, S.M., Samsonov, V.L., Khatlamadzhiyan, A.E.: Processing of fuzzy graphic images in intelligent computer vision systems on railway transport. In: Proceedings of 9th International Conference on Application of Information and Communication Technologies (AICT), pp. 118–122. IEEE (2015)
14. Adadurov, A.S., Bushuev, R.Y., Dolgiy, A.I., Khatlamadzhiyan, A.E.: Post kompleksnogo kontrolya kak innovacionnyj podhod k diagnostike hodovoj chasti vagona. Vagony i vagonnoe hozyajstvo. **4**(44), 24–27 (2015). (The complex control station as innovative approach to diagnostics of the undercarriage)

Intelligent Methods for State Estimation and Parameter Identification in Fuzzy Dynamical Systems

Sergey Kovalev, Sergey Sokolov and Alexander Shabelnikov

Abstract This paper presents a new approach for state estimation and parameter identification in fuzzy dynamical systems. The basis of the proposed approach is adaptive network calculation model of the fuzzy prior and posterior estimates of system state variables taking place in consecutive time steps. The optimization of model parameters based on modified simplex algorithm is also proposed. Presented method for parameter identification has also a set of new properties, such as ability of integration in the expert systems, higher level of potential accuracy and possibility of real-time identification. Example of optimal parameter estimation for fuzzy dynamical system is considered and results of the experiments are provided. Experiments show that estimations of identified parameters obtained on the basis of adaptive network applied in dynamical systems of Sugeno type does not deviate from real values by more than in 10 %.

Keywords Fuzzy dynamical system · Conditional membership function · Prior fuzzy distribution · Posterior fuzzy distribution · Adaptive network model · Parametric identification

1 Introduction

Recent methods for control of complex dynamical objects, which work is connected with uncertainty, are based on the analytical models, which are represented in form of differential and recurrence equations. Analysis of related publications [1–4] shows that the major part of researches uses traditional methods, which has a set of limitations, such as requirement of subordination to normal distribution law, usage

S. Kovalev (✉) · S. Sokolov
Rostov State Transport University, Rostov-on-Don, Russia
e-mail: ksm@rfniias.ru

S. Kovalev · A. Shabelnikov
JSC "NIIAS", Rostov branch, Rostov-on-Don, Russia
e-mail: info@rfniias.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_29

of traditional mean-square criteria for estimation of the optimality of model parameters, usage of the simplest linear models for measurer systems, etc. Here, the questions, which deal with integration of the empirical knowledges of human experts into poorly formalized process models characterized by incompleteness, imprecision and contradictory [5] as well as presence of fuzzy and subjective factors affected on parameter estimation, are still practically unexplored [6–8].

Nowadays, intelligent models are more preferable for modeling of poorly formalized objects. These models are based on knowledges, the main class of which is fuzzy dynamical systems (FDS) [2–4]. The basis of FDS is formalization of the empirical experience and knowledges of human experts, which is represented in linguistic form via fuzzy logic tools. To make FDS be practically usable in controlling systems, the development of effective construction methods as well as estimation and state correction algorithms is required. It is also important to adapt FDS for the real conditions.

This paper considers the decision of general problems, which are referred to identification, prediction and estimation of FDS states, which describe behavior of poorly formalized dynamical objects.

2 Model of Representation and State Estimation for FDS

Among of many well-known ways for discrete-time FDS representation [9–11], the most simple one is FDS construction in form of recurrence equation [7]:

$$x_{k+1} = F(x_k) \quad (k = 0, 1, \ldots, N), \tag{1}$$

where x is the state from the state space X of FDS, F is the fuzzy mapping $x_k \rightarrow x_{k+1}$, which is given by membership function (MF) $\mu_F(x_k, x_{k+1})$. This function is also convenient to be represented in form of conditional MF $\mu_F(x_{k+1}|x_k)$.

Real applications consider both FDS and the measuring system, which realize the transformation of states from X into the external observations from Z taking into account affecting noises. Thus, practically useful model of FDS may be represented considering measurer errors and fuzzy noises in form of the following system:

$$\begin{cases} x_{k+1} = F_k(x_k, \varepsilon_k) \\ z_k = S_k(x_k, \delta_k) \end{cases} \quad k = 0, 1, \ldots, N, \tag{2}$$

where F_k is the state equation for FDS; S_k is the nonlinear function of measurer work; x_k is the internal state of a system; ε_k is the fuzzy system noise described by a defined MF μ_{ε_k}; δ_k is the fuzzy measurer error described by a defined MF μ_{δ_k}; k is the discrete time index.

State space X for (2) is the set of values characterizing the position of a FDS in an observed time step and playing the role of initial conditions for the future system behavior. However, real system has uncertainties and, thus, dependency between

current and future values is represented by a fuzzy variable. Estimation and correction of this dependency are very important for dynamical objects control in the area of fuzzy modeling. The task of estimation and correction is defined as follows.

Let the initial information about FDS be presented in form of MF $\mu(x_0)$ and the observed states be presented by vector of observations $Z_k = (z_0, z_1, \ldots, z_k)$ from time interval $[t_0, t_k]$. It is required to predict fuzzy state x_{k+1}, which is defined by MF $\mu(x_{k+1})$. FDS correction is performed by specifying the MF for the determined fuzzy value of x_{k+1}, when z_{k+1} is observed at time step t_{k+1}.

3 Recurrent Algorithm for FDS State Estimation

Estimation and correction for the FDS states are performed based on determination and matching prior information and posterior one characterized by conditional MFs $\mu(x_{k+1}|Z_k)$ и $\mu(x_{k+1}|Z_{k+1})$. The conditional MFs are determined by the following recurrence procedure.

Let the FDS be nonstationary system with discrete time presented by (2). Let initial state MF $\mu(x_0)$ be a priory determined. Measurement errors, noises and states are independent fuzzy variables [12].

Based on assumption that $Z_{k+1} = (Z_k, z_{k+1})$, MF $\mu(x_{k+1}|Z_{k+1})$ can be presented as

$$\mu(x_{k+1}|Z_{k+1}) = \mu(x_{k+1}|Z_k, z_{k+1}). \tag{3}$$

Conditional fuzzy variables $\mu(x_{k+1}|Z_k$ and $(x_{k+1}|Z_{k+1})$, which belong to (3), are independent because fuzzy noises ε_k and δ_k affecting on FDS (determining $(x_{k+1}|Z_{k+1})$) and measurer (determining $(x_{k+1}|Z_{k+1})$), respectively, are also independent. Thus, Eq. (3) can be defined as follows:

$$\mu(x_{k+1}|Z_k, z_{k+1}) = \mu(x_{k+1}|Z_k) \,\&\, \mu(x_{k+1}|z_{k+1}). \tag{4}$$

Conditional MF of fuzzy variable $(x_{k+1}|z_{k+1})$ may be expressed by using measurer function from (2) via the MF of fuzzy noise:

$$\mu(x_{k+1}|z_{k+1}) = \mu(\delta_{k+1}) \quad \left(\delta_{k+1} = S_{k+1}^{-1}(x_{k+1}, z_{k+1})\right). \tag{5}$$

Since nonlinear mapping S_{k+1}^{-1} from (5) is commonly multivalued, the fuzzy estimation for $\mu_{\delta_{k+1}}(S_{k+1}^{-1}(x_{k+1}, z_{k+1}))$ takes maximum possible value via Zadeh extension principle [13]:

$$\mu_{\delta_{k+1}}(S_{k+1}^{-1}(x_{k+1,}z_{k+1})) = \sup_{\Delta = S_{k+1}^{-1}(x_{k+1}, z_{k+1})} \mu_{\delta_{k+1}}(\Delta). \tag{6}$$

Merging (6) and (5), the following equation can be got:

$$\mu(x_{k+1}|z_{k+1}) = \sup_{\Delta=S_{k+1}^{-1}(x_{k+1},z_{k+1})} \mu_{\delta_{k+1}}(\Delta). \tag{7}$$

Conditional MF $\mu(x_{k+1}|Z_k)$ from (4) describes fuzzy mapping $\Phi : Z_k \to x_{k+1}$, which can be represented in form of the composition of fuzzy mappings:

$$\Phi = (Z_k \to x_k) \circ (x_k \to x_{k+1}). \tag{8}$$

Conditional MF $\mu(x_k|Z_k)$ characterizes fuzzy mapping $Z_k \to x_k$ as well as fuzzy conditional MF $\mu(x_{k+1}|x_k)$ characterizes fuzzy mapping $x_k \to x_{k+1}$. As a result of composition, we get MF $\mu(x_{k+1}|Z_k)$ for fuzzy mapping $\Phi : Z_k \to x_{k+1}$:

$$\mu(x_{k+1}|Z_k) = \sup_{x_k}[\mu(x_k|Z_k) \;\&\; \mu(x_{k+1}|x_k)], \tag{9}$$

where $\mu(x_{k+1}|x_k)$ is expressed via fuzzy noise based on the state equation (2):

$$\mu(x_{k+1}|x_k) = \sup_{\Delta=F_k^{-1}(x_{k+1},x_k)} \mu_{\varepsilon_k}(\Delta). \tag{10}$$

If expression (10) is merged with Eq. (9), MF $\mu(x_{k+1}|Z_k)$ can be calculated as

$$\mu(x_{k+1}|Z_k) = \sup_{x_k}[\mu(x_k, Z_k)\& \sup_{\Delta=F_k^{-1}(x_{k+1},x_k)} \mu_{\varepsilon_k}(\Delta)]. \tag{11}$$

Considering (9) and (11), the final recurrence relations for finding the posterior MF of the fuzzy state of FDS at certain step (k + 1) can be expressed in following form:

$$\begin{cases} \mu(x_{k+1}|Z_{k+1}) = \mu(x_{k+1}, Z_k)\& \sup\limits_{\Delta=S_k^{-1}(x_{k+1},z_{k+1})} \mu_{\delta_{k+1}}(\Delta) \\ \mu(x_{k+1}|Z_k) = \sup\limits_{x_k}[\mu(x_k|Z_k)\& \sup\limits_{\Delta=F_k^{-1}(x_{k+1},x_k)} \mu_{\varepsilon_k}(\Delta)] \end{cases}, \tag{12}$$

Initial information for implementation of (12) is presented by $\mu(x_0|z_0)$, which is taken in the form of initial fuzzy state $\mu(x_0)$ determined by the problem statement.

4 Optimal FDS Parameters Estimation

The task of parameter estimation refers to the problem of parameter identification, when the structure of FDS is a prior determined.

Let the structure of FDS is given in form of system (2), where nonlinear state function F_k and measurer one S_k depend on the set of uncertain parameters at each

time step. The parameters are presented by vector A_k and B_k, respectively. In this case, model of non-stationary FDS with uncertain parameters is expressed by the following system:

$$\begin{cases} x_{k+1} = F_k(x_k, A_k, \varepsilon_k) \\ z_k = S_k(x_k, B_k, \delta_k) \end{cases} \quad k = 0, 1 \ldots, N. \tag{13}$$

Let the observation set be presented in form of vector $Z = [z_0, z_1, \ldots, z_k]$, which is determined on time interval $[t_n, t_m]$. It is required to calculate parameters of FDS A_k and parameters of measurer B_k, which make the system behavior be similar to experimental observations as more as possible.

To formalize the term "as more as possible", the criterion of identification quality is introduced. This criterion characterizes the rate of correspondence between FDS and experiments. It is calculated based on the matching of prior MF $\mu(x_{k+1}|Z_k)$, which reflects the fuzzy estimation of state from X at the time step t_{k+1} considering the observation of z_k at time step t_k, and posterior MF $\mu(x_{k+1}|Z_{k+1})$, which reflects the fuzzy estimation of FDS state at time step t_{k+1} considering the observation of z_{k+1}. Difference between prior and posterior MF is expressed by fuzzy error e_k of current estimation of FDS. MF of the error has the following form:

$$\mu(e_k, A_k, B_k, B_{k+1}) = \sup_{x_{k+1}} \mu(x_{k+1}|z_k, A_k, B_k) \,\&\, \mu(x_{k+1} + e_k|z_{k+1}, B_{k+1}). \tag{14}$$

Optimality criterion J can be presented by any fuzzy criterion in form of non-linear dependency on both conditional and posterior MFs. Particularly, it is convenient to use minimum of deviation for MF of fuzzy estimation error e_k from its model function defined on interval $[e_{min}, e_{max}]$, i.e.:

$$J_k(A_k, B_k, B_{k+1}) = \int_{e_{\min}}^{e_{\max}} (r(e_k) - \mu(e_k|, A_k, B_k, B_{k+1}))^2 \, de_k, \tag{15}$$

where e_k is the current error of the estimation, $\mu(e_k|A_k, B_k, B_{k+1})$ is the MF of fuzzy estimation error (14), $r(e_k)$ is the model function, which is chosen according to the specifications of an identification task.

Estimation problem is concluded in calculation of such vectors A_k and B_k, which minimize criterion J, i.e.

$$\min_{A_k, B_k, B_{k+1}} J = \min_{A_k, B_k, B_{k+1}} \int_{e_{\min}}^{e_{\max}} (r(e_k) - \mu(e_k|A_k, B_k, B_{k+1}))^2 de_k. \tag{16}$$

5 Adaptive Network Model of FDS

Determination of the conditional MF for fuzzy state and optimization of the FDS parameters are made by using both adaptive network model (ANM) and iterative algorithm described below. The basis of ANM is the process of a prior and a posterior MF calculation for each time step from the interval $[t_n, t_m]$ and also their matching based on chosen criterion J.

Let output signal (or observed state of a system) z_k is observed at certain step $t_k \in [t_n, t_m]$. Then, the fuzzy state of FDS at time step t_{k+1} can be calculated based on composition of fuzzy relations $S_k^{-1} : z_k \to x_k$ and $F_k : X_k \to X_{k+1}$ determined by measurer equation and state one (13), respectively. Conditional MF $\mu(x_k|x_k)$ is determined for fuzzy relation S_k^{-1} based on measurer equation taking into account fuzzy noise δ_k:

$$\mu(x_k|z_k, \mathrm{B}_k) = \mu_{\delta_k}(S_k(x_k, \mathrm{B}_k) - z_k). \tag{17}$$

Conditional MF $\mu(x_{k+1}|x_k)$ is defined for fuzzy mapping F_k based on state equation taking into account fuzzy noise ε_k:

$$\mu(x_{k+1}|x_k, \mathrm{A}_k) = \mu_{\varepsilon_k}(F_k(x_k, \mathrm{A}_k) - x_{k+1}). \tag{18}$$

According to (17) and (18), MF $\mu(x_{k+1}|z_k)$ has the following form for composition $S_k^{-1} \circ F_k$:

$$\mu(x_{k+1}|z_k, \mathrm{A}_k, \mathrm{B}_k) = \sup_{x_k} \left[(\mu(x_k|z_k, \mathrm{B}_k) \,\&\, \mu_{\varepsilon_k}(F_k(x_k, \mathrm{A}_k) - x_{k+1})\right] \tag{19}$$

Expression (19) is a prior MF characterizing the fuzzy value of state from X considering observed state z_k at step t_k.

To calculate posterior MF $\mu(x_{k+1}|z_k)$, output signal z_{k+1} should be assumed. The posterior MF for fuzzy state x_{k+1} is calculated based on measurer equation (13) considering fuzzy noise δ_{k+1}:

$$\mu(x_{k+1}|z_{k+1}, \mathrm{B}_{k+1}) = \mu_{\delta_{k+1}}(S_{k+1}(x_{k+1}, \mathrm{B}_{k+1}) - z_{k+1}) \tag{20}$$

Calculations for the MFs and criterion J can be shown in form of network structure representing feedforward calculation illustrated by Fig. 1.

Figure 1 shows that each element implements separated stage of transformations for input signals z_k and z_{k+1} into conditional MF $\mu(x_{k+1}|z_k, \mathrm{B}_k, \mathrm{A}_k)$ and $\mu(x_{k+1}|z_{k+1}, \mathrm{B}_{k+1})$.

Output block calculates J according to the given input signals. Identification of the FDS parameters is performed by handling parameters of A_k of FDS and both B_k and B_{k+1} of measurer using backpropagation method [14] and modified Nelder-Mead simplex method [15].

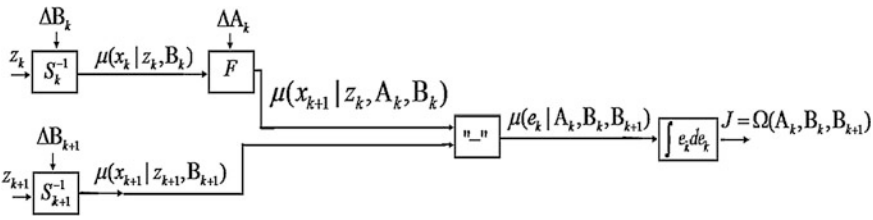


Fig. 1 General structure of ANM

6 Example

Implementation of above described method can be illustratively shown on the example of identification of the following FDS:

$$\begin{cases} x_k = f(x_{k-1}, a_{k-1}) + \varepsilon_k = a_{k-1} \cdot x_{k-1}^{1.7} + \varepsilon_k \\ z_k = s(x_k) + \delta_k = b_k \cdot x_k^{1.2} + \delta_k \end{cases}$$

where a_{k-1} is the identified parameter of the system; b_k is the parameter of an observer; ε_k is the fuzzy noise presented in the system, which is represented by Gauss MF with zero mean and variance $\sigma_\varepsilon = 0.05$; δ_k is the fuzzy noise presented in the measurer, which is represented by Gauss MF with zero mean and variance $\sigma_\delta = 0.22$.

Let a_{k-1} equal 2, c_k equal 1.2 for all k. Then, MFs have the following forms:

$$\mu(\varepsilon_k) = \exp\,(-\frac{1}{2 \cdot 0,5} \cdot \varepsilon_k^2).$$

$$\mu(\delta_k) = \exp\,(-\frac{1}{2 \cdot 0,22} \delta_k^2)$$

To describe the optimality criterion, the minimization of the MF deviation for fuzzy estimation error e_k is used:

$$J = \min_{a_k} \int_{e_{\min}}^{e_{\max}} \left(r\,(e_k) - \mu(e_k, a_k)\right)^2 de_k.$$

Model function of error is determined on interval $[e_{\min}, e_{\max}] = [-1, 1]$ of its changing as follows:

$$\begin{cases} r(e) = e + 1 & if \quad -1 \le e < 0 \\ r(e) = -e + 1 & if \quad 0 \le e \le 1 \end{cases}$$

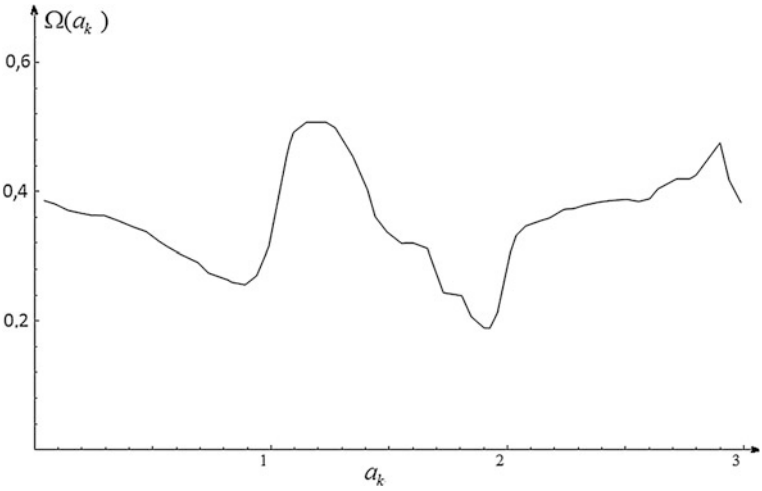


Fig. 2 Dependency between J and a_k for $k = 37$

Let initial state x_0 equal 0.8 and the calculation of $\sup_{x_k}(*)$ be provided with discretization step $\Delta = 0.05$ for x. Let also the integral from (15) be performed numerically utilizing quadratic formulas with step $\Delta = 0.05$ and infinite limits be changed by finite ones satisfying finite requirements for estimation algorithm ($x_{min} = 0$, $x_{max} = 4$). Here, the Nelder-Mead method [9] together with ANM optimizing a_{k-1} plays the main role for providing both the satisfactory computational speed and the required accuracy of results. Presented example consider the imitation of noises be generated programmatically using standard package of *Mathematica* software.

Calculation results are presented on figures below. Figure 2 illustrates curve of dependency between J and identified parameter a_k, when $k = 37$. The curve shows that the minimum of J is placed near to real value of $a_{k-1} = 2$.

Figure 3 presents the dependency between parameter a_k, which is required to be determined, and iteration step k of Nelder-Mead algorithm. Curve illustrates that a_k come around its real value and it deviate from this no more than by 10 %, when k increases.

The set of 400 experiments was performed to experimentally test efficacy of proposed approach. In the experiments, fuzzy Sugeno models with various numbers of unknown parameters (from 3 to 9) represent the FDS. Results show that the calculated estimations of a_k differ from their real values no more than by 10 % in the major part of samples (more than 95 %). Moreover, results proved that identified parameters of FDS are approximately converged to their real values after 20 −30 iterations of algorithm in the major part of samples. Small number of required iterations shows the practical possibility of using ANM for FDS identification in real time mode.

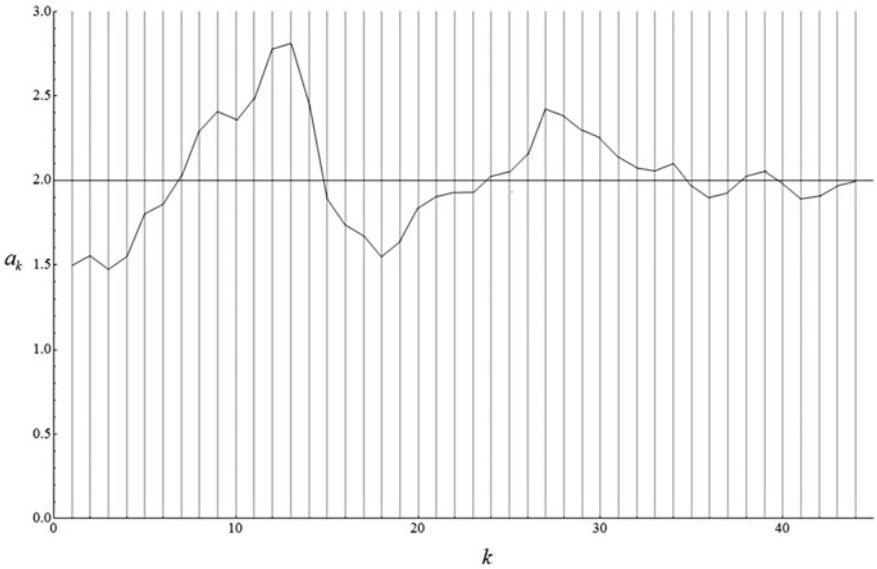


Fig. 3 Dependency between a_k and iteration number

7 Conclusions

This paper presents the new approach for estimation of states and identification of parameters in fuzzy systems describing the dynamics of poorly formalized processes. The proposed method utilizes adaptive network model and modified simplex algorithm. Experimental test of the method shows that determined parameters deviated no more than by 10 % from their real values in the major part of samples. Proposed approach for parameter identification has also the set of new properties, among which possibility for integration in the system of expert knowledges, higher level of accuracy versus traditional techniques and possibility of identification in real time are presented.

Acknowledgments The reported study was funded by RFBR, according to the research project (Grants No. 10-07-00158 and No. 10-01-00058).

References

1. Grop, D.: Metody identifikatsii system (The methods for system identification). Mir, Moscow (1979)
2. Lyung, L.: Identifikatsiya sistem. Teoriya dlya pol'zovatelya (System identification. Theory for user. Nauka, Moscow (1991)

3. Pashchenko, F. F.: Vvedeniye v sostoyatel'nyye metody modelirovaniya sistem. Autentifikatsiya nelineynykh system (Introduction to wealthy methods of system modeling. Autentification of nonlinear systems). Finansy i statistika, Moscow (2007)
4. Guo, R., Qin, Z., Li, X., Chen, J.: Interacting multiple model particle-type filtering approaches to ground target tracking. J. Comput. **3**(7), 23–30 (2008)
5. Pospelov, D.A.: Logiko-lingvisticheskiye modeli v sistemakh upravleniya (Logical-linguistic models in control systems). Energoizdat, Moscow (1981)
6. Gordon, N.J., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: IEE Proceedings F (Radar and Signal Processing), vol. 140, no. 2, pp. 107–113. IET Digital Library, London (1993)
7. Doucet, A., Godsill, S., Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering. Stat. Comput. **10**(3), 197–208, Springer, Heidelberg (2000)
8. Shteynberg, S.Y.: Identifikatsiya v sistemakh upravleniya (Identification in control systems). Energoatomizdat, Moscow (1987)
9. Kudinov, Y.I., et al.: Nechetkiye modeli dinamicheskikh protsessov Fuzzy models of dynamical processes). Nauchnaya kniga, Moscow (2007)
10. Terano, T., Asai, K., Sugeno, M.: Applied Fuzzy Systems. Academic Press, New York (2014)
11. Petrov, B.N., Ulanov, G.M., Goldenblat, I.I., Ulyanov, S.V.: Teoriya modeley v protsessakh upravleniya. Nauka, Moscow (1978)
12. Liu, B., Liu, B.: Theory and Practice of Uncertain Programming. Physica-verlag, Heidelberg (2002)
13. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning. Inf. Sci. **8**(3), 199–249 (1975)
14. Werbos, P.J.: Beyond regression: new tools for prediction and analysis in the behavioral sciences. Ph.D. thesis, Harvard University, Cambridge (1974)
15. Nelder, J.A., Mead, R.: A simplex method for function minimization. Comput. J. **7**(4), 308–313 (1965)

Safety Process Management Based on Software Risks Assessment for Intelligent Railway Control System

Vladimir D. Vereskun and Maria A. Butakova

Abstract In the paper, we propose new complex method devoted to the design of safety process management for intelligent transport systems, taking into consideration software risks assessment. Differences between the proposed method and widely known ones consist in the fact that it applies to a wide range of programmable intelligent control systems. A detailed explanation of required several steps for proposed method is implemented. The benefits of proposed software risks assessment for some class of intelligent railway control system are shown.

Keywords Safety management · Process management · Intelligent transportation system · Railway control system

1 Introduction

The railway transport is the leading element of transportation system of Russia carrying out now 86 % of cargo and 43 % of passenger traffic. Improvement of quality of transport services, to efficiency and safety of transport process is provided with application of the modern information and telecommunication technologies. The main aim of the development of information and telecommunication technologies on railway transport at present is a creation of a new class of systems with intelligent control facilities. These capabilities enable the radical turn from individual automation technologies and management operations to an integrated system that allows management of transportation processes in real time, implementation of comprehensive planning and the actual control of the whole railroad industry. *JSC* "*Russian Railways*" is developing such set of intelligent facilities, which is named *Intelligent Railway Control System* (*IRCS*). The prime directions of development of

V.D. Vereskun · M.A. Butakova (✉)
Rostov State Transport University, Rostov-on-Don, Russia
e-mail: butakova@rgups.ru

V.D. Vereskun
e-mail: vvd@rgups.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_30

IRCS are control of rolling and traction stock, infrastructure management, safety of railroad services and operations at stations and railway sections.

The basic elements of *IRCS* technological platform were defined as the set of technological modules enabling to realize the following functions:

- real-time train operations controlling;
- spatial and temporal synchronization of all business processes of the *JSC* "*Russian Railways*";
- uniform co-ordinate and temporal information space on the railway network;
- total situation awareness of dispatching staff;
- unification of information storage about technological processes and decision-making support in automated workplaces.

System integration of the above-specified basic functions of the *IRCS* in the form of a complex of hardware, software, information and telecommunication resources, mathematical models allows starting development and deployment of artificial intelligent methods in safety management tasks of railway infrastructure and transportation processes. The described conditions require the use of new methods of industrial safety, especially for constantly created new information and control systems with intelligent features. Among others, the least attention in scientific research was paid to the risks associated with faulty *IRCS* operations, that caused by software bugs and faults. Our work fills the gap and proposes new complex method devoted to safety management, which based on process analysis of risks, through the development of a new artificial neural network-based software reliability model.

The rest of the paper is organized as follows. In Sect. 2 we analyze the actual results of *IRCS* deployment and integrated methodology "*URRAN*" of risks and life cycle cost assessment on railway transport. We found it necessary because some years ago the *IRCS* and *URRAN* projects were adopted and now are used by *JSC* "*Russian Railways*". In Sect. 3 we give consideration to process safety management and software risk assessment during its life cycle. These aspects have been insufficiently researched, though there are ways to impact on the dependability of *IRCS* subsystems at design time. Section 4 is devoted to our proposed method and some illustrative numerical results.

2 Experience of *IRCS* and *URRAN* Projects

As we mentioned above the project *IRCS* was started a few years ago, in 2010, and now we can estimate some of its results of deployment. This project is the first intelligent system introduced directly in the production of *JSC* "*Russian Railways*". The *IRCS* project is coordinated by *JSC NIIAS* [1] and is being a key area of their research and development activity. Regarding the application of artificial intelligent methods and technologies the development of *IRCS* is based on the following principles:

- Usage of a single ontology and implementation platform. It allows continuous scaling and adapting the system.
- Usage of an adapting planning based on multi-agent technologies. It allows performing the effective correction in real time of various plans.
- Usage of a process-oriented approach. It ensures the completeness of the technological processes automatization and the elimination of the functional duplication.

In 2013, as a pilot project, and in 2015 into real traffic control at the "*Oktyabrskaya*" railway (in Russia) the complex *IRCS* tasks were introduced for the planning train operations. Nowadays, intelligent functions of *IRCS* have exploited at several locations: a range *Saint-Petersburg-Moscow*, a range *Saint-Petersburg-Buslovskaya*, a range *Saint-Petersburg-Nevel*, a primary railroad track of *East-Siberian Railway*, a main railroad track of *Baikal-Amur Mainline* (in Russia). *JSC* "*Russian Railways*" has plans to deploy *IRCS* on most railways in 2018. The experience of implementing *IRCS* shows that one of the central problems is the organization of intelligent decision support in controlling the traffic of trains in compliance with safety requirements. Such requirements with other constraints and their interaction with central part of *IRCS* are shown in Fig. 1.

Among other constraints from Fig. 1, we especially highlight the tasks, which are concerned with ensuring of railway safety. By this aim, *JSC* "*Russian Railways*" started *URRAN* project [2]. In large part, the *URRAN* uses the *RAMS* (*Reliability*, *Availability*, *Maintainability* and *Safety*) methodology [3], standards EN 50126 and IEC 62278. Besides, the *URRAN* project is aimed to one of the priority tasks of innovative development—reduction in life-cycle cost of infrastructure and rolling stock, subject to a high level of hardware reliability and the required level of the transportation process safety. Some practical results of *URRAN* project implementation in early stages, also methodology normative, standards and guidelines applied for the cost management of life cycle processes in *JSC* "*Russian Railways*" one can find in [4–6].

From a technical point of view, *URRAN* is a decision support system, which aims to help staff make decisions in difficult conditions for a full and objective analysis of their field of activity. Depending on the nature of controlled processes and *IRCS* subsystems, different intelligent methods are used in decision-making support. They include information search, intelligent data analysis and data mining, knowledge acquisition from databases, cognitive and simulation modeling, artificial neural networks, evolve and evolutionary calculation, genetic algorithms, swarm intelligence, artificial immune systems, and others. A whole system architecture of information technology of complex management of reliability, safety, risks and resources with *URRAN* subsystem is described in Fig. 2.

A central part, corporate data storage of described system architecture in addition to direct data storage performs converting different data interchange formats and accessing to large volumes of document-oriented and semi-structured data [7]. By information which is had in data storage, is organized the functioning of two systems, namely: *URRAN* and a *System* of *Analysis Railway Traffic Safety* and

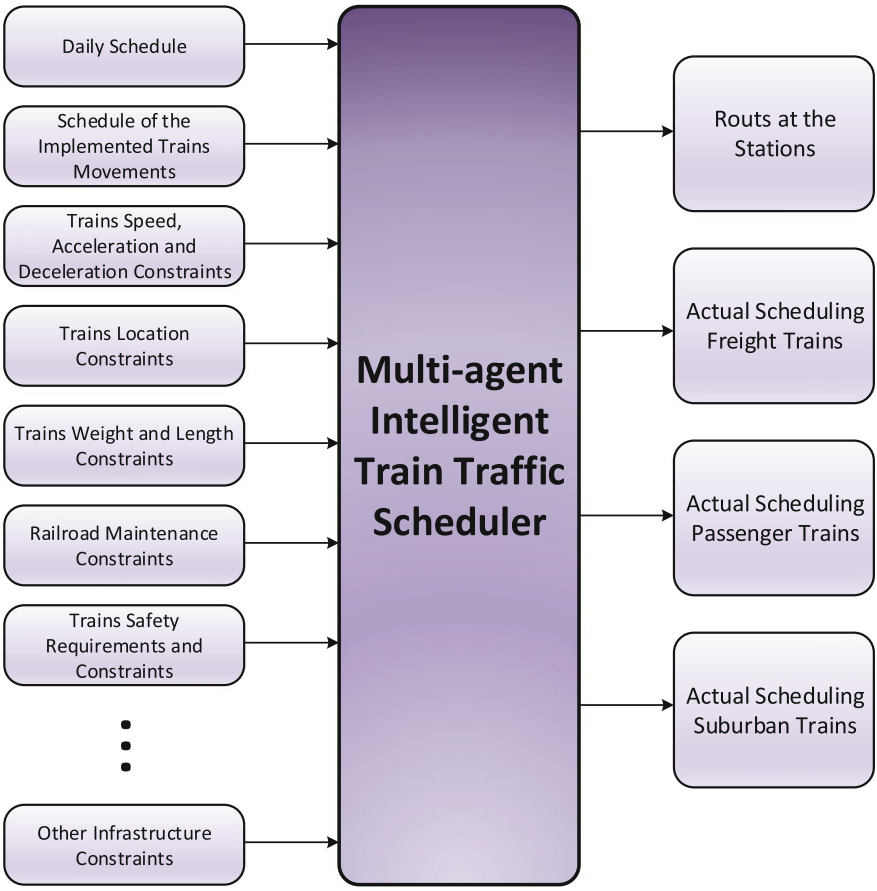


Fig. 1 The central part of *IRCS* multi-agent intelligent train traffic scheduler and its information constraints

Decision Support. These two systems enable the forming, measuring and managing the set of quantity measured parameters, such as reliability and risks metrics, the set of safety metrics and the life cycle costs. As a result, we have a unified coordinate system, allowing to assess the whole set of suggested decisions about transportations processes, which executed on railway enterprises. In practice, because of *URRAN* project, *JSC* "*Russian Railways*" uses the united methodology, dealing with processes of monitoring, registering, storing and processing staff faults and technical failures, coming from the real-time automated control systems, railway infrastructure, and traction and rolling stock parameters and locations.

Next, we briefly describe the main functions, that concerning to forming of risks metrics. The *System* of *Technical Failures Monitoring* and *Analysis* integrates with railway industry automated control systems in part of information about registered technical failures. The *System* of *Staff Faults Monitoring* and *Analysis* acquires

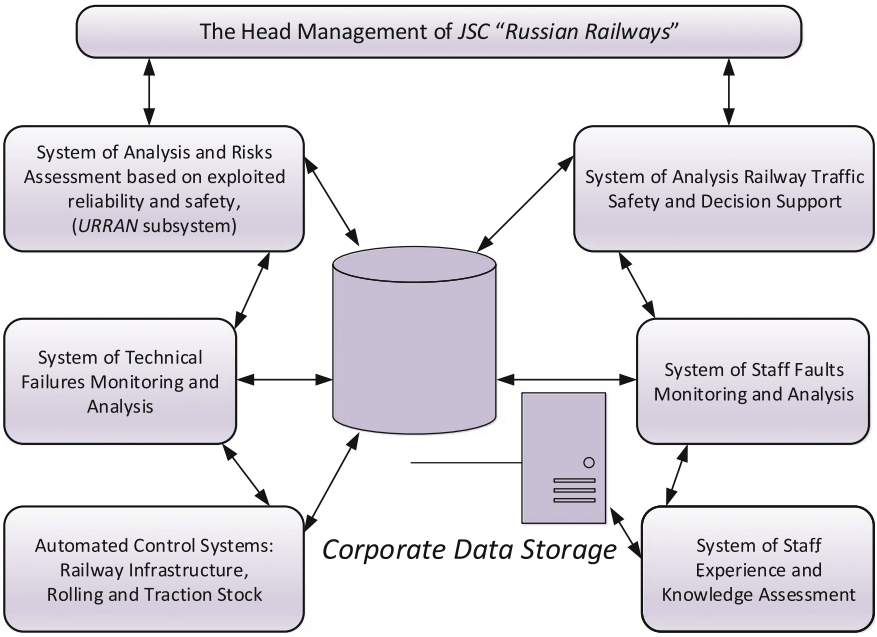


Fig. 2 System architecture of information technology of complex management of reliability, safety, risks and resources on railway transport

actual data from such events as a traffic control safety, investigations of violations of traffic safety conditions, the analysis of security requirements. Moreover, finally, *System* of *Staff Experience* and *Knowledge Assessment* is designed to obtain an objective evaluation of the stand-alone level of staff, which involved in railway transportation operations.

As a result, of implemented analysis, we conclude this section with following findings. *JSC* "*Russian Railways*" is interested in development and deployment of intelligent technologies and systems concerning reliability, safety and management of risks management in maintaining life cycle of railway infrastructure. The *RAMS* methodology, standing behind the *URRAN* project, has advantages such as completeness of the reliability analysis, the reliance on international standards and accurate assessment of risks at different levels. On another hand, we found some drawbacks. In *RAMS EN* and *IEC* standards, as well as national *GOST R* standards the risks of software malfunction have not been accounted. Only *EN* 50129 and *IEC* 62279 standards contain recommendations and nomenclature requirement of processes on software development and assurance for railway applications. In complex analysis does not take into account the durability of software. This circumstance does not allow to link durability and safety of the intelligent automation objects, also does not allow to assess the risks arising from the operation of the software and as a result of the update, as well as correctly assess the limit values of software-controlled objects. The cost of the life cycle of the operating software is

estimated in a separation from hardware reliability and safety of object of intelligent automation as these aspects are not included in *RAMS* methodology. The *RAMS* methodology is well suited for design stages and production of intelligent automation subsystems and is practically not developed for stages of their operation, updating, and utilization. For *IRCS* the risks and safety management of software-controlled intelligent subsystems at stages of their operation and modernization represents the main and important task.

Given the aspects mentioned above, in the following section, we will turn to the development of an approach of safety process management based on software risks assessment for *IRCS*.

3 The Proposed Approach

First of all, we adopt the following general risks management principles, also suited to intelligent automation systems:

- the decision connected with risk management has to be economically expedient and not make a bigger negative impact, than possible damage from consequences of actual realization of risk;
- risk management has to be coordinated with corporate strategy of the intelligent enterprise automation;
- decisions in risk management have to be based on sufficient amounts of credible information;
- decisions in risk management have to consider objective characteristics of an environment in which the intelligent system is functioning;
- risk management has to provide the analysis of the efficiency of earlier made decisions and correction of the principles and methods of management.

In Fig. 3 the general scientific view of our approach, adapted from [8] is presented.

As it follows from Fig. 3 the management of risk in safety process is an interactive process for such kinds of activity as risk assessment and risk processing. An interactive approach to risk assessment allows increasing depth and specification of an evaluation at each iteration. It is important that in a time of all process of safety risks management and risk processing, the exchange of information about incidents [9] between the head management and operational staff has to be performed.

The main difference of our approach is based on existence in the scheme a unit named "*Risks from Software Updates*". Let's examine how it impacts on safety and risks. The vast majority of subsystems as a part of *IRCS*, which realizes methods of artificial intelligence are implemented in the form of software-controlled devices. Surely, software plays increasingly important role in artificial intelligent methods implementation. At present, a large number of *IRCS* subsystems have been

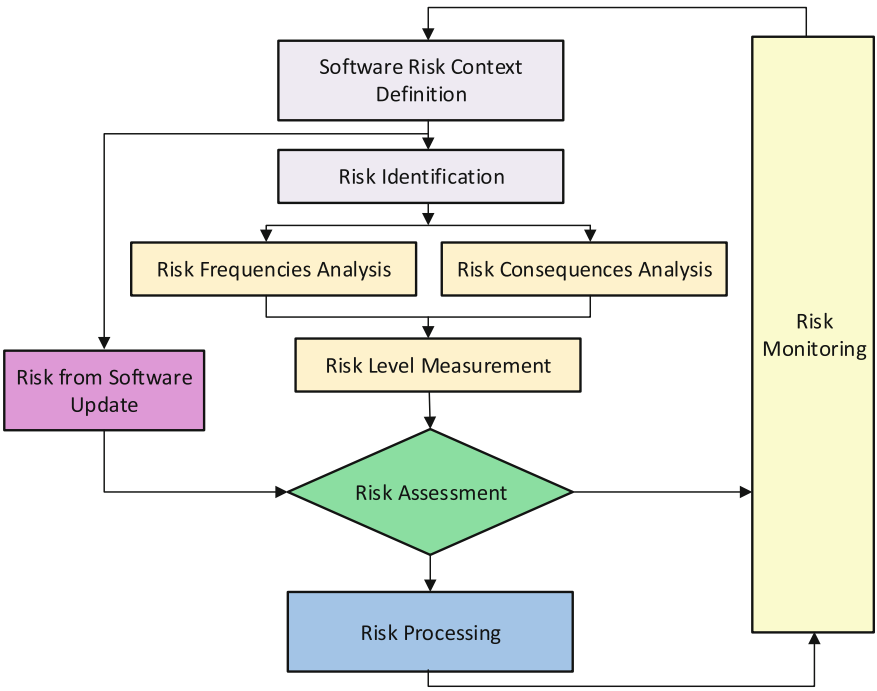


Fig. 3 Our diagram of software risk management

designed, and plenty of applications have been put to use with complex software controlling functions. Moreover, such class of systems as *IRCS* is commonly assumed, that they are safety-critical systems. The main percentage of hardware failure is due to material aging, short circuits, weather, and climatic conditions, wear while software is not subject to these factors, but surely, may contains latent faults. So we can confidently approve that the software reliability directly impacts on the risk assessment, as shown in the diagram in Fig. 3.

Let's consider more properly the possible mathematical background of a "*Risks from Software Updates*" unit given software reliability theory. The importance of software reliability was clearly defined in early 1972nd [10], and then the extensive research studies on software reliability have been performed, for example [11]. Almost more than 200 software reliability models were proposed by the researchers and software experts, but we can divide them into two broad classes: models based on failure history and models based on data requirements. In another point of view, one can divide software reliability models also into two broad classes: parametric and non-parametric models. It is widely known fact, which parametric models are based on the assumptions of underlying distributions. Non-parametric models do not require any presumptions of parameters to be estimated. It is possible to think

that the non-parametric software reliability models are far beyond the parametric models in application to intelligent software. However, this is not entirely true and depends on a careful definition of upper and lower boundaries of the parameters in the non-parametric models, because the goodness of fit may be sufficient even in incorrect low and high values.

Such sequence performs the parametric software reliability model calculation. In our use case, we considered the following assumptions. Also, software reliability models have been grouped into two classes of models—concave and *S*—shaped. Debugging and other testing actions in real release software production and operation environment are not always performed perfectly. For example, new software errors invalidate the fault-correction activity, or fault-removal is not fulfilled precisely due to incorrect analysis of debug results. Because we deal with the new releases and software updates, we choose the imperfect debugging models among other software reliability models which we describe like that. Moreover, that is important, we must consider the software of *IRSC* not only by itself but as part of the production system.

Observed software failure data to be *S*-shaped. Among *S*-shaped models the *Non-Homogeneous Poisson Process* (*NHPP*) are the most adequate to the real application. *NHPP* is the whole family of stochastic models used to model software reliability growth. Let M be the size of potential software release, $N(t)$ be the number of software updates by the time t, p is a coefficient of faults are experienced in upgrades and q is a number of faults we have been detected. Process N is started with at time $t = 0$ and is described by the following non-linear differential equation.

$$\frac{dN(t)}{dt} = [M - N(t)]\left[p + q\frac{N(t)}{M}\right];\ t \geq 0. \tag{1}$$

The equation number (1) has an exact solution when $M \to \infty$ and we try to find $F(t)$ be the fraction of a number of adoptions of software release, and we note M cannot be infinity. Next we use the essential technique to numerical solving of (1). The $F(t)$ is an *S*-shaped function that has on an inflection point.

$$\frac{dF(t)}{dt} = (1 - F(t))(p + qF(t);\ t) \geq 0,\ F(0) = 0,\ t \geq 0,\ p > 0,\ q \geq 0. \tag{2}$$

From (2) it following

$$\frac{1}{F'(t)} = \frac{1}{1 - F(1)}\frac{1}{p + qF(t)} = \frac{1}{p + q}\frac{1}{1 - F(t)} + \frac{q}{p + q}\frac{1}{p + qF(t)}$$
$$\Leftrightarrow (p+q)t + const1 = -\ln(1 - F) + \ln\left(\frac{p}{q} + F\right) \Leftrightarrow \frac{1 - F(t)}{\frac{p}{q} + F(t)} = const2e^{-(p+q)t},$$

were $const2 = \left(\frac{q}{p}\right)$ with $F(0) = 0$.

$$1 - F(t) = e^{-(p+q)t} + \frac{q}{p}e^{-(p+q)t}F(t) \Leftrightarrow 1 - e^{-(p+q)t} = F(t)\left(1 + \frac{q}{p}e^{-(p+q)t}\right).$$

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \left(\frac{q}{p}\right)e^{-(p+q)t}}. \tag{3}$$

The inflection point of (3) can be found computing its second derivative. We start with (1) differential equation.

$$\frac{dF(t)}{dt} = (1 - F(t))(p + qF(t)); \quad t \geq 0, F(0) = 0$$

$$\Leftrightarrow \frac{1}{F'(t)} = \frac{1}{1 - F(t)}\frac{1}{p + qF(t)} = \frac{1}{p+q}\frac{1}{1 - F(t)} + \frac{q}{p+q}\frac{1}{p + qF(t)}$$

$$\Leftrightarrow \frac{-F''}{(F')^2}(p+q) = \frac{F'}{(1-F)^2} - \frac{F'}{\left(\frac{q}{p} + F\right)^2}.$$

If we equate the second derivative to zero, one can find the location of the inflection point.

$$F''(t_*) = 0 \Rightarrow (1 - F) = \left(\frac{p}{q} + F\right) \Rightarrow F(t_*) = \frac{1}{2}\left(1 - \frac{p}{q}\right) \Rightarrow t_*$$

$$= \frac{-\ln\left(\frac{p+q}{4q}\right)}{p+q}.$$

It is hard to direct calculate of Eq. (3). However, we receive this solution as a test case. Now we switch to the simplified procedure, when p and q the parameters of our interest be the same interpretation, parameter M be non-random and non-infinite, but fixed (expected failure number), n denote failure number has been found, and λ denotes the intensity of the failures appearance.

The finite-state software reliability model with parameters above is

$$\lambda_n = (M - n)\left(p + q\frac{n}{M-1}\right); n = 0, \ldots, (M-1).$$

With initial conditions and normalization, we solved equations below.

$$p_0' = -\lambda \quad p_0' = -\lambda_0 p_0;$$
$$p_n' = \lambda_{n-1}p_{p-1} - \lambda_n p_n; n = 1, \ldots, (M-1);$$
$$p_M' = \lambda_{M-1}p_{M-1};$$
$$\sum_{n=0}^{M} p_n(t) = 1; p_0(0) = 1; p_n(0) = 0; n = 1, \ldots, M.$$

Thus, we can find the average number of detected faults by time t:

$$F_M(t) = \frac{1}{M}\sum_{n=0}^{M} np_n(t) = \sum_{n=0}^{M} p_n(t)\frac{n}{M};$$
$$F_M(t) = \frac{1}{M}\sum_{n=0}^{M} np_n(t) = \frac{1}{M}\sum_{n=0}^{M}\sum_{k=n+1}^{n} p_k(t) = \frac{1}{M}\sum_{n=0}^{M}\left(1 - \sum_{k=0}^{n} p_k(t)\right) \quad (4)$$
$$\Leftrightarrow F_M(t) = \frac{M+1}{M} - \frac{1}{M}\sum_{n=0}^{M}\sum_{k=0}^{n} p_k(t).$$

Also, one can differentiate (4) on t on both sides

$$\frac{dF_M(t)}{dt} = \sum_{n=0}^{M} p_n(t)\frac{\lambda_n}{M}.$$

Moreover, finally, we received this equation, such as done of [12]

$$\frac{dF_M(t)}{dt} = (1 - F_M(t))(p + qF_M(t)) - \frac{q}{M-1}\left(F_M(t)(1 - F_M(t)) + M\mathrm{Var}\left(\frac{n}{M}\right)\right). \quad (5)$$

Equation (5) contains variance that can be calculated using numerical methods, but in our case, it would be rather difficult.

To use advantages of parametric and non-parametric software reliability models, risks management and real production approach we below propose a combined approach with analytical model and artificial neural network. We construct a combined calculation model contains an artificial neural network having single input neuron in the layer of entry. We put single output neuron in the output layer and two or more neurons in the hidden layer. In the hidden layer we determine the number of neurons by the number of the base model selected to construct the neural network combination model.

It means that next update of software release may cause or may not cause software faults. In our model we assume that the number of the neurons in the hidden layer is equivalent to the number of software updates, e.g. if we have two updates, then hidden layer consists of two neurons, if we have three updates, then hidden layer consists of three neurons and so on. In Fig. 4 such network is presented.

Regarding the mathematical representation the proposed in Fig. 4 artificial neural network produces the output as conventional feed forward neural network. Of course, we must keep the several steps are required to construct the artificial neural network properly: analyze characteristics, build the proper activation functions, train the network by data sets, and an estimate of parameters.

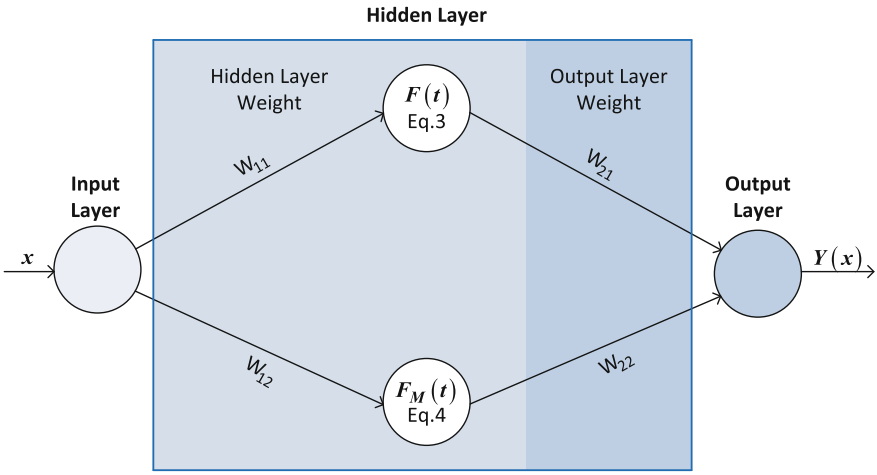


Fig. 4 Combined artificial neural network model

Table 1 Comparison of models

	Generalized NHPP			Goel-Okumoto			Proposed model		
	R^2	RMSE	AIC	R^2	RMSE	AIC	R^2	RMSE	AIC
Data set 1	0.9902	3.826	18.33	0.9913	2.548	16.61	0.9937	0.2955	5.93
Data set 2	0.9893	5.638	26.62	0.9902	3.602	22.91	0.9702	0.9602	11.08

Proposed combined neural network has the output as follows

$$Y(x) = W_{21}\left(\left(1 - e^{-(p+q)t}\right) \Big/ \left(1 + \frac{q}{p}e^{-(p+q)t}\right)\right) + W_{22}\left(\frac{M+1}{M} - \frac{1}{M}\sum_{n=0}^{M}\sum_{k=0}^{n} p_k(t)\right).$$

We use such performance metrics [11] as *Coefficient of determination* (R^2), *Root Mean Square Error* $(RMSE)$ and *Akaike Information Criterion* (AIC). Table 1 describes the results.

The summary from Table 1 demonstrates that proposed neural networks model to give a better fit that two other models. We analyze the performance of our proposed approach with two generalized software reliability models [11], namely *Generalized NHPP* model and *Goel-Okumoto* model.

4 Conclusions

The main aim of the research offered in the paper is to propose new complex method devoted to the design of safety process management for intelligent transport systems, taking into consideration software risks assessment. We examined in

details the experience of *IRCS* and *URRAN* projects and found that *JSC* "*Russian Railways*" devotes much attention to the development of projects related to risk assessment of railway infrastructure. However, the analysis revealed that insufficient attention is paid to the software safety risks estimated in the program-controlled subsystems in the *IRCS*. Our work fills the gap and proposes new complex method devoted to safety management, which based on process analysis of risks, through the development of a new artificial neural network-based software reliability model. Statistical criterions numerically validate some benefits of our approach.

Acknowledgments This work was supported by the Russian Foundation for Basic Research (Grants: No. 15-01-3067-a and No. 16-07-00888-a).

References

1. Research and Design Institute for Information Technology, Signaling and Telecommunications on Railway Transport. http://www.vniias.ru/en/images/stories/docs/niias_leaflet_en.pdf
2. Gapanovich, V.A., Shubinsky, I.B., Rozenberg, E.N., Zamyshlyaev, A.M.: System of adaptive management of railway transport infrastructure technical maintenance (URRAN Project). In: Reliability: Theory & Applications, №2 (37), vol. 10, pp. 30-41 (2015)
3. Stapelberg, R.F.: Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design. Springer, London (2009)
4. Gapanovich, V.A., Zamyshlyaev, A.M., Shubinsky, I.B.: Some issues of resource and risk management on railway transport based on the condition of operational dependability and safety of facilities and processes (URRAN project). In: Dependability, №1, pp. 2-8 (2011)
5. Gapanovich, V.A.: URRAN system. Universal decision-making support tool. In: Zheleznodorozhny transport, №10, pp. 16-22 (2012)
6. Shubinsky, I.B., Zamyshlyaev, A.M.: Main scientific and practical results of URRAN system development. In: Zheleznodorozhny transport, №10, pp. 23-28 (2012)
7. Yants, V.I., Chernov, A.V., Butakova, M.A., Klimanskaya, E.V.: Multilevel data storage model of fuzzy semi-structured data. In: Soft Computing and Measurements (SCM), 2015 XVIII International Conference, vol. 1, pp. 112–114. IEEE Press, New York (2015)
8. ISO/IEC 27005:2011: Information Technology—Security Techniques—Information Security Risk Management, ISO (2011)
9. Chernov, A.V., Butakova, M.A., Karpenko, E.V.: Security incident detection technique for multilevel intelligent control systems on railway transport in Russia. In: 2015 23rd Telecommunications Forum Telfor (TELFOR), pp. 1–4 (2015)
10. Jelinski, P., Moranda, P.: Software reliability research. In: Statistical Computer Performance Evaluation. Academic Press, New York and London, pp. 465–464 (1972)
11. Lyu, M.R.: Handbook of Software Reliability Engineering. IEEE Computer Society Press and McGraw-Hill Book Company (1996)
12. Niu, S.-C.: A Piecewise-Diffusion model of new product demands. Oper. Res. **54**(4), 678–695 (2006)

An Approach to Interactive Information Processing for Situation Awareness About Incidents in Railway Infrastructure Management System

Vladimir D. Vereskun, Maria A. Butakova and Olga V. Ivanchenko

Abstract The paper devoted to designing an approach to maintaining a situation awareness about incidents in *Railway Infrastructure Management System* (*RIMS*), which is operated by *JSC* "*Russian Railways*". In the paper, we provide an analysis of information model of *RIMS* and define the flowchart of interactive information processing for situation awareness about incidents. A novel approach to the strategy of information granulation about incidents and design of the human-computer interface for interactive information processing within *RIMS* is developed. The paper contains an implemented human-computer interface for mobile devices connected to *RIMS* and statistics data examples about incidents and situation awareness of staff.

Keywords Interactive information processing · Situation awareness · Incident awareness · Human-computer interface

1 Introduction

The term "Situation Awareness" is now known long enough, but because of the appearance of information technologies, it began to sound different. The concept means the possibility of obtaining a complete and accurate set of required for the decision information about the situation in real time, including its nature and characteristics. The need for full-ownership situation exists in a variety of areas where there is a large amount of information flow and a high degree of risk, i.e. where a wrong decision can have serious consequences. Scientific and systematic

V.D. Vereskun · M.A. Butakova (✉) · O.V. Ivanchenko
Rostov State Transport University, Rostov-on-Don, Russia
e-mail: butakova@rgups.ru

V.D. Vereskun
e-mail: vvd@rgusp.ru

O.V. Ivanchenko
e-mail: olka-pozitiff@mail.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_31

study of situational awareness began with the work [1], which systematically defined an integrated definition of the highlighted concept. Situation awareness is a three-stage process: (1) perception of the elements in the environment; (2) comprehension of the current situation; (3) projection of future status. In next paper [2], Endsley explored the questions concerning various situation awareness measures and has revealed some principles of *SAGAT* (*Situation awareness global assessment technique*). In general, we can say that the papers above gave rise to quite a fruitful bright area with the advent of scientific monographs [3, 4].

Situation awareness occupies a tactical level, but the next term "Incident awareness" is applied to the strategic level of decision support. The incident analysis is quite a broad concept and applies to all kinds of circumstances associated with the actual incident. It is used as a means of incident mitigation measures and management. Events and incidents in this context are hypothetical, at the level of probabilities. Scenarios are developed as a rule when the situation of the incident outcome has not happened yet. Scenarios for adaptation and response are planned and implemented by the probability of their occurrence and negative impact.

Interactive computations [5] involve principles of designing a human-machine interface, especially for intelligent systems. Such systems operate with a considerable amount of data from various sensors, programmable agents, networked devices and the quality and quantity of information significantly affect a staff awareness in decision-making. A large amount of data about incidents does not mean that the user has more incident awareness. On the contrary, the user may be misled about the actual situation. Proper human-machine interface must present information in such a form as to ensure the fulfillment of the desired goals and objectives needed to the user. The registering and management functions about rail incidents are performed by the unified corporate automated infrastructure management system of *JSC* "*Russian Railways*", or briefly, the *Railway Infrastructure Management System* (*RIMS*). *RIMS* is a tool to address management problems and information support business processes of the current operational infrastructure content *JSC* "*Russian Railways*". Nowadays, *JSC* "*Russian Railways*" is continuously implementing of the development of multilevel intelligent control systems on transport, which realize various new classes of control algorithms and technologies. *RIMS* is a core part of any subsystem within intelligent control systems on transport, which performs a data storage about numerous railway infrastructure objects, planning repair jobs, violation safety incidents, and other incidents [6]. The rest of the current paper is organized as follows. Section 2 contains related work and analysis of information model of *RIMS*. Then we define an interactive incident processing cycle. Section 3 proposes an approach to the design of the human-computer interface for interactive information processing within *RIMS*. Section 4 present an implemented human-computer interface for mobile devices connected to *RIMS* and some statistics data about incidents and situation awareness.

2 The Information Model and General Incident Processing

The information model of railway infrastructure, which is shown in Fig. 1 consists of three parts: (1) *Storage of Infrastructure Objects* (*SIO*); (2) *Incident Management System* (*IMS*); (3) *System of Incident Processing* (*SIP*). *SIO* is the first and main component, which is designed to store the characteristics of infrastructure objects and relationships with their classification by purposes and functions. *SIO* contains a detailed description and relationships among various rail infrastructure objects, rail track schemes, their characteristics, and parameters, sequences of states of infrastructure objects, as well as their dependencies at logical and physical levels.

The second component of the information model is *IMS*, which is aimed at automation and management of technical failures and other incidents from various rail process intelligent and control systems. *IMS* performs such tasks as information support for incidents monitoring and accounting, formulation of the incidents specifications, the supply of the information about the ways of troubles resolving and timing control of removing the negative effects of incidents. The third component of the information model is an *SIP*, that is intended for automation of control processes of incidents monitoring and failures and troubles eliminations at the whole of the railway infrastructure objects.

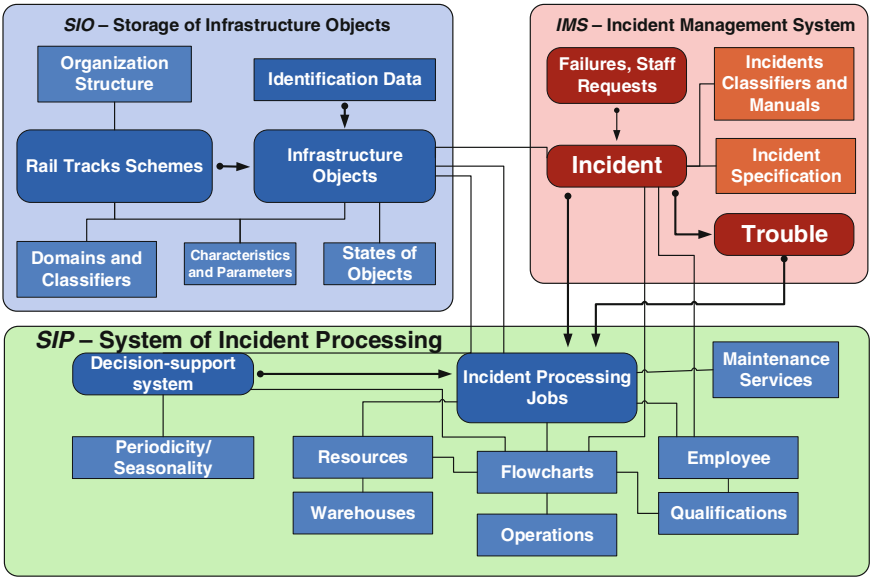


Fig. 1 The information model of *RIMS*

These three components are viewed as unified platform for a specialized database of rail infrastructure objects. The process of incidents control is directly related to the process of monitoring and diagnosis. The incidents processed can have a different base. They may be based on the data from automatic monitoring and diagnosis system, and also on the data entered manually. The incidents as themselves may be regarded as important and requiring action to restore and requiring actions for adjusting maintenance plans. The *RIMS* as a whole is as a business-critical system. *RIMS* operation does not affect the operation of other systems. The above system operates 24 h of the day, 7 days of the week and 365/366 days of the year without a time of maintenance.

We describe the *RIMS* as a system for a situation awareness and below in Fig. 2 we reviewed and designed a general flowchart of incidents processing cycle within *RIMS*.

Several rail departments involved into interactive information processing cycle about incidents, namely: railway track and engineering and civil engineering buildings; electrification and electricity supply department; automatics and telemechanic department, and, finally, the rolling stock department. They perform the following functions, concerning the registering and monitoring incidents: the forming and maintaining the unified information model of the exploited rail infrastructure; the forming of the unified description of objects and the set of relations among them; the forming and the managing of unified normative and corporate help documentation, including geolocation data; provision of the unified user interface to processed datasets.

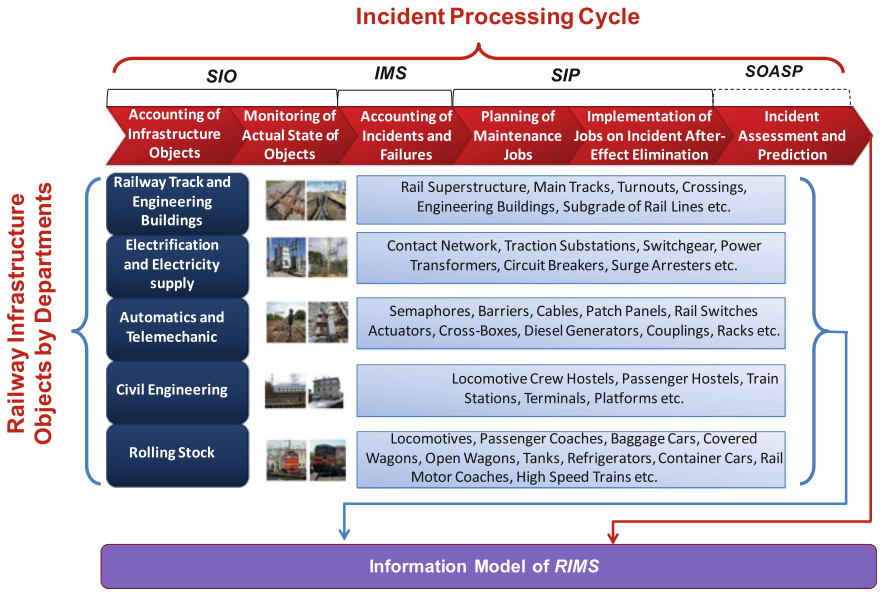


Fig. 2 A general flowchart of incidents processing cycle

The *RIMS* has a powerful hardware to maintain the incidents processing cycle above. The hardware consists of

- two database servers, *IBM zSeries* with parameters adopted by *JSC* «*Russian Railways*» and connected in the fault-tolerant cluster;
- two application servers (each of them contains a dual *Intel Xeon* 3 *GHz* processor, 16 *Gb RAM*, 4*146 Gb *Ultra-fast RAID*, 2*1*Gbit/s* network interfaces), which is forming a network load balancing cluster;
- one report server and one web server with parameters the same as above, connected to the high-speed local area network;
- over 12,000 workstations of various employees of *JSC* «*Russian Railways*».

The software of *RIMS* has been designed using the *IBM Maximo* platform with *Oracle Database Server*, *IBM WebSphere* as an application server and web portal. The workstation of *RIMS* is a thin client with only requirements a *Java Runtime Environment* and web browser.

3 Proposed Approach

In the previous section, we briefly outlined a technical architecture of *RIMS*, its hardware and software platform and general incident processing cycle. Next we pay more detailed attention to principles of interactive human-computer processing of incidents. Our approach will be using the human-centric granular computing [7, 8] principles. The proposed approach involve several strategic principles of information granulation for situation awareness as following.

1. The source of incident granulation:
   (a) The granule of events category "*Complex*", registered in the objects by Departments;
   (b) The granule of events derived from automation monitoring systems by Departments, the category "*Technical*";
   (c) The granule of incidents initiated by various employees, the category "*Request*".
2. The incident registration granulation:
   (a) The granule of registration unification—the same treatment is related to only one incident;
   (b) The granule of registration singularity—a single incident cannot be linked more than one source of incident granulation;

3. The incident processing granulation:
   (a) The granule of processing execution—a uniquely *Responsible Officer* for "*Technical*" incident;
   (b) The granule of processing control—a uniquely *Senior Officer* for "*Request*" incident;
   (c) The granule of processing workgroups—a *Head of the Workgroup* is responsible for "*Complex*" infrastructure incidents.
4. The efficiency metrics granulation:
   (a) The granule of infrastructure incidents solution rapidity;
   (b) The granule of user requests incidents solution rapidity;
   (c) The granule of information awareness about incidents grouped;
   (d) The granule of mean time to maintenance jobs about incidents grouped.

The main flowchart of proposed approach to interactive information processing about incidents in Rummler-Brache notation [9] is shown in Fig. 3. The

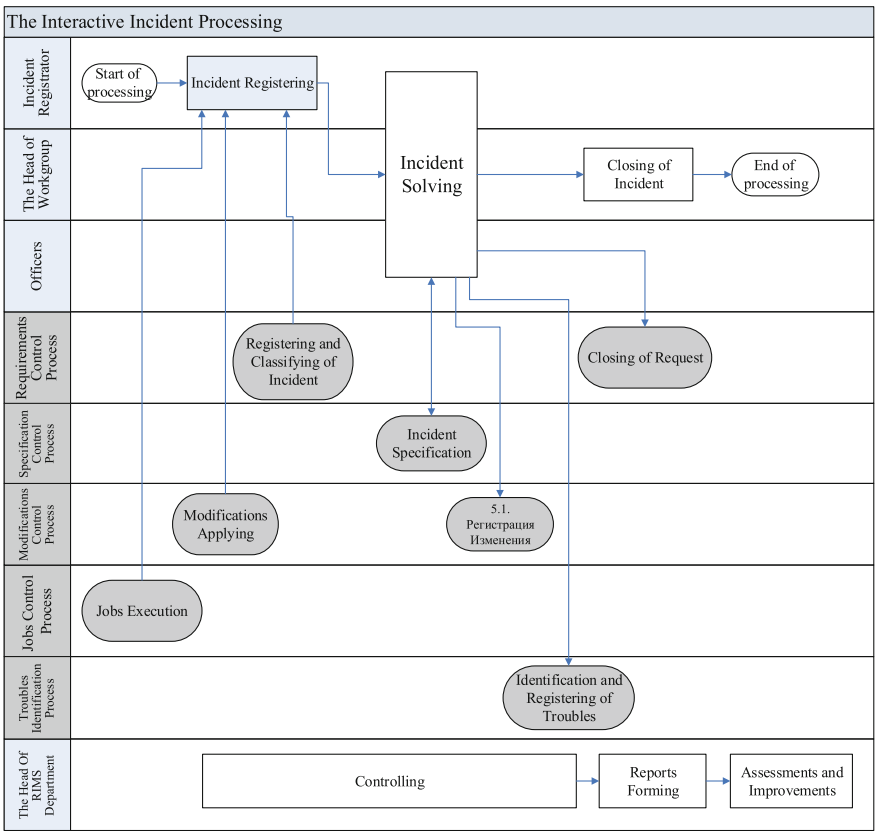


Fig. 3 A general flowchart of incidents processing cycle

methodology used to detail the business processes and is a modification of the widely known *WFD* and *DFD* standards. The diagram of the business process is divided into tracks. Each track corresponds to the structural unit or positions involved in the business process. Both operations are performed a structural subdivision, located in its area of the track. Rummler and Brache proposed [9] to divide the chart on the track on the following principles: the first track—to display the cooperation of the customer and process; the latter path—to display support external processes; the central walkway—relevant structural units or positions involved in the process.

Let us detail some math background of proposed approach. As we mentioned above, our approach is based on granular computing principles. We consider granulation problem as topology problem what distinguishes our approach from the others.

Let us expound the topological approach related to the notion of *similarity*, which is one of the basic when considering the information granules. The approach is applicable in situations that were expressed in the discrete nature (e.g., when the many universal courses and admitted). Let U—the arbitrary universal set and $k, n \in N$.

Also, we define in U the fuzzy set $C_{\mu}^{(n)}$, which has a type n, $C_{\mu}^{(n)} = \{(u, \mu(u)) : u \in U\}$, $n \in N$, where $\mu(u)$—is arbitrary mapping the U into the set of fuzzy sets U of type $n - 1$.

Definition 1 Let's define that the fuzzy set $C_{\mu_1}^{(n)}$ is *k-similar* to a fuzzy set $C_{\mu_2}^{(n)}$ if in cuts of these sets coincide, at least, k elements.

The relation k-similarities is the *tolerance* relation on a set $[0, 1]_n^U$. It is easy to be convinced that so that the relation to be k-similar was the equivalence relation on $[0, 1]_n^U$, is necessary also enough that the set U was final and the number of its elements equaled k. Let's note that situations when tolerance appears equivalence, there is no need to exclude from consideration though in such cases, and there is a weakening of "illegibility" of sets. At further specification of the thoughtful questions, the relation of equivalence can also be demanded.

Definition 2 Let's define that the fuzzy set $C_{\mu_1}^{(n)}$ is *k-similar* to a fuzzy set $C_{\mu_2}^{(n)}$ if cuts of these sets differ no more than $k - 1$ elements.

The relation *k-similarity* also is the *tolerance* relation on a set $[0, 1]_n^U$. At $k = 1$ similarity of fuzzy sets means coincidence of their cuts. For final universal sets of the concept of similarity and similarity of fuzzy sets coincide, if N—the number of all elements U, then so that sets $C_{\mu_1}^{(n)}$ and $C_{\mu_2}^{(n)}$ (at any fixed n) were *k-similar*, is necessary and enough that they were $(N - k)$-*similar*. If U it is infinite, then two k_1-*similar* to some k_1 sets $C_{\mu_1}^{(n)}$ also $C_{\mu_2}^{(n)}$ can not be k_2-*similar* at any k_2. At the same time, two k_2-*similar* to some k_2 sets are k_1-similar at any k_1. Thus, the relation of similarity is, generally speaking, narrowing of the relation of similarity.

Traditional creation of the theory of fuzzy sets uses the set-theoretic algebraic structure of a lattice. It is natural to attract to cases in point and the topological structures in the most general view expressing the concept of proximity.

Definition 3 Let's define that fuzzy sets $C_{\mu_1}^{(n)}$ and $C_{\mu_2}^{(n)}$ are *ε-proximity* $(\varepsilon > 0)$ if the distance between cuts of these sets in the considered metric functional space does not surpass number ε.

The relation *ε-proximity* also is the *tolerance* relation on a set $[0,1]_n^U$. For example, addressing one of the standard metrics, it is possible to say that fuzzy sets $C_{\mu_1}^{(n)}$ and $C_{\mu_2}^{(n)}$ are *ε-proximity* if in cuts of these sets the absolute value of a difference of degrees of the accessory of each element $u \in U$ does not surpass ε. Such form of proximity belongs to the topology of uniform convergence.

The relations entered above are *k-similarities*, *k-* and *ε-proximity* we will unite *similarity* ε one term—*a resemblance*.

Definition 4 *An information granule* in $[0,1]_n^U$ is called its subset Γ such that any two elements belonging to it are constitutes a *resemblance*.

Each subset of a granule consisting, at least, of two elements is also a granule.

Definition 5 The granule in $[0,1]_n^U$ is called *finished* if it does not contain in any other granule.

Thus, the finished granules are the maximum elements in the partially ordered set (rather set-theoretic inclusion) of all granules in $[0,1]_n^U$, that is that and only those subsets $[0,1]_n^U$ which coincide with the crossing of classes of tolerance of all elements belonging to them.

The set of all fuzzy sets in any three-element set U is geometrically identified with a single cube of space R^3. For the relation of similarity the class of tolerance of any fuzzy set representing some point of a cube is an association of three pieces which are lying in this cube, passing through this point and such that one of them is parallel to abscissa axis, another—ordinate axes, the third—axes of z-coordinates.

4 Some Practical Results

In this section, we present some practical results concerned to the human-computer interface of mobile devices connected to *RIMS*. To show the scale of the control task incidents on the railways in Fig. 4, we present statistics on reported incidents in recent months in 2015. Next, In Fig. 5 a couple of sample screenshots of the interactive incident processing on a mobile device are presented.

The mobile interface also consists much-supporting functions for operational staff, such as information about incident location tracking, detailed incident specification, incident processing status, etc.

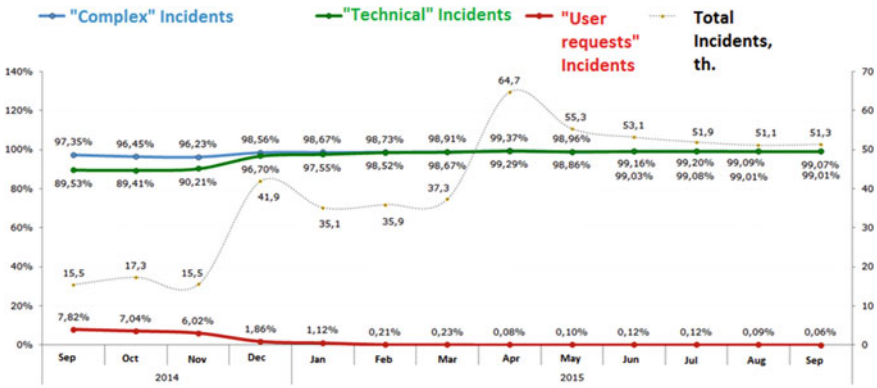


Fig. 4 The types of *RIMS* incidents registered in 2015

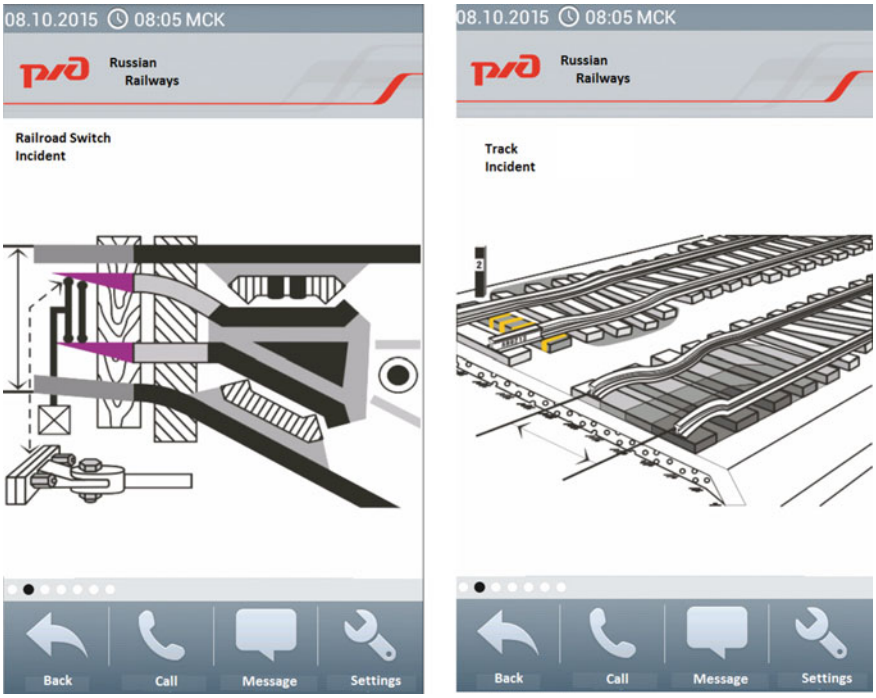


Fig. 5 A human-computer mobile interface to interactive incident processing

5 Conclusion

Finally, we conclude the paper by describing some advantages of our approach. The first advantage consists the fact that proposed approach is consistent with the general structure of new classes of modern intelligent transportation systems, such as *RIMS*, which is exploited by the *JSC* "*Russian Railways*" nowadays. The second advantage is that we use a strategy and principles of information granulation. This strategy allows us to maintain a situation awareness about a huge number of incidents, of the order 70 thousand various incidents in a month. Moreover, finally, as the mathematical background of our approach, we applied original topology definitions of fuzzy granules similarity, proximity and as a whole concept is a tolerance and resemblance relations of information granulation.

Acknowledgments This work was supported by the Russian Foundation for Basic Research (Grants: No. 15-01-3067-a, No. 16-01-00597-a, No. 16-07-00888-a).

References

1. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **37** (1), 32–64 (1995)
2. Endsley, M.R.: Measurement of situation awareness in dynamic systems **37**(1), 65–84 (1995)
3. Endsley, M.R., Bolte, B., Jones, D.G.: Designing for Situation Awareness: An Approach to Human-Centered Design. Taylor & Francis, London (2003)
4. Endsley, M.R., Garland, D.G. (eds.): Situation awareness analysis and measurement. CRC Press, Atlanta, GA (2001)
5. Goldin, D., Smolka, S., Wegner, P. (eds.): Interactive Computation: The New Paradigm. Springer, Berlin (2006)
6. Chernov, A.V., Butakova, M.A., Karpenko, E.V.: Security incident detection technique for multilevel intelligent control systems on railway transport in Russia. In: 2015 23rd Telecommunications Forum Telfor (TELFOR), pp. 1–4 (2015)
7. Yao, Y.: Artificial intelligence perspectives on granular computing. In: Pedrycz, W., Chen, S.-M. (eds.) Granular Computing and Intelligent Systems: Design with Information Granules of Higher Order and Higher Type. Springer, Berlin, pp. 17–34 (2011)
8. Bargiela, A., Pedrycz, W.: Toward a theory of granular computing for human-centered information processing. IEEE Trans. Fuzzy Syst. **16**(2), 320–330 (2008)
9. Rummler, G.A., Brache, A.P.: Improving Performance. How to Manage the White Space on the Organization Chart, 3rd edn. Jossey-Bass a Wiley Imprint, Wiley (2013)

Stochastic Traffic Models for the Adaptive Train Dispatching

Vladimir Chebotarev, Boris Davydov and Aleksandr Godyaev

Abstract Adaptive dispatching allows you to adjust the movement of vehicles which improves transport service and reduces costs. Optimal dispatching decisions are taken on the basis of modeling the disturbed traffic. The paper presents the foundations of the stochastic model creation and the methodology of its use in online traffic rescheduling. The authors investigate the base principles of the train traffic stochastic models which are used for the effective dispatching. The paper propose the practical method of macro regulation the train flow which prevents local conflicts between services based on the probabilistic forecasting the headway. Computational results based on experimental data confirm effectiveness of the stochastic model when predicting and solving conflicts in the mixed traffic on the main line.

Keywords Train traffic · Stochastic model · Operative management · Rescheduling

1 Introduction

Stochastic modeling of train traffic is widely used for the creation of effective annual (normative) timetable or when searching for ways to develop the rail infrastructure. Modeling allows you to obtain a stable schedule of through proper allocation of trains on the network and by the optimal distribution of time surpluses. Predictive model of the traffic in a large railway network has an approximate character, since used much averaged baseline conditions of functioning. There are often situations in the real train traffic that differ greatly from the averaged option.

V. Chebotarev
Computing Centre, Far Eastern Branch, RAS, Khabarovsk 680021, Russia
e-mail: chebotarev@as.khb.ru

B. Davydov (✉) · A. Godyaev
Far Eastern State Transport University, Khabarovsk 680021, Russia
e-mail: dbi@rambler.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_32

This is due to the presence of many independent influencing factors. The conflicts between trains are local in character and tend to propagate in a confined time-distance space. Conflict resolution is carried out by online regulation that provides the train manager. You must have an adaptive control system to implement a qualitative adjustment of movement in disturbed conditions. This system uses a large volume of the actual information and makes a decision on conflict prevention.

Rescheduling when used the adaptive control begins at a moment when happens one of the events, such as coming the signal of the train trajectory deviation. In this case, the problem involves the detection of first the nearest conflict (in time) and then, in determining the chain of secondary conflicts. The train traffic is modeled using probabilistic directed graph when there are the random impacts. Such a representation of the process leads to the unification of operations that used for cumulative delay formation on the basis of partial distributions the activity intervals. The research has shown the convolution function of the initial distribution is the basis of any particular act of delay formation. This method is used in the prior probabilistic analysis of delay propagation which appears due to disturbed process of connections. In reality, the majority of disturbances occur in the movement of trains at open tracks. Therefore there is a need to create a comprehensive model which adequately reflects all kinds of intertrain conflicts. This problem is considered in Sects. 3 and 4 of this paper. The process of delay formation is represented as a *tree of interferences*, nodes of which are the acts of conflicts between trains at stations or at open tracks. The problem of local modeling the train conflicts is significantly different from the creation of traffic model on a largely network. The main features are a small number of trains that are experience troubles and a presence of actual implementation the influencing factors.

Adaptive train traffic control involves the use of two fundamentally different strategies. The first type of strategy used in the strict traffic regulation, which is necessary for transit the passenger trains and the freight trains with rigid timetable. The second kind of management is advisable to apply to freight traffic during periods of weakening the intensity of the train flow. This strategy consists in operative change the parameters of movement the train groups which carry out the train manager. The manager can enlarge the departure interval of freight trains. Time margins improve the local situation, i.e. reduce the number of nonscheduled stops and their total duration. The problem of modeling the economical mode of freight train traffic and determining its temporal place in the schedule is investigated in Sects. 5 and 6.

2 Literature Review

The majority of published papers are considering the rescheduling as a deterministic problem of train traffic on the congested network. The paper [1] provides an extensive overview of recovery models and algorithms for real-time rescheduling.

The main types of problems which are solved in searching for optimal adjustments the disturbed flow are conflict detection and resolution (CDR) and train speed coordination (TSC) [2]. In these problems, the best order of train movement determined to reduce the deviations. Problem of finding the optimal schedule is very time-consuming. Most frequently, the branch and bound algorithm is used as the optimization framework to trains' rescheduling. To speed up the solution often apply different heuristics and greedy algorithms. One of the most efficient algorithms described in [3].

The well-known works employ microscopic models that describe the process of moving in detail, but more time is needed to determine the optimal solution. Many authors have used more productively mesa-model [4], which deals with the running operations on relatively long sections and with the stops at stations. It allows you to develop a schedule for a large rail network in a shorter time.

The main disadvantage of deterministic models used for rational dispatching, is insufficient consideration the risk of accidental disturbances. This drawback is largely compensated by using the stochastic models. In one of the first papers considering this problem [5] proposed analytical method for delay calculation by use the convolution of distributions the input and current disturbances. The statistical model of train traffic on main line created in [6]. The approach is based on regression models, built using training sequences for the main indicators such as speed, duration of train processing etc. The model of formation the delays in the train packet using a probabilistic approach proposed by Carey et al. [7], determines the total running time as a sum of the partial random intervals at the elements of site.

The most developed stochastic model of propagation the delay is created in the paper [8]. This work examines the influence of the random scattering of running time on the passenger waiting process at the interlinking. The authors use a discrete representation of the probability distribution. Stochastic model which comes really close to our approach is in [9]. The paper uses the mesoscopic model and stochastic activity graph. The main purpose of the research is to determine the right cumulative distribution function which is the result of convolution the initial distributions.

The analysis shows that there are insufficiently studied issues in the problem of stochastic modeling of train traffic. Effective implementation of adaptive traffic management requires development of an overall probabilistic model.

3 The Tree of Interferences as the New Mesa-Model of Delay Propagation

Rescheduling models which describe the passenger connections and the train departure process at the stations are developed in many papers. In these models, the nodes of an activity graph which reflect the process within the station corresponds

to operations of arrival (B_{arr}) and departure (B_{dep}) the trains. The nodes of the graph represent the events of the beginning and end of the short stop, if examined the model of the conflict at the open track. To combine both of these models, we are using a new approach that involves an identical description of the processes the delay accumulation at the stations and the interstations. Nodes of a simplified activity graph correspond to conflicts of trains. The method of constructing a graph based on the mesoscopic model of the traffic process [4]. The train movement which disturbed by random influences, we present as *a tree of interferences*. Such a model is a graph whose nodes represent the competition between the trains for the occupation the element of infrastructure in conflict places (*operating points*-according to the terminology of [8]). Some block-sections or level crossings at the stations and on the open trucks may represent these conflict points. It is usually assumed that conflict situations can be of two types. The first type corresponds to the primary (input) disturbances and other to the secondary delays. Deviations from the schedule, which is directly caused by the difficulties related to only one train, belong to the primary disturbances. In particular, the delay of arrival to the boundary point between the external and internal networks is considered to be the primary one. The main part of primary disturbances which appears inside the network is caused by equipment failures. Impact of the train-leader on the following train is manifested in the conflict points of second type. This situation arises, for example, if speed of the following train exceeds the same for train-leader.

The proposed model differs from the graph used in a number of papers to simulate the situation on the railway section (we specify the paper [2], as the example). The difference lies in the fact that the model includes only the points at which the train scheduled activity is disrupted. There are some conflict situations, or forced (preventive) change of the running mode are present at these points. The last activity performed to eliminate the delay or satisfy requirements for limiting the speed.

Often there is a type of conflict which occurs between a pair of trains at the boundary station due to the late arrival of the leading train (see Fig. 1a) or of both the trains (Fig. 1b). The primary delay can be spread over the next trains (Fig. 1c). Thickened lines (that is, the arcs of the graph) reflect the running or the dwelling activity, which are implemented with the delays regarding the timetable. The length of each arc corresponds to the duration of the operation (or the group of operations). This value is called *an operating time*. Train 2 gets additional delay after the act of interaction with the train 1. This causes a delay of scheduled train 3 (Fig. 1c).

Figure 2 shows the components of the tree which correspond to the appearance of secondary delays due to the speed difference of trains running. Figure 2a illustrates the delay propagation on the trains 2 and 3 due to unscheduled stop of the train 1 at a site or at a station (at the point s_1). Conflict situations when rapidly moving train catches up a slower one at a certain point, depicted on the fragment of Fig. 2b.

The tree of interferences also includes components that reflect the forced change a mode of train run due to driver activity or the dispatcher order. In Fig. 2c you can

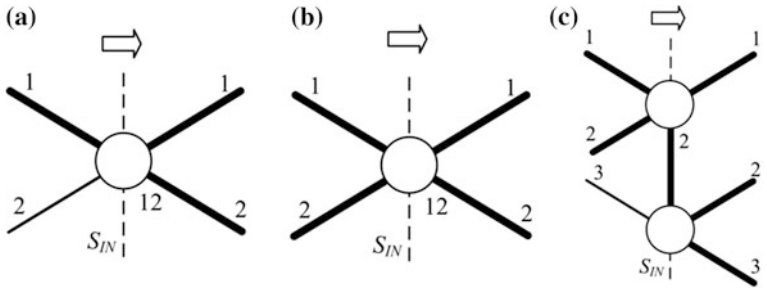


Fig. 1 The element of conflict on the border of the section. The *arrow* indicates the direction of train movement

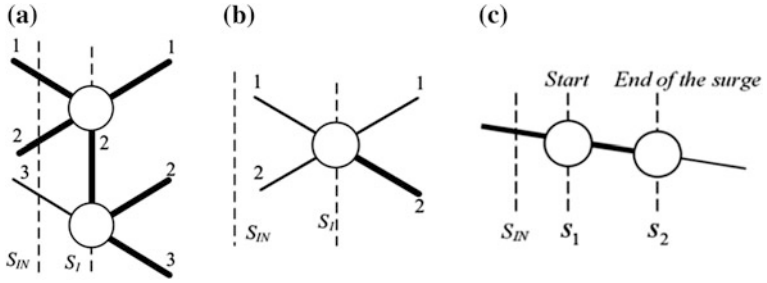


Fig. 2 Elements of passing conflict and of forced change the speed mode

see a model element in which nodes correspond to moments of switching-on (s_1) and switching-off (s_2) the enhanced train speed. The delay disappears at the point s_2 due to local high speed movement of the train.

The example shown in Fig. 3 reflects the delay propagation across the network comprising various types of conflicts. The control site lies to the left of the input station S_{IN} (see Fig. 3a). When the late arrival of the leading train 1 occurs, conflict with the followed train 2 arises. This train acquires a secondary delay. The schedule is dis-rupted causing a new conflict, namely, the simultaneous arrival of the trains 2 and 3 in the intersection of *A*. The train 3 acquires a secondary delay as the result of conflict resolved. This situation causes the propagation of disturbances the transfer of passengers at the point *B* and as a consequence, the occurrence of delay the train 4 (Fig. 3b).

Search for points at which the delays are changed due to conflicts, makes by solving a chain of the two-train problems. This realizes the following algorithm. You have to determine the nearest train (by the time) which may fall under the influence of the disturbance. This operation is performed when a signal about the actual deviation from the schedule comes. This may be a following train which moves after the delayed one, or the potentially conflicted train at the crossing point.

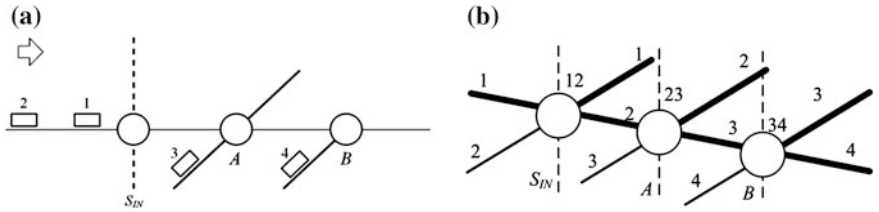


Fig. 3 Fragment of the tree illustrating a propagation of delays at stations and on intersections Trains 3 and 4 are coming from lines which differ from the arrival directions of trains 1 and 2

The number of problems depends on amount the crossings which intersect the path of belated train.

4 The Generic Stochastic Model of a Train Traffic

Here we describe the mechanism of formation the train delays and the headways in the presence of random variations the operation times. Running time on the section is formed by summing the values of duration the running and stopping operations. We assume that the trains are moving in accordance with the schedule in the absence of disturbances. Each of the scheduled operation time includes the minimum allowable time and some supplements.

Influencing factors that are disrupted movement of the train, shifts the actual start and stop times of operations relative to schedule. The train departure from the station (or from the control point) in this case is delayed and scattered within a certain interval Δt^{dep} (see Fig. 4).

The train complies with the scheduled travel time t_S in the absence of random influences at each the elementary sites. In reality, the running time on the section AB is exposed to random influences. These actions cause changes of speed or emergence of short stop what leads to the scattering of arrival time. This time is the sum of random variables, jointly reflecting a scattered time of the train departure and running activities. This is depicted in the graph of Fig. 4 as a set of "packets" the train traces.

The work [9] shown that the accumulation of delays can be described by a probabilistic operational graph. Each node of the graph corresponding to the planned stop. Parameters of the actual state are determined by a set of previous operations. Some of them are the running operations at the segments adjoining the station (arcs t_{runi} in Fig. 5), while others (arcs t_{dwelli}) represent the passenger transfer. Random moment of departure from the station B is determined by set of their arrival times together with the strategy of changing the order of train departure by having the delays.

The mentioned paper shows the convolution of the distributions of two random variables $F_{ac1}(t)$ and $F_{ac2}(t)$ is the basic operation for the analysis of disturbances:

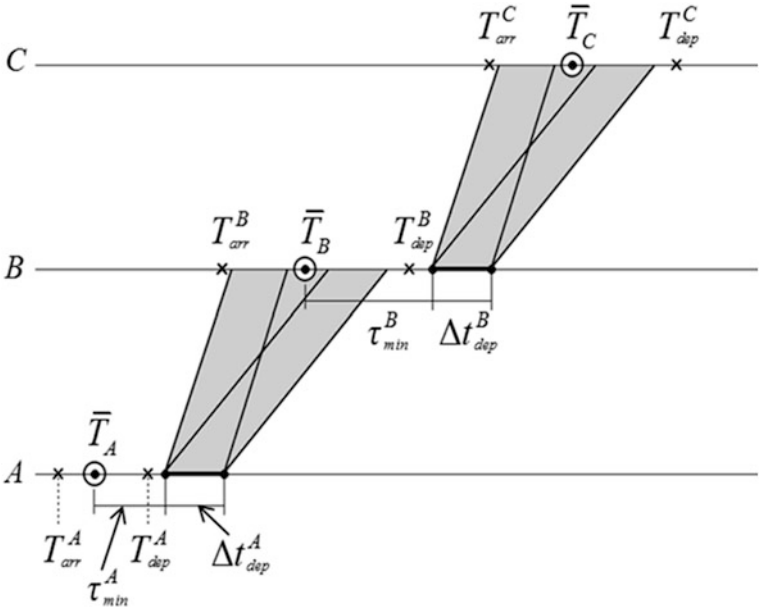


Fig. 4 Graphic interpretation of the stochastic model of a train traffic ×—scheduled times of arrival and departure; ⊙—predicted arrival times

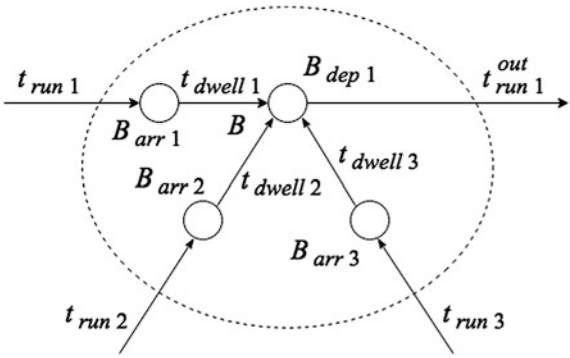


Fig. 5 Fragment of the graph, which reflects the impact of the interlinking

$F_{sum}(t) = F_{ac1}(t) * F_{ac2}(t)$. These variables reflect operating delays at the interconnection. We offer the generic stochastic model in which a convolution operation is considered as a universal (basic) element in the formation of partial delay. A framework of processing the two random variables is independent of the type the train interaction and of the chain of operations. A convolution operation is performed as in the evaluation of the results that interaction, and in determining the delays on the station.

The restrictions which are defined by the regulations of processes management are involved in the forming of partial distributions. There are such a restrictions as

the minimum values of speed and of headways at the open tracks. Dwelling times at the station is also restricted by the limit values. Now we show the scheme for calculating the output probability distribution of moments of departure the train 1 in the situation displayed by the graph fragment on the Fig. 5:

$$F_{out1}(t) = [(F_{run1}(t) * F_{dwell1}(t)) * (F_{run2}(t) * F_{dwell2}(t))] * (F_{run3}(t) * F_{dwell3}(t))$$

The structure of this formula emphasize the fact that the basis for each elementary calculation is two-train model, and composition of distributions is determined regardless of what kinds of operation are processed together.

5 Two Strategies for Conflict Detecting and Resolution

Prevention of conflicts between trains during disturbances is a major problem of operative traffic management. Rescheduling in different flows of trains may have a certain type. The main requirements are to ensure the accuracy of the arrivals and the quality of the transfer process if is regulated the flow of passenger trains. The same requirements apply also to freight trains, are moving on a hard schedule. These conditions assume the use of a control strategy in which the manager detects and resolves every conflict individually. This framework we call *the strategy of local management*.

At the main railway with mixed traffic, there are the periods when are moving large packets of conventional freight trains. Quite often there are periods of 2-3 h, when the intensity of freight flow is low. During these periods there are additional time reserves, which can eliminate many conflicts. An increase the train running time is one of the main methods to reduce the level of delays [10]. The regulatory action is that the dispatcher purposefully assigns the group of trains that travel in economy mode. The combination of economic and intensive modes is illustrated in Fig. 6.

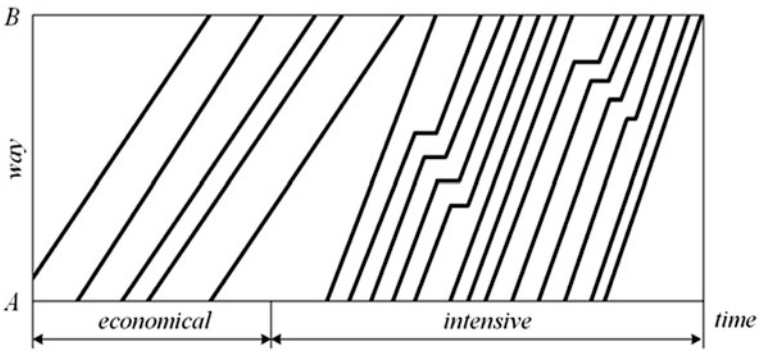


Fig. 6 Fragment of schedule with the economical and intensive train traffic

We are showed the using a stochastic model of train traffic to solve the problem of detection and resolution the set of conflicts. There are use the model the propagation of disturbances at the chain of train that occur after the occurrence of a random primary delay. We consider the methodology for analyzing the features of train flow and rational choice of a traffic mode as *the on-line integral flow adjustment.*

6 The Boundary Interval for Appointment the Economy Mode

We consider the problem of propagation of a nonscheduled small stop having a probability density $g(t)$, when a dense flow of trains is moved. A primary delay can generate a chain of nonscheduled deviations. Because the moments of the train departures usually deviate from scheduled ones, they can be considered as random ones. We assume the departure time intervals between the trains following are independent and have the same density function $\psi(t)$.

We solve the following problem: for arbitrary $\boldsymbol{k} \geq 2$ to find the distribution function $W_k(t) \equiv W_{B,k}(t)$ of the arrival interval between $(k-1)$th and kth trains to the B. Regularities which are described in Sect. 3 are used to create model of delay propagation. We deduce the expression for the output distribution in the following form: $W_k(t) = I(t > t_0) \left\{ \int\limits_{-\infty}^{t-t_0} \psi(z)\, dz + \int\limits_{-\infty}^{\infty} \left[\int\limits_{t-t_0}^{\infty} \left(\int\limits_{z+u-t+t_0}^{\infty} g(x)\, dx \right) \psi(z)\, dz \right] \psi^{*(k-2)} (u)\, du \right\} 2 \leq \boldsymbol{k} \leq \boldsymbol{n}$, where t_0 is the least safe time interval between trains,

$\psi^{*m}(u) = \left(\psi^{*(m-1)} * \psi\right)(u)$, $(f_1 * f_2)(u) = \int\limits_{-\infty}^{\infty} f_1(u-y) f_2(y)\, dy$ is the convolution of two arbitrary densities f_1 and f_2.

Let, as the example, $g(t) \geq I(t > t_0)\lambda e^{-\lambda t}$, $\psi(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-T)^2/2}$,

$$I(x \in A) = \begin{cases} 1 & if \quad x \in A, \\ 0 & if \quad x \notin A, \end{cases}$$

where T is a scheduled departure interval and $\boldsymbol{\sigma}$ is sufficiently small: $4\sigma + \lambda\sigma^2 < T$

Graphs of the functions $W_{\boldsymbol{k}}(t)$ (Fig. 7), calculated taking the real-type distributions mentioned into account. The curves show that with increasing the number of delayed train the form of distribution approaches to a single jump. This demonstrates the decreasing trend of secondary delay up to its complete disappearance.

Now we consider an example that defines the critical headway to assign the economical mode of freight flow. Express headway T_{dep} as the sum of the unknown parameter T and a fixed amount of the minimum safe interval t_0: $T_{dep} = T + t_0$.

Consider the following scenario as an example for estimate the duration of critical departure interval T_{dep} at the station A. It is expected an increase in the flow

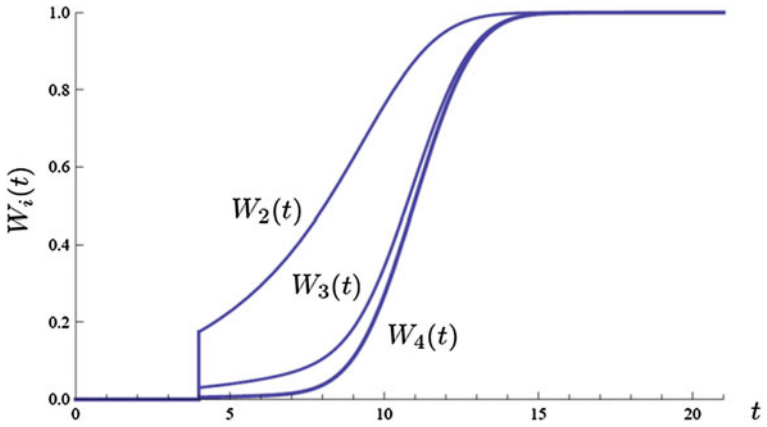


Fig. 7 Graph of distribution the output headway for different numbers of delayed trains

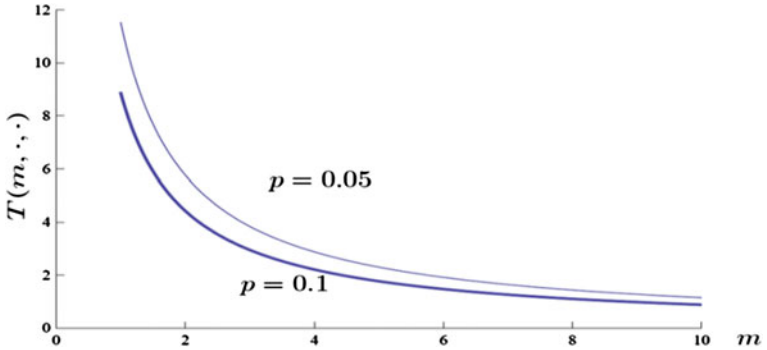


Fig. 8 Dependence of headway on the number of expected unscheduled delays

density of freight trains on the railway section. Manager must appoint such an interval, that the amount of delays m_k is not exceed two units with a probability p less than 0.1. We show an additional interval T in the case when $T_j = \text{const}$ and the exponential behavior of $g(t)$ must satisfy the following condition $T \geq (m\lambda)^{-1}\ln(p)^{-1}$. Graph of the function $T(m_k)$ where m_k is the number of permissible secondary de lays are shown in Fig. 8. Dependences obtained for the real parameters which are valid for the main railway directions of Russian railways.

The calculation shows the time margin (to the value of safe interval t_0) is equal to 4.5 min for $m = 2$. Total duration of the minimum interval between trains is 9.5 min. It should be noted that the headway of heavy freight trains at the Russian railways lies in the range from 10 to 14 min. Obviously, this decision is due to the need to obtain a small amount of unplanned delays. Indeed, when m_k is equal to 1, the interval which is calculated by our technique, represents 13 min.

7 Conclusions and Future Work

The described approach to stochastic modeling of trains allows you to create algorithms for online management. The most important feature of these algorithms is the adaptive adjustment of parameters of the model, taking into account the actual situation. This approach is used to solve practical problems for intensity regulation the freight trains flow. The calculated critical headway values are correspond well to the actual ones. In a further study is necessary to develop an algorithm for the forming the statistical feature of the main model elements such as running and dwelling times.

References

1. Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J.: Overview of recovery model sand algorithms for real-time railway rescheduling. Transp. Res. Part B **63**, 15–37 (2014)
2. D'Ariano, A., Pranzo, M., Hansen, I.A.: Conflict resolution and train speed coordination for solving real-time timetable perturbations. IEEE Trans. Intell. Transp. Sys. **8**(2), 208–222 (2007)
3. Törnquist Krasemann, J.: Design of an effective algorithm for fast response to the rescheduling of railway traffic during disturbances. Transp. Res. Part C **20**, 62–78 (2012)
4. Hansen, I., Pachl, J.: Conclusions. In: Railway Timetable and Traffic. Analysis, Modelling, Simulation, Hamburg: Eurailpress, pp. 209–211 (2008). ISBN: 978-3-7771-0371-6
5. Muhlhans, E.: Berechnung der Verspatungsentwicklung bei Zugfahrten. Eisenbahntechnik Rundschau **39**(7/8), 465–468 (1990)
6. Shapkin, I.N., Usipov, R.A., Kozhanov, E.M.: Modelling of train operation on the basis of multi-factor valuation the process operations. VestnikVNIIZhT. **4**, 30–36. Шапкин, И.Н., Юсипов, Р.А., Кожанов, Е.М.: Моделирование поездной работы на основе многофакторного нормирования технологических операций. Вестник ВНИИЖТ, **4**, 30–36 (2006)
7. Carey, M., Kwiecinski, A.: Stochastic approximation to the effects of headways on knock-on delays of trains. Transp. Res. Part B **28**(4), 251–267 (1994)
8. Berger, A., Gebhardt, A., Müller-Hannemann, M., Ostrowski, M.: Stochastic delay prediction in large train networks. In: Caprara, A., Kontogiannis, S. (eds.) 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems. ATMOS 2011, pp. 100–111 (2011)
9. Bueker, T., Seybold, B.: Stochastic modelling of delay propagation in large networks. J. Rail Transp. Plann. Manag. **2**(12), 34–50 (2012)
10. Davydov, B.I., Chebotarev, V.I.: Optimal modes of the freight train traffic. Transp. Sci. Technol. Manag. **1**, 65–67. Давыдов, Б.И., Чеботарев, В.И.: Оптимальные режимы движения потока грузовых поездов. Транспорт: наука, техника, управление, **1**, 65–67 (2015)

Usage of Digital Image Processing Methods in the Problem of Determining the Length of the Rail Joints

Anatoly Korobeynikov, Vera Tkalich, Sergey Aleksanin and Vladimir Polyakov

Abstract Methods and techniques of digital image processing are commonly used for various problems of defectoscopy. Usage of digital image processing methods for automated determination of length of the rail joint is described. There are many of such methods. Among them are filtering methods, image enhancements methods, deblurring methods, methods or regulation, morphological filtering methods, edge detection methods, image analysis methods. Automated procedure for determination of length of the rail joint has been proposed. Also specific example of determination of length of the rail joint has been adduced.

Keywords Defectoscopy · Rail joint · Blur image · Image analysis · Convolution · Deblurring · Image enhancement · Image processing toolbox · Morphological filtering

1 Introduction

Nowadays many millions of different rails like tram and train ones are laid in the world. Some of these tracks relates to specific transportation facilities, such as subways, crane runways, carriers, trolleys, etc. Vulnerable point of tracks is a rail joint, which has various gaps. Further impact forces by passing rolling stock reaches the maximum value due to the lack of continuity in the joints of the track way. In other words, a sufficiency large wheel impact occur, resulting in decreased reliability of the track structure and carriages moving thereon.

A. Korobeynikov · V. Tkalich · S. Aleksanin · V. Polyakov (✉)
ITMO University, Kronverksky Pr., 49, 197101 Saint Petersburg, Russia
e-mail: v_i_polyakov@mail.ru

A. Korobeynikov
e-mail: office@izmiran.spb.ru

A. Korobeynikov
SPbF IZMIRAN, University Emb, Building 5, letter. B., 199034 Saint Petersburg, Russia

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_33

The rail ends in the joins deflect under load and form an angle. Thus, wheels hit the surface of the receiving rail ends at a slight angle to the vertical direction. The horizontal component of these forces creates running-away in the direction of the force directly proportional to the load.

However, the horizontal component of the force may not cause a noticeable running-away of rails (i.e. longitudinal movement of the rail by passing train wheels), because receiving rail ends are already pressed by collided wheels. The main reason for the running-away from hits in the joints lies in the fact that track way shakes and obtained partial discharge of the temperature stress happened in the receiving rail. Due to this, it changes the length and the opposite not clamped end slightly slips forward.

Based on the foregoing, the control of the length of rail joints is a prerequisite for ensuring the safe operation of rolling stock.

2 Digital Processing Methods Used for Image Enhancement in the Problems of Defectoscopy

Reliable identification of defects depends on the quality of the analyzed image. Poor quality can lead to misidentification or unreliability of the characterization of defects. In this case, reliable quality enhancements of such images is an important task that must be addressed in a low contrast and low sharpness images.

Image Enhancement—is the process of manipulating the image, in which it becomes more suitable for a particular application than the original one It is important the word "particular" because it establishes from the outset that the methods of image enhancement are problem-oriented [1, 2]. For example, a method that is quite useful for improvement of X-ray imaging, may not be the best approach to improve satellite images taken in the infrared range of the electromagnetic spectrum.

There is no general theory of image enhancement. If the image is processed to visual interpretation, then the appraisal how well a particular method works provides an observer. Improvement methods are very diverse and using so many different approaches to image processing, that it is difficult to collect meaningful set of appropriate techniques to improve in one article without making a separate extensive research.

The term filtration is quite often used for image enhancement, taking not only the removal of or compensation for noise and interference, but also directly to the selection of image information on the characteristics of locally heterogeneous objects. But it is necessary to consider that methods of image enhancement often lead to distortion of information about the objects present to them, for example, contrast enhancement and edge enhancement can lead to distortion of the shape and size of the object flaw, which is unacceptable.

A possible solution to this problem is to use not methods of image enhancement, but filtering techniques of useful signal, that can be interpreted here as image defects. The signals from the defects can be interpreted as two-dimensional local inhomogeneities of the final non-stationary stochastic signal. Then the problem of local inhomogeneities filtering image noise background occurs.

Formally, the process of distortion of the original image $f(x, y)$ can be presented in the following way:

$$g(x, y) = H(f(x, y)) + \eta(x, y), \tag{1}$$

where
$g(x, y)$ distorted image;
$\eta(x, y)$ additive noise;
$H(*)$ distorting operator

Based on (1) it is possible to formulate the problem as follows:
Initial point:

- distorted image $g(x, y)$;
- information about the operator $H(*)$ and its main parameters$\eta(x, y)$.

Required:

- an approximate image $f'(x, y)$ as close to the original image.

2.1 Blur Processing

In real terms during solving the problem of determining the length of the rail joints due to quite a high speed movement of the platform with the installed imaging equipment, there is often a distortion of the digital image, called blur.

It is widespread that the operation is irreversible and blur information is irrevocably lost as each pixel is converted into a spot, everything is mixed. Moreover a uniform color obtained throughout the image with a large radius of blur. It is not so, because all the information is distributed to a certain law and can be unambiguously restored with some reservations. The only exception is the edges of the image width of the blur radius where a full recovery is not possible.

In the process of distortion each pixel of the original image is converted into a spot in the case of defocus and in the segment for the case of a simple blur. It can be also presented as each pixel of distorted image is "aggregating" from a neighborhood pixels of the original image. All of this is superimposed on each other and as a result there is a distorted picture. Thus distorted function (or distortion) is the law how one pixel is blurred or aggregated to. Other synonyms are PSF (Point spread function), the core of distorting operator, kernel, and others. The dimension of this function is usually less than the dimension of the image.

Operation applying a distorting function to another function (an image in this case) is a convolution, i.e. some area of the original image is minimized to a single pixel of a distorted image. It is designated as an operator "*". Operation, which is opposite to convolution, called deconvolution and the solution of this problem is quite trivial.

A mathematical formulation of this problem should be given to improve the image (removal of a blur).

Let the camera with CCD-matrix makes uniform and rectilinear displacement (shift) by the time τ at a speed $v = const$ along a certain direction by the amount $\Delta = v\tau$. A fixed rectangular coordinate system $\eta O\xi$ should be defined. The direction of the axis ξ is compatible with the direction of the shift. In this case, the image on the CCD-matrix will become blurred (offset shifted) along the axis ξ. After that the problem of the restoration (reconstruction) of the real (non-distorted image) arises by using known the blurry image, the direction and amount of blur Δ.

One more movable rectangular coordinate system xOy should be defined for the mathematical formulation. Let the initial coordinate system xOy coincides with the coordinate system $\eta O\xi$, and the axis x is directed along the axis ξ. Further assume that on the selected point of CCD-matrix with coordinates (x, y) exposure time τ is projected continuous set of points P with abscissas $\xi = x$ starting from $\xi = x + \Delta$ with different intensity $f(\xi, y)$. The integrated intensity $g(x, y)$ at a point (x, y) equal to the sum (integral) of intensities $f(\xi, y)$, $\xi \in [x, x + \Delta]$:

$$g(x,y) = \frac{1}{\Delta}\int_{x}^{x+\Delta} f(\xi,y)d\xi. \tag{2}$$

Factor $\frac{1}{\Delta}$ placed before integral to performs $g(x, y) \rightarrow f(x, y)$ in the absence of blur ($\Delta \rightarrow 0$). Furthermore in the case of constant image ($f(x, y) = const$) condition $g(x, y) = const$ will always be correct.

The Eq. (2) can be presented as follows:

$$\frac{1}{\Delta}\int_{x}^{x+\Delta} f(\xi,y)d\xi = g(x,y). \tag{3}$$

The Eq. (3) is the base for the task of reconstruction of blurred images [3–5], where the function $g(x, y)$ is the measured image, axis x is directed along the blur, Δ—the value of blur, $f(x, y)$—the desired intensity distribution in the undistorted image (means, of the intensity that would have been in the absence of blur, when $\Delta = 0$).

The Eq. (3) is a homogeneous integral equation of the Volterra type relating to $f(\xi, y)$ for each fixed value y, which plays the role of a parameter. In other words, (3) there is a set of one-dimensional integral equations. Therefore, given the noise in the image, (3) can be written as [5]:

$$\frac{1}{\Delta}\int_{x}^{x+\Delta} f(\xi)_y d\xi = g_y(y) + \delta y, \tag{4}$$

where δy—noise

Based on the terminology of article [6], this equation is the classic equation of Volterra type I, since both variable limit of integration.

The integral Eq. (4) can also be called unusual, since it does not contain the kernel explicitly (or the kernel can be considered as $\frac{1}{\Delta} = const$).

It is necessary to draw attention, that value Δ is often not known a priori and is usually determined by the selection based on visual assessment of the values obtained for a number of decisions $f(\xi, y)$ [7]. In addition, it is possible to determine the magnitude and direction of the blur of the strokes in the picture, especially if at least one of these strokes are the result of blurring bright points in the image.

As a result solution to the integral Eq. (3) can be defined (more precisely, the set of equations), undistorted (real) image $f(x, y)$ can be reconstructed from a distorted snapshot-image $g(x, y)$ by selecting the axis x direction (along the blur) and the amount of blur Δ, i.e. by defining the parameters of the blur.

2.2 The Choice of Method Solution in the Problem of Image Reconstruction

The computational experiments have been made to select a method for deblurring task.

The image shown in Fig. 1 was taken for the experiment. This image has been processed the procedure of blurring the length of 10 pixels. The one percent of Gaussian noise was added to the result (Fig. 2). Reconstructing the image is done by two modifications of the quadrature method with Tikhonov regularization: 1—with blurred edges and overdetermined system of linear algebraic equations (SLAE), 2—with truncated edges and certain SLAE.

The relative standard deviation $\sigma_{rel} = (\alpha)$ was calculated in an experiment to quantify the error of the reconstructed image of the original by the formulas:

$$\sigma_{rel}(\alpha) = \frac{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}[f_\alpha(i,j) - f(i,j)]^2}}{\sum_{i=1}^{n}\sum_{j=1}^{m} f(i,j)^2}$$

Fig. 1 Original image

Fig. 2 Noisy image

where

α regularization parameter;
$f(i, j)$ original image;
$f_\alpha(i, j)$ reconstructed image at a given α.

Figure 3 logarithmically (more clear in this case, of visual perception) shows the dependence of σ on the regularization parameter α. Method 1 represented by curve 1, and, accordingly, method 2 represented by curve 2. As can be seen from the graph, the optimum value α for the method 1 equals $\alpha_{1opt} = 0.1$ and $\sigma_{rel}(\alpha_1) = 0.26$. The optimum value α for the method 2 equals $\alpha_{2opt} = 0.02$ and $\sigma_{rel}(\alpha_2) = 0.11$.

According to the deblurring results is recommended to apply the method 2 with regularization parameter $\alpha = 0.02$.

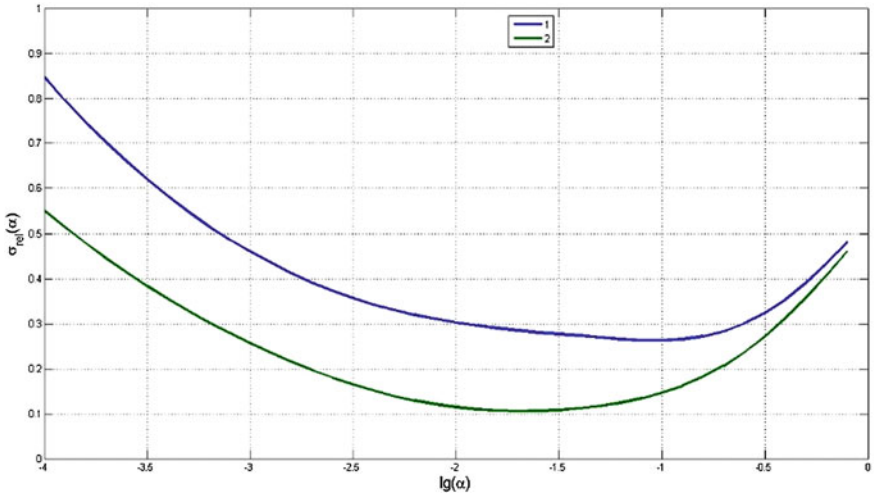


Fig. 3 Results of computational experiment

3 Automated Determination of the Length of the Rail Joint

Let review the developed method of digital image processing. It allows operator serving the instrumentation automatically determine the length of the rail joints. This technique is based on a morphological image processing techniques, namely dilation and erosion operations, as well as their different set of consecutive impacts called morphological closing and morphological opening operation [8].

It should be noted that the measurement of the length of the image in pixels occurs. Therefore, to convert to meters, or derivatives thereof, requires binding. In the present experiment, it will be equal to the width of the railhead 75 mm.

3.1 The First Stage

In the first stage the crop of the original image (Fig. 1) processed in order to highlight areas where there is a railway joint (Fig. 4).

Further, "thermal treatment" to minus 95 °C applied to better highlight the area of rail joint 95 °C (Fig. 5).

The first stage finalized with binarization, i.e. transfer of "cold" image in black-and-white (binary) image with a threshold 0.09 (Fig. 6).

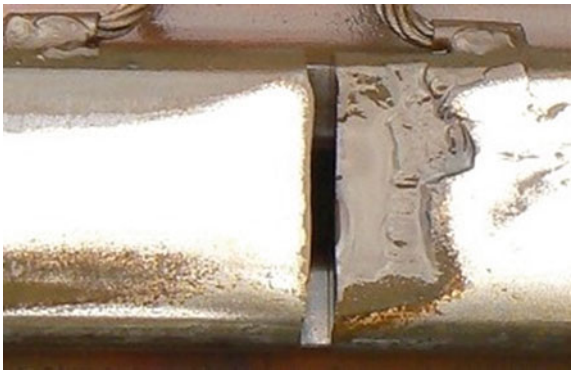


Fig. 4 Cropped image

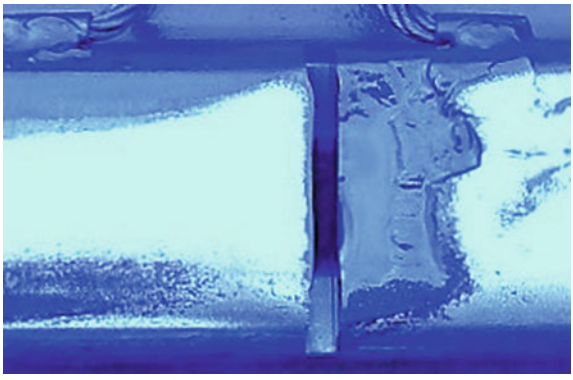


Fig. 5 "Cold" image

Fig. 6 Binary image

Fig. 7 Result of the morphological processing stage

3.2 The Morphological Processing Stage

At this stage the consistent application of morphological opening and closing operations to remove dark spots in the light area and the removal of the bright spots in a dark area. A generator of "Line" is taken to do this. This choice is dictated by the assumption that the area of rail joint can be considered almost rectangular.

Then the connected components are counted. In this example, the number was 21. In other words, 21 binary image areas are detected.

After that there is the calculation of these areas, except the area marked as 0 (zero). It is considered background. A region with a maximum area is taken. The intensity of the remaining areas translated into the background. The result of this step is presented in Fig. 7.

3.3 Determination of the Length of the Rail Joint

At this stage the maximum Euclidian distance is calculated in the pixel area in the X-axis shown in Photo 6 based on the assumption of a rectangular region of the rail joint. Result is 12.5 px.

Let use the assumption that format is the same for all images and translate to the absolute values of length. In this case, the width of the head of the rail is 75 mm corresponds to 155 px. Then the length of one pixel in the image always corresponds to 0.48 mm. These data can always be obtained before measurements. Thus in the image shown in Photo 1 determination of the length of the rail joint is 6 mm.

4 Conclusion

This article deals with the usage of digital image processing for automated determination of the length of the rail joint. Briefly discussed formulation of a method for deblurring task on the image. This problem occurs because the dynamics of the camera. The solution to this problem is proposed to conduct on the basis of

regularization method. A numerical example of the calculation of the regularization parameter is presented.

The stages of the automated determination of the length of the rail joint are described. Specific example of the image processing is adduced.

Based on the considered methods automated procedure for solving the problem of determination of the length of the rail joints can be created. This will increase the efficiency of operator serving the instrumentation. This, ultimately, will necessarily lead to an increase in the safety of rolling stock.

Also all numerical experiments were done using MATLAB system [9].

References

1. Korobeynikov, A.G., Aleksanin, S.A., Perezyabov, O.A.: Automated image processing using magnetic defectoscopy. ARPN J. Eng. Appl. Sci. **10**(17) 7488–7493 (2015). ISSN: 1819-6608
2. Grishencev, A.J., Korobejnikov, A.G.: Metody i modeli cifrovoj obrabotki izo-brazhenij, 190 s. Politehnicheskij universitet, Sankt-Peterburg (2014). ISBN: 978-5-7422-4892-7
3. Vakushinskij, A.B., Goncharskij, A.V.: Nekorrektnye zadachi. Chislennye metody i prilozhenija, 199 s. Izd-vo, MGU, Moscow (1989)
4. Vasilenko, G.I., Taratorin, A.M. Vosstanovlenie izobrazhenij, 304 s. Radio i svjaz', Moscow (1986)
5. Tihonov, A.N., Goncharskij, A.V, Stepanov, V.V.: Obratnye zadachi obrabotki fo-toizobrazhenij//Nekorrektnye zadachi estestvoznanija, S. 185–195. Izd-vo MGU, Moscow (1987)
6. Turchin, V.F.: Reshenie uravnenija Fredgol'ma I roda v statisticheskom ansamble gladkih funkcij. Zh. vychislit. matem. i fiziki, T. 7. № 6. S. 1270−1284 (1967)
7. Jagola, A.G.. Koshev, N.A.: Vosstanovlenie smazannyh i defokusirovannyh cvetnyh izobrazhenij//Vychislitel'nye metody i programmirovanie, T. 9, S. 207−212 (2008)
8. Gonzales, R.C., Woods, R.E., Eddins, S.L.: Digital Image Processing Using MATLAB, 827 pp., 2nd edn. Gatesmark Publishing (2009)
9. CHarl'z Genri EHdvards, Dehvid EH. Penni. Differencial'nye uravneniya i problema sobstvennyh znachenij: modelirovanie i vychislenie s pomoshch'yu Mathematica, Maple i MATLAB = Differential Equations and Boundary Value Problems: Computing and Modeling, 3-e izd. M.:«Vil'yams» (2007). ISBN: 978-5-8459-1166-7

Automatic Control of Train Brakes with Fuzzy Logic

Yurenko Konstantin

Abstract Automation of control pneumatic brakes allows improve safety by eliminating the human factor and to ensure energy-optimized operation of train motion. Subsystem of automatic control pneumatic brakes is part of the on-board autopilot system of the train. Based on the analysis of existing relay systems of automatic control of the brake is proposed and studied the control algorithm that uses nonlinear feedback from pressure in the brake cylinders and derivative of error speed control. To calculate the feedback signal used mathematical apparatus of fuzzy logic. Computational experiments were conducted using a software simulator of driver of the freight locomotive that is part of the teaching-research laboratory and training complex "Virtual railway RSTU". In paper investigated the stabilization mode the velocity of the train on long downhill with a slope of 9 % with application of the pneumatic brakes. Shows efficiency of the proposed algorithm in comparison with known.

Keywords Train brakes · Control · Fuzzy logic · Simulation · Computing experiment

1 Introduction

On Railway transport pneumatic brake is one of the major devices for ensuring safety of trains. It's using in Russian Railways regulates by the Rules of technical maintenance of the brake equipment and control of the brakes of railway rolling stock [1]. In the process of driving freight and passenger trains pneumatic brakes are used to implement different modes of reference of the train.

These include: reduction in the rate to perform permanent or temporary speed restrictions, defined by the lights (white or yellow); stopping braking at the station

Y. Konstantin (✉)
Rostov State Transport University, Rostov-on-Don, Russia
e-mail: ki-yurenko@yandex.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_34

or before forbidding signal of a traffic light; to stabilize the speed when driving downhill and the emergency braking.

Brake control requires consideration of the plan and profile path, the location of trains on the sections, the characteristics of the braking freight trains related to the effect of "braking wave", the restrictions of acceleration and the longitudinal dynamic forces in the train, and other factors. Execution these functions directly relates to traffic safety. Also should strive to reduce braking losses accumulated the kinetic energy and the motion time.

To reduce the influence of the human factor in the security and taking into account a numerous factors in the implementation of energy-optimal mode of conducting the train uses the automatic brake control subsystem [2], which are part of a complex hierarchical locomotive control system.

2 Automatic Control System of Train Brakes

2.1 Analysis of Existing Systems

The problem of brake control automation dedicated many works both in Russia and in other countries [2–7]. However, their analysis shows that most of them are devoted to the brake control automation speed and commuter trains as well as passenger trains. They use electric braking with continuous or step control or electro-pneumatic braking. However, the pneumatic braking freight train has its own characteristics that must be considered in the design of structures and simulation system. The braking of freight train evolves over time and over the length of the train. When setting the emergency or service brake mode begins pressure drop in the brake pipe, which extends to the finite speed of the locomotive to the tail of the coach. This causes sequential activation of brakes in the train cars. Similarly done and brakes release.

Currently, pneumatic brake control function in case of violation of traffic safety have onboard security CLUB and the SAUT, in the normal mode brake control, brake system control subsystem onboard the train automatic driving [2].

2.2 Functional Diagram of the System

Structure. General functional diagram of control system of pneumatic brakes of the train is shown in Fig. 1.

General functional diagram of the system controlling the pneumatic brakes of the train are shown in Fig. 1. The program block that is part of an onboard autopilot based on the information of speed limit, the coordinates of stations and traffic signals, stored in the onboard database, and codes of signals calculates the braking

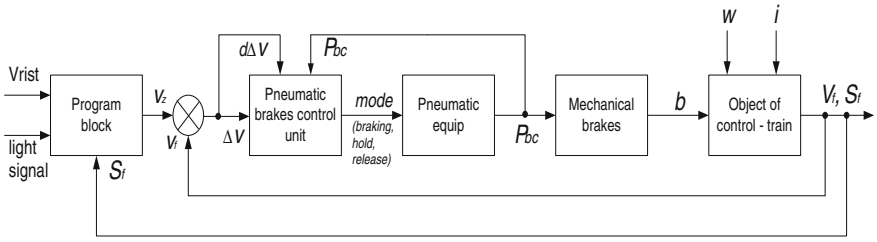


Fig. 1 Functional diagram of the automatic control system of train brakes

programs $V(S)$. The error of speed regulation is defined as the difference between the desired and actual speeds:

$$\Delta V = V_z - V_f \tag{1}$$

In addition, calculate the change (or derivative) of the error

$$d\Delta V = \Delta V_i - V_{i-1}. \tag{2}$$

Control unit pneumatic brakes is a relay element, which on basis of ΔV and additional correction signals: $d\Delta V$ and the current value of the pressure in the brake cylinders P_{bc} sets the work mode brakes—braking, hold or release. Depending on the mode of operation of the brake equipment is provided filling or emptying of brake cylinders. This determines the value of P_{bc}. Depending of P_{bc} makes the braking force $\boldsymbol{b}$, which along with the main forces of the resistance $\boldsymbol{w}$ and the additional resistance from the slope and the curves of the way $\boldsymbol{i}$ determines the nature of the motion of the train [8].

Relay control. In accordance with a known algorithm of the relay element [2] the mode of brake operation is calculated according to the following Eqs. (3), (4)

$$R(u) = \begin{cases} brake, if & u > b_1, \\ hold, if & b_2 \le u \le b_1, \\ release, if & u < b_2, \end{cases} \tag{3}$$

$$u = \Delta V + k_1 \cdot P_{bc} + k_2 \cdot d\Delta V, \tag{4}$$

where k_1 and k_2 are the parameters of corrective feedback on P_{bc} and $d\Delta V$; b_1 and b_2 are parameters defining the switching characteristics of the relay element. This algorithm is used in the SAUT system. To use it requires accurate selection of parameters k_1, k_2 and b_1, b_2.

However, the selected values for one mode may not provide the necessary quality control in other modes. So, in [2] considered the use of this relay element for braking to reduce speed. To maintain speed using the pneumatic brakes on the slope may be requires other settings.

This, in particular, is shown a imperfection of rule (4) based on the linear relationship. Therefore, improvement of the structure, operating principles and algorithms of automatic brake control system is an urgent task.

2.3 *Fuzzy-Logic Controller*

Input and output. In this paper we propose to use the mathematical apparatus of fuzzy logic and fuzzy inference method Mamdani [9]. In this case u for the formula (3) will find as,

$$u = \Delta V + K_{cfb}(d\Delta V, Pbc), \tag{5}$$

where K_{cfb}- corrective feedback. Thus obtain a fuzzy-logic controller (FLC) with two inputs and one output (Fig. 2).

Linguistic variable. Introduce the linguistic variables corresponding to a simple variable $d\Delta V$, P_{bc}, and K_{cfb} with the following membership functions (see Figs. 3, 4 and 5), there NB—negative big, NM—negative medium, NS—negative small, ZE—zero equation, PS—positive small, PM—positive medium, PB—positive big.

Fig. 2 Inputs and output FLC

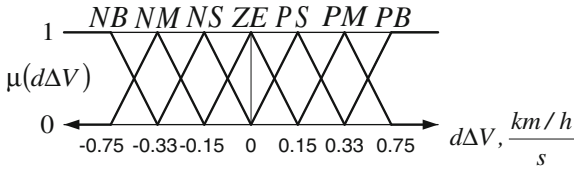


Fig. 3 Term-sets for linguistic variable $d\Delta V$

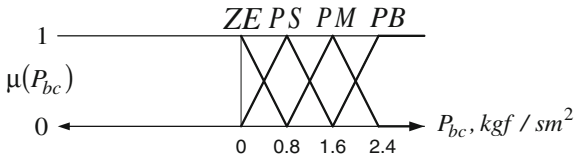


Fig. 4 Term-sets linguistic variable P_{bc}

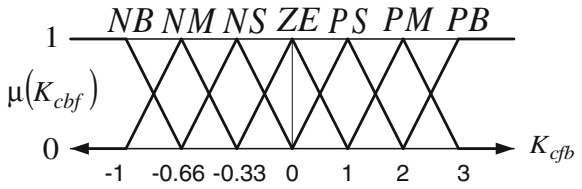


Fig. 5 Term-sets linguistic variable K_{cfb}

Table 1 Rules for fuzzy-logic controller

№	$d\Delta V$	P_{bc}	K_{cfb}
1	ZE	ZE	ZE
2	NS	ZE	NS
3	NM	ZE	NM
4	NB	ZE	NB
5	PS	ZE	PS
6	PS	PS	PM
7	PS	PM	PB
8	PM	PS	PB
9	PM	PM	PB
10	ANY	PB	PB

Table of rules for FLC. For the fuzzy controller Mamdani developed the following rule base (Table 1).

2.4 Computational Experiment

The goal of computational experiment was to compare at the same parameters of the object (freight train) quality of control, that achieved in system with a known calculation algorithm corrective actions and proposed in this paper based on fuzzy logic controller.

To simulate the data were collected on the site of "Rostov-Tovarny" - "Kiziterinka" North-Caucasian railway. This section of track is a downhill slope with 9 %. In carrying out computational experiments used the software simulator freight locomotive driver, a member of the teaching and research laboratory and training complex "Virtual Railway RSTU" [10]. it implements the model of train motion including calculation longitudinal dynamic response [11]. The graphics of velocity and pressure in the brake cylinders in each case are shown in Fig. 6.

Analysis of the simulation results shows that in the first case, the average speed on the descent was 55.9 km/h, while the second—57 km/h. Thus, due to the effect of the proposed approach achieved electricity savings in traction due to the fact that in the future less energy will be spent on the acceleration of the train. Another

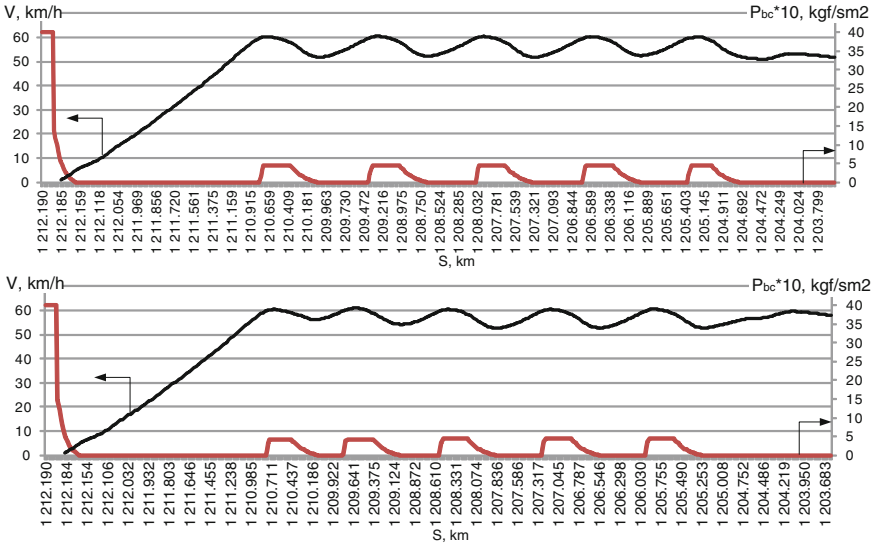


Fig. 6 Graphics velocity and pressure for use relay regulator with linear feedback (*upper*) and feedback with fuzzy logic (*low*)

qualitative result is reduced wear of bandage wheel pair by reducing the total braking forces. However, quantify the results needs further investigation.

3 Conclusions

1. The results demonstrate the fundamental possibility of increasing the efficiency of pneumatic freight train brake control system through the use of non-linear correcting feedback on the pressure in the brake cylinders and the derivative of the speed control error which calculated using the mathematical apparatus of fuzzy logic.
2. A useful effect compared with the method of calculating the corrective effect linearly in stabilization mode speed on downhill is to increase the average speed, which helps to reduce power consumption due to increased time reserve for the remaining motion. An additional qualitative effect is to reduce wear bandages wheel sets due to the overall reduction of braking forces.
3. Positive results of the study reveal potential application of fuzzy logic to control in other modes of conducting a freight train.

References

1. Rules of technical maintenance of the brake equipment and brake control of the railway rolling stock, 178 p. Rostov-on-Don (2015)
2. Nikiforov, B.D., Golovin, V. N., Kutiev, U.G.: Automation control braking of trains, 263 p. Transport, Moscow (1985)
3. Sankar, G., Saravana Kumar, S.: Fuzzy logic based automatic braking system in trains. In: India International Conference on Power Electronics, pp. 383–387, Chennai. IICPE (2006)
4. Yasunobu, S., Miamoto, S., Ihara, H.: A Fuzzy control for train automatic stop control. Transactions of the Society of Instrument and Control Engineers, vol. E-2, no. 1, 1 Sep 2002
5. Mamun, B.I.R., Jubayer, J., Hafizah, H., Badariah, B.: Subway train braking system: a fuzzy based hardware approach. Am. J. Appl. Sci. **8**(7), 740–747 (2011)
6. Yurenko, K.I., Sapunkov, A.N., Fandeev, E.I.: Automatic brake control of train based on the mathematical apparatus of fuzzy logic in the autopilot system. Visnik of the East Ukrainian Nnational University V. Dal. Tech. science. A Series of Transport. Part 2, no. 5 (176). pp. 22–29. Lugansk (2012)
7. Yurenko, K.I., Fandeev, E.I: The device of automatic control of train brakes and the method of its implementation. Patent RU 2540212 B 60T 8/172. Appl. 09/07/13. Published in Bull. No. 4 10 Feb 2015
8. Grebenyuk, T.P.: Rules of brake calculation, 115 p. Intex, Moscow (2004)
9. Mamdani, E.H., Assilian, S.: An Experiment in linguistic synthesis with fuzzy logic controller. Int. J. Man-Mach. Stud. **7**(1) 1–13. (1975)
10. Vereskun, V.D., Kolesnikov, V.I., Sukhorukova, N.N.: Educational research laboratory complex "Virtual Railway" In: 3rd International Scientific and Engineering. Conference on Computer Simulation-2002. pp. 203–209. STU, Saint Petersburg (2002)
11. Blokhin, E.P., Manashkin, L.A.: Train dynamics (nonstationary longitudinal oscillations), 222 p. Transport (1982)

Part VI
Technologies for Industrial and Autonomous Robots

Indoor LiDAR Scan Matching Simulation Framework for Intelligent Algorithms Evaluation

Jaromir Konecny, Michal Prauzek and Jakub Hlavica

Abstract Localization in mobile robotics is one of the most challenging concerns, taking into account the demand on perfect accuracy and quick response. However, high-performance approaches in conjunction with cutting-edge technologies are not necessarily applicable in every case, and thus an optimized localization algorithms suitable for implementation in low-end hardware applications are to be favorable to fill the market niche. Simulation framework, introduced in this contribution, is capable of performing simulations of systems with LiDAR and model an ambient environment by means of user-defined vector maps. Modeled laser sensor is SICK LMS 100. The framework, developed in C# language, enables the user to generate laser scans from user-defined vector maps and trajectories. Scans can subsequently be used for simulations. Computational method considered in this study is particularly Scan Matching.

Keywords Localization · Scan matching · Simulation framework · SLAM

1 Introduction

Accurate and efficient localization is one of the major aspects in mobile robotics. It enables the robot to determine its position in diverse environment (e.g. waste rock exploring [8]). There are basically two localization approaches: obtaining the exact position by receiving signal from a beacon, such as Finger Printing algorithms [12],

J. Konecny (✉) · M. Prauzek · J. Hlavica
VSB - Technical University of Ostrava, 70833 Ostrava-Poruba,
Czech Republic
e-mail: jaromir.konecny@vsb.cz
URL: http://cbe.vsb.cz

M. Prauzek
e-mail: michal.prauzek@vsb.cz

J. Hlavica
e-mail: jakub.hlavica@vsb.cz

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_35

or employing localization subsystem that provides the robot with information about its ambient environment. The latter is investigated in detail within the scope of this study. Commonly used devices for localization are, for instance, wheel sensors (odometers) and ultrasonic or optical rangefinders [6].

Optical rangefinders create two dimensional laser scans of the ambient environment and provide very accurate measurements with high sampling frequency. Unfortunately such data processing is considerably demanding and in most cases requires high computational power [11] in order to facilitate orientation in large-scale maps [2] or in high-speed navigation applications [5].

Control systems of mobile robots are mostly embedded systems with low consumption, having limited power and memory [9] and employing low-cost sensors [3]. For these reasons, there is a need to develop and optimize fast and efficient algorithms applying modern methods. Currently, there are several simulation and evaluation frameworks, for instance Open Source Software Platform for Web-based Management, Visualization and Analysis of LiDAR Data [7] or Framework for Applying Point Clouds Grabbed by Multi-beam LiDAR in Perceiving the Driving Environment [10].

Our study introduces an extension in form of Simulation framework that facilitates verification and development of intelligent method for indoor environment laser scans processing.

This paper is arranged in five sections. Section 2 describes essential principles of relative localization. Section 3 introduces the developed Simulation framework. Section 4 comments on results and license policy. Finally, the major conclusions and the focuses of future work are presented in Sect. 5.

2 Background

2.1 Intelligent and Adaptive Scan Matching Methods

There are several Scan Matching methods suitable to use for relative localization, from which particularly Intelligent Scan methods are being applied recently. These localization methods are based on learning and adaptive algorithms, namely, Perimeter-Based Polar Scan Matching (PB-PSM) [4] or Adaptive Iterative End Point Fitting (IEPF) [1]. Evaluation is performed in individual iteration steps in time-invariant environment [14]. Consequently, it is necessary to use a tool able to generate laser scans on a defined trajectory and include exact description of the presented task.

2.2 LiDAR Sensor

Simulation framework, described in this contribution, uses principally a 2-D mathematical model of laser sensor SICK LMS 100 in order to generate data [13].

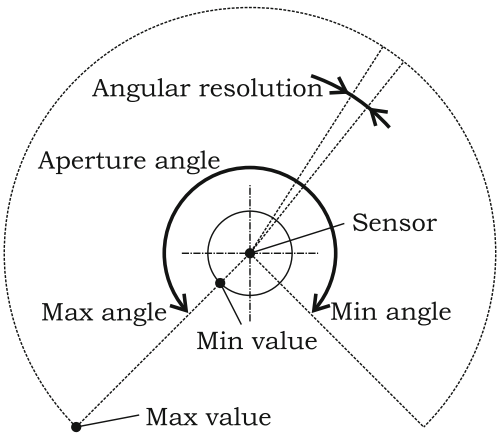


Fig. 1 LiDAR measurement model

Measuring beam of this sensor oscillates in horizontal plane at the angle of 270° (ranging from −45° to 225°) and measures distance to the nearest obstacle. Angular resolution is 0.25°. The minimum measurable distance is 500 mm while the maximum is limited to 20 000 mm. Figure 1 depicts LiDAR measurement model describing operating range and angular resolution. The output from sensor is a clockwise vector $\boldsymbol{R} = (R_0, R_1, \ldots, R_k)$, where k is the number of measured distances (for SICK LMS 100 k is equal to 1082) and R_i is distance to the nearest obstacle within the angle $(-45 + 0.25 \cdot i)^\circ$.

2.3 LiDAR Scan Model

To create the laser scan described in Sect. 2.2, a model of environment in form of a vector map needs to be defined. Map $\boldsymbol{M}$ is a vector of objects O_i. Object can be any geometric primitive, such as a segment line. Laser scan generating process (represented in Fig. 2) is described as follows:

(a) Model total detection area using the following parameters: operating range and aperture angle.
(b) Position the scan onto the map with objects according to the coordinates of sensor's shift and rotation.
(c) Crop the detection area with respect to the position and shape on the map.
(d) Shift and rotate the scan back into the origin of coordinates, delete non-cropped detection area, edit the points with minimum distance and add a noise.

To be able to perform this sequence, a set planar operations have to be applied. In particular, conversion into Cartesian coordinates system and affine

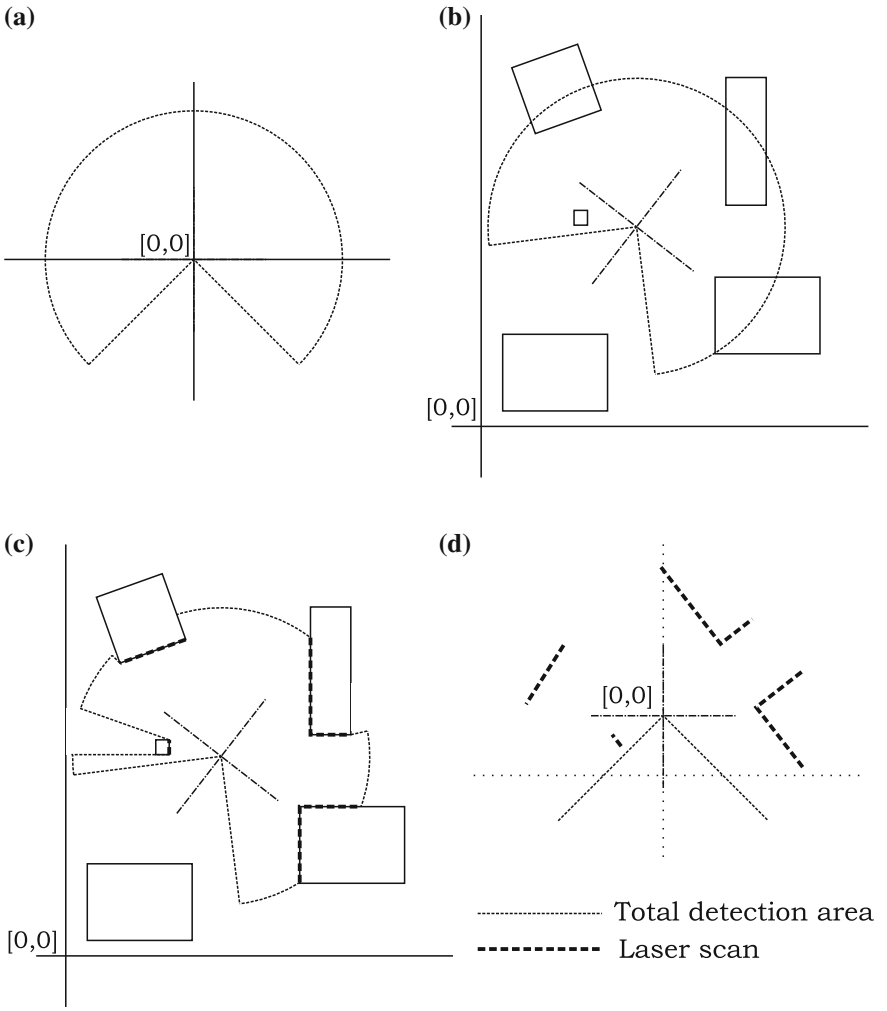


Fig. 2 Laser scan generating model. **a** Model of total detection area, **b** Shift and rotation of total detection area in the map according to coordinate system and rotation of the sensor, **c** Cropped detection area with respect to object on the map, **d** Shift and rotation back into the system of coordinates, deletion of non-cropped detection area, editing of the point with minimum distance, and noise addition

transformation. The conversion from polar coordinates system into Cartesian is described by formula 1:

$$\mathbf{C} : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r(\Phi) \cdot \cos \Phi \\ r(\Phi) \cdot \sin \Phi \end{pmatrix}, \tag{1}$$

Table 1 Vector map XML structure

Element	Sub-elements	Attributes
Map	Base, entities, routes	–
Entities	Entity	–
Base	–	Position
Entity	Points	Name, type
Points	Point	–
Point	–	Loc
Routes	Route	–
Route	Track	Name
Track	–	Start, steer, speed, steps, type

where r is distance of the point from the origin of the coordinates, Φ is an angle, x and y are Cartesian coordinates.

Affine transformation, $\mathbf{T}(\boldsymbol{p})$, based on vector of parameters $\boldsymbol{p} = (p_x, p_y, \omega)$, is defined as:

$$\mathbf{T}(\boldsymbol{p}) : \begin{pmatrix} b_x \\ b_y \end{pmatrix} = \begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix} \cdot \begin{pmatrix} a_x \\ a_y \end{pmatrix} + \begin{pmatrix} p_x \\ p_y \end{pmatrix}, \tag{2}$$

where (a_x, a_y) and (b_x, b_y) are the coordinates of a point, $(x,\ y)$, A and B, respectively.

White noise is added in the final step. The noise is there to illustrate an imperfection of the sensor. The said white noise is manifested in the measured value in form of measured distance error. Random number generator is used to generate the white noise. The maximum magnitude of the noise is, considering the employed sensor, set to 30 mm.

3 Scan Matching Simulation Framework

Scan matching simulation framework is a tool that enables the user to create laser scans of LiDAR and model an ambient environment by means of user-defined maps. Vector map files contain a set of defined trajectories that are applied in a device with mounted LiDAR. Laser scans, stored in `CSV/DSV` files, are generated from user-defined vector maps and user-defined trajectories. These scans can then be utilized in testing and validation of intelligent algorithms.

3.1 User-defined Vector Maps

Vector maps are stored in `XML` files, and therefore can be generated and edited using commonly available tools. Table 1 describes structure of vector map. Two

components represent elements in vector map file. The first component (`entity`) is an individual object on a map. This component supports polygonal chain (`polyline`). The map of obstacles is therefore composed of a set of polygonal chains. The second component (`routes`) comprises predefined trajectories which are followed by a device (e.g. robot) having laser sensor.

Two approaches are implemented when it comes to definition of trajectory. Absolute positioning is characterized in that the sensor is shifted from a defined start point (or from its previous position in case the point is not specified) to an end point following a straight line. The individual shifts are executed in defined steps with constant length.

Relative positioning, on the other hand, is characterized in that the sensor is shifted from a defined start point (or from its previous position, respectively) at specified speed in defined steps with constant length towards direction given by the orientation of the sensor.

The selection of these positioning methods principally depends on kinematic model of the robot whose motion is defined.

3.2 Vector Maps Processing

Computer processing of vector map $M = (O_1, O_2, O_3, \ldots, O_k)$ generally utilizes object-oriented feature, particularly `Interface`. Vector map is constituted of array of elements `IBeam`. `IBeam` is an interface implementing `CutBeam` method.

```
bool CutBeam(DVector start,
             DVector end,
             out DVector intersection);

returns:       true, when intersection is found
start:         start LiDAR position
end:           end point of particular laser beam
intersection:  calculated intersection
```

Every vector map object implements interface `IBeam`, and is therefore a generator of laser scan, capable of handling with any other object on map, that can return an intersection with laser beam. Hence, each object on vector map can become an obstacle to the laser beam.

3.3 Laser Scan Generating Algorithm

The algorithm can be described as follows:

1. Create scan $\boldsymbol{S} = (A_0, A_1, \ldots, A_k)$, where k is the number of measured beams.
   - For measured angle Φ_i of laser scan $\boldsymbol{S}$: $A_i = \mathbf{T}(\boldsymbol{p}) \cdot \mathbf{C} \cdot [l_{\mathrm{MAX}}, \Phi_i]$, where $\boldsymbol{p} = (p_x, p_y, \omega)$, l_{MAX} is maximum measured distance of laser sensor, ω is rotation and $P = (p_x, p_y)$ is position of laser sensor. p_x is x position and p_y is y position.
2. For each point A_i of laser scan $\boldsymbol{S}$:
   - For each object O_j on map $\boldsymbol{M}$:
   - Find an intersection with line having start point in P and end point A_i with map object O_j.
   - If the intersection is found, substitute point A_i with the said intersection I.
3. Shift and rotate the cropped scan $\boldsymbol{S}$ into origin of coordinates $[0, 0]$ using affine transformation $\boldsymbol{T}(\boldsymbol{p})$, where $\boldsymbol{p} = (-p_x, -p_y, -\omega)$
4. Aggregate noise $N : w(t)$
5. For each point A_i from $\boldsymbol{S}$ calculate the distance from [0, 0] and insert it into array $\boldsymbol{R} = (R_0, R_1, \ldots, R_k)$
6. For each point R_i from $\boldsymbol{R}$: IF $R_i < l_{\mathrm{MIN}}$, THEN $R_i = l_{\mathrm{MIN}}$, where l_{MIN} is the minimum distance measurable with the sensor.
7. For each point R_i from $\boldsymbol{R}$: IF $R_i = l_{\mathrm{MAX}}$, THEN $R_i = 0$.

3.4 Output Data

Scan is stored in `CSV/DSV` file. Structure of one row of the `CSV/DSV` file is defined by formula 3:

$$x; y; \Phi; R_1; R_2; R_3; \ldots; R_k, \tag{3}$$

where values are separated by semicolon, x and y represent the sensor's position on map, and R_k are individual distances of laser scan.

The output data are arranged in directory structure. Every directory includes `CSV/DSV` file containing created values and `XML` file containing identification of said values along with user's description of the data.

4 Results

The application is composed of one `.exe` file and was developed in C# programming language in .NET Framework 4.5.1. It is executable in Windows 7 platform or higher. This application collaborates with the directory-structured database along with directory containing maps (`XML` files).

4.1 Framework Functionality

The described application enables user to open a predefined or user-defined vector maps, and also examine already generated laser scans. Collaboration with external tools allows to draw maps in AutoCAD and subsequently generate output in `XML` file using extension. AutoCAD extension, however, is not part of Simulation framework. Using the said maps, user can simulate motion of a robot on map and then generate laser scans.

Figure 3 shows application's functionality preview of robot's motion in a room. The robot is illustrated with a white arrow; the modeled trajectory is represented by the thick gray line while the thick black line represents the edges detected by the sensor. The resulting laser scan is then created from detected edges.

The application is also capable of recording scans from the connected laser sensor SICK LMS 100 stored in `CSV/DSV` file. When connecting the laser sensor the user needs to manually enter its exact position and space orientation. Neither Scan Matching nor SLAM algorithm are implemented due to the fact that data are

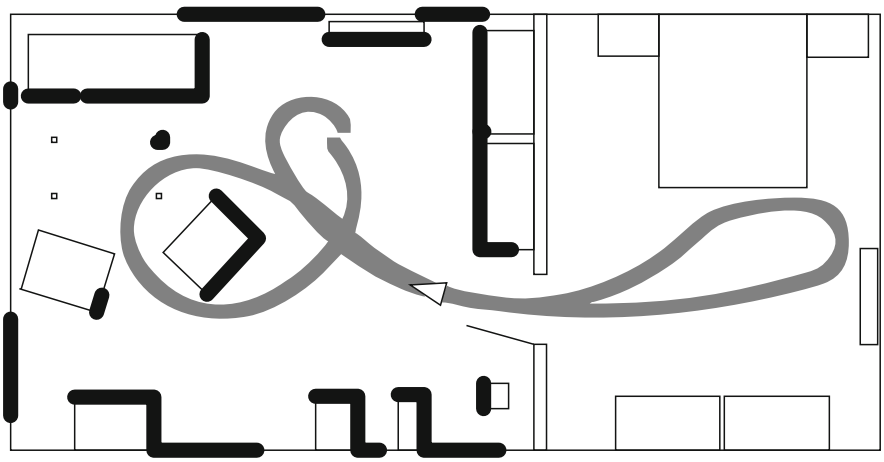


Fig. 3 Graphical representation of robot's motion on map following defined trajectory. *Thin black lines* represent walls and furniture, *thick gray line* illustrates robot's trajectory and *thick black line* depicts the resulting laser scan

primarily used for validation of the algorithm, and therefore the information has to be inputted manually.

4.2 Downloading and License Policy

Simulation framework is available at http://homel.vsb.cz/~kon430/smsf/, including source codes, compiled executable application, reference maps and generated laser scan models stored in `CSV/DSV` file. It is licensed under the Academic Free License version 3.0. Cite this paper in references in case you use this Simulation framework.

5 Conclusions and Future Work

This study presents a Simulation framework based on Scan Matching method. It is principally designed for testing and verification of an algorithm implementing the above mentioned method. However, the framework is also applicable for tests and simulations of localization algorithms that utilize two dimensional LiDAR. Resulting laser scans are stored in CSV/DSV files. Simulation framework has been developed in C# language and is available for download of a compiled program and source codes, along with sets of sample data and basic maps.

Future work will now focus on extending the capabilities of the framework, particularly by supplementing arc-shaped and elliptical objects to the map. This way, the rounded objects will be supported and performance of the framework will be enhanced. Also, considering the fact that there is currently no tool for map design, the framework will be extended to provide a possibility to create user-defined maps.

Acknowledgement This work was supported by the project SP2016/162, 'Development of algorithms and systems for control, measurement and safety applications II' of Student Grant System, VSB-TU Ostrava.

References

1. An, S.Y., Kang, J.G., Lee, L.K., Oh, S.Y.: SLAM with salient line feature extraction in indoor environments. In: 11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010, pp. 410–416 (2010)
2. Bosse, M., Zlot, R.: Map matching and data association for large-scale two-dimensional laser scan-based SLAM. Int. J. Robot. Res. **27**(6), 667–691 (2008)
3. Choi, M., Choi, J., Chung, W.: Correlation-based scan matching using ultrasonic sensors for EKF localization. Adv. Robot. Utrecht, 1495–1519 (2012)

4. Friedman, C., Chopra, I., Rand, O.: Perimeter-based polar scan matching (PB-PSM) for 2D laser odometry. J. Intell. Robot. Syst. Theory Appl. **80**(2), 231–254 (2015)
5. Furukawa, T., Dantanarayana, L., Ziglar, J., Ranasinghe, R., Dissanayake, G.: Fast global scan matching for high-speed vehicle navigation. In: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Oct 2015, pp. 37–42 (2015)
6. Ge, S., Lewis, F.L.: Autonomous Mobile Robots. CRC/Taylor (2006)
7. Jung, J., Pijanowski, B.: LiDARHub: a free and open source software platform for web-based management, visualization and analysis of LiDAR data. Geosci. J. **19**(4), 741–749 (2015)
8. Konecny, J., Kelnar, M., Prauzek, M.: Advanced waste rock exploring by mobile robot. Appl. Mech. Mater. **313–314**, 913–917 (2013)
9. Konecny, J., Prauzek, M.: Implementation of orienteering methods for advanced autonomous robot. Lect. Notes Eng. Comput. Sci. **1**, 378–382 (2013)
10. Liu, J., Liang, H., Wang, Z., Chen, X.: A framework for applying point clouds grabbed by multi-beam LiDAR in perceiving the driving environment. Sensors (Switzerland) **15**(9), 21931–21956 (2015)
11. Liu, Y., Zhang, H.: Towards improving the efficiency of sequence-based SLAM. In: IEEE International Conference on Mechatronics and Automation (ICMA), 2013, pp. 1261–1266 (2013)
12. Machaj, J., Brida, P.: Impact of radio map simulation on positioning in indoor environtment using finger printing algorithms. ARPN J. Eng. Appl. Sci. **10**(15), 6404–6409 (2015)
13. SICK AG: LMS100, http://www.sick.com/group/EN/home/products/product_news/laser_measurement_systems/Pages/lms100.aspx
14. Sprunk, C., Tipaldi, G., Cherubini, A., Burgard, W.: LiDAR-based teach-and-repeat of mobile robot trajectories. In: IEEE International Conference on Intelligent Robots and Systems, pp. 3144–3149 (2013)

Tail-Assisted Active Controller of a Mobile Unmanned Aerial Vehicle

Andrey S. Solovyov and Valery A. Kamaev

Abstract This paper introduces a control scheme for a quadcopter with 3-degree of freedom tail. There are some works in manipulating mobile robot stability at this moment. However, in the field of aerial vehicle control, it needs an advanced investigation to present an experience of flight correction in one time with moving the manipulators. We present the results of modeling of a control scheme for a quadcopter with manipulator and analysis of efficiency.

Keywords Mobile · Manipulating · Unmanned · Aerial · Vehicle · Maneuver · State · Quadcopter · Model

1 Introduction

The large segment of last research papers is studying of robust moving of quadcopter at various environment conditions. The main direction of that research segment is the application of various techniques of attitude stabilization, influence configuration of devices mechanics and their flight control style. The novel point of view is that the complex dynamic system Body-Payload has to be investigated. However, the payload can be kinematically active equipment as well as passive. In the first case, the movement of quadcopter can be nonstable and not robust.

The development of various use cases for quadrotors producing new system component—manipulators. The problem of robust movement with manipulator is already studied for ground mobile robots. In some works [4–6] authors are applying their solutions to compensate deviated properties of the dynamic and static model.

Several authors have examined manipulation process of UAV in the context of sustainable trajectory tracing and attitude position. However, periods of flight when quadrotor is not using manipulator processed only in stabilization plate. In this paper, we are considering manipulator as a mechanical "tail" in flight periods when

A.S. Solovyov (✉) · V.A. Kamaev
Volgograd State Technical University, Volgograd, Russia
e-mail: a.s.solovyov@gmail.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_36

it's not used for the intended purpose. For example, we are analyzing periods of transportation to and from the object of manipulation. A static and kinematic model of "tail" is described in Sect. 2. The simulation of introduced model is presented in Sect. 3. The dynamic system presented in this paper successfully showed positive results in adaptive control in-flight maneuver.

2 Mathematical Model

2.1 Manipulator Kinematic Model

For describing the kinematic model of a chain of the manipulator, we choose Denavit-Hartenberg parameters (DH-parameters). The set of parameters values described in Table 1. Parameters θ, d, α, and r are given in standard DH convention.

Parameters q1, q2, r1, r2 and re are joint angles and constant distance between joints respectively. To simplify the model we set r1 = r2 = re. Details of parameters are described in Fig. 1.

In Fig. 1 Xb-Yb-Zb system is body local coordinate system; Xw-Yw-Zw—world system; the 1 and 2 are indexes of joints; 0—is a point of fixed contact with quadcopter body.

According to direct kinematics, result transform function obtained by multiplying the transformation matrices:

$$T_b^e = T_b^0 T_0^2 T_2^e \tag{1}$$

Described transformation (1) made in quadrotor local frame. If we need to calculate result E position, we should apply additional transformation T_w—world frame matrix [1]. The model of manipulator dynamics is obtained from the RR model using the recursive Newton-Euler approach [1, 2]. Total equations are calculated by adding quadcopter dynamics to the base frame [3]. In addition, to simplify our model we set up conditions only for low-level attitude aircraft [4]. The necessary forces are obtained after forward step calculations of joint speed and accelerations, applying Newton-Euler laws.

Table 1 Denavit-Hartenberg parameters for the tail model

Link	θ	d	r	α
B-0	π/2	0	0	0
0	0	0	0	0
1	$q_1 - \pi/2$	0	r_1	$-\pi/2$
2	q_2	0	r_2	π/2
E	0	0	r_e	0

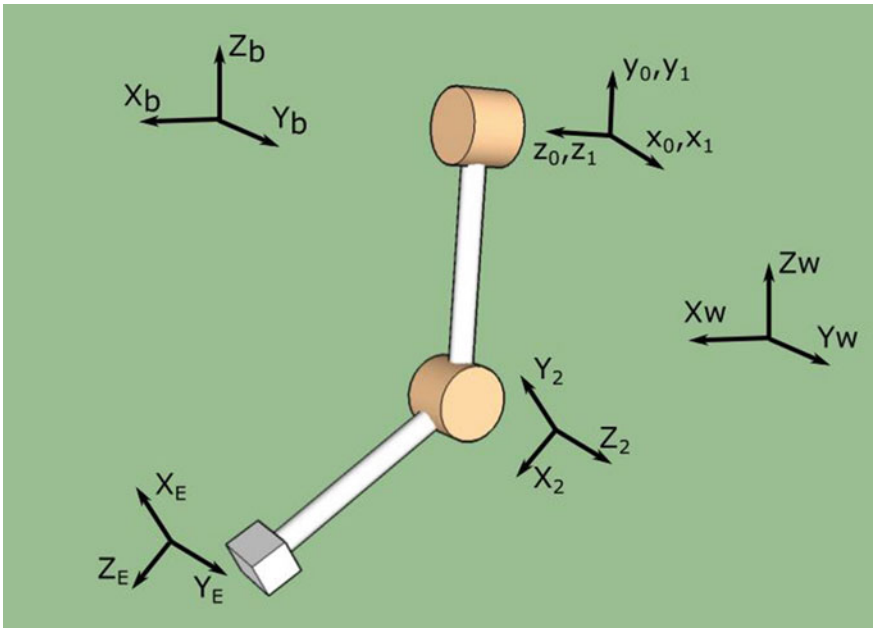


Fig. 1 Reference frames for system body-tail

2.2 *Main Body Dynamics*

As it was mentioned in the introduction of this paper, we are focusing on researching of manipulator dynamics and its influence on quadcopter dynamics and robust characteristics. That is why we are dropping aerodynamic effects.

The quadcopter dynamics are introduced in many works by now [1, 4]. Because we represent manipulator dynamics through the recursive Newton-Euler method, it is possible to setup body motion as a separate model from the manipulator. In [1], the authors present the following model:

$$\begin{bmatrix} F \\ \tau \end{bmatrix} = \begin{bmatrix} m_q \mathrm{I} & 0 \\ 0 & \mathrm{J}_q \end{bmatrix} \begin{bmatrix} \overrightarrow{\ddot{\vartheta}} \\ \overrightarrow{\dot{\omega}} \end{bmatrix} + \begin{bmatrix} 0 \\ \overrightarrow{\omega_q} \times \mathrm{J}_q \overrightarrow{\omega_q} \end{bmatrix} \tag{2}$$

The propulsion system torque and thrust can be estimated using NACA-standardized thrust and torque coefficients C_T and C_Q as rotor thrust and torque respectively. We used measuring results for various propellers used in unmanned aerial vehicles described in [5].

$$C_T = \frac{T}{\rho n^2 D^4}, C_Q = \frac{Q}{\rho n^2 D^5} \tag{3}$$

The air density marked as ρ; n is rotor speed; D is a rotor radius. Forces and torques of each propeller calculated to standard quadrotor propulsion system equations [1] as shown in (4–7):

$$\overrightarrow{F_{tot}} = \overrightarrow{T^1} + \overrightarrow{T^2} + \overrightarrow{T^3} + \overrightarrow{T^4} \tag{4}$$

$$\tau_x^{tot} = \tau_x^2 + \tau_x^3 - \tau_x^1 - \tau_x^4 \tag{5}$$

$$\tau_y^{tot} = \tau_y^3 + \tau_y^4 - \tau_y^1 - \tau_y^2 \tag{6}$$

$$\tau_z^{tot} = \tau_z^2 + \tau_z^4 - \tau_z^1 - \tau_z^3 \tag{7}$$

The stability analysis of given dynamic model (body and manipulator chain) is utilized to establish stability criteria for the complete system. Most of the works in this area operates with synchronized quadrotor center if mass and geometric center. As it is shown in Figs. 2 and 3, the center of mass Q_{CM} has downward offset in the Z-direction. Movement of the manipulator also moves the center of mass of our system.

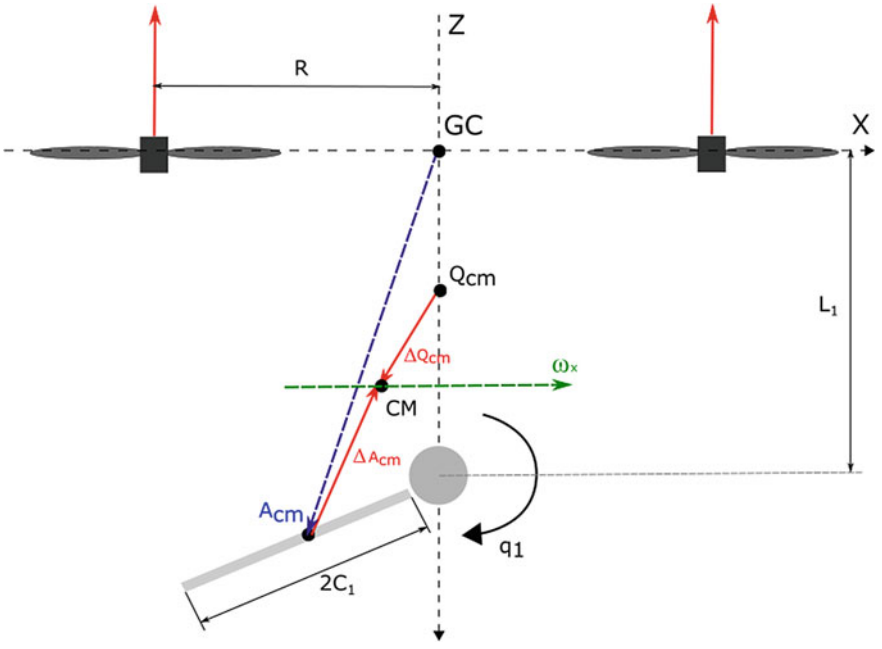


Fig. 2 Kinematic scheme of system quadcopter-manipulator in Z-X

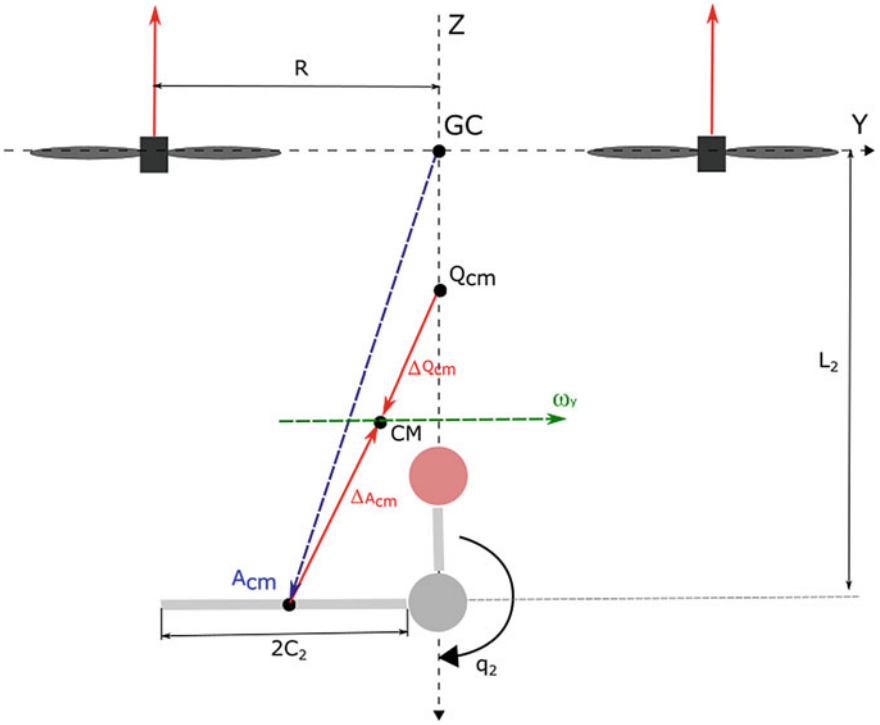


Fig. 3 Kinematic scheme of system quadcopter-manipulator in Z-Y

Kormela and team in [1] present a simplified kinematic model for moving arms. The simplification of the model allows describing varying center of mass (CM):

$$CM = \frac{Q_{cm}m_Q + A_{cm}m_A}{m_Q + m_A} \tag{8}$$

In (8) Q_{CM} is a vector of a quadcopter center of mass; A_{CM}—vector of manipulator CM; m_A—mass of manipulator; m_Q—mass of quadcopter body. Vector A_{CM} is changing due to manipulator movements. Using Parallel axis theorem the overall moment of inertia CM changes can be delivered:

$$I_{CM} = I_Q + I_A + m_Q \Delta Q^2 + m_A \Delta A^2 \tag{9}$$

where ΔQ, ΔA represent the center of mass of each body with respect to the overall center of mass CM (10).

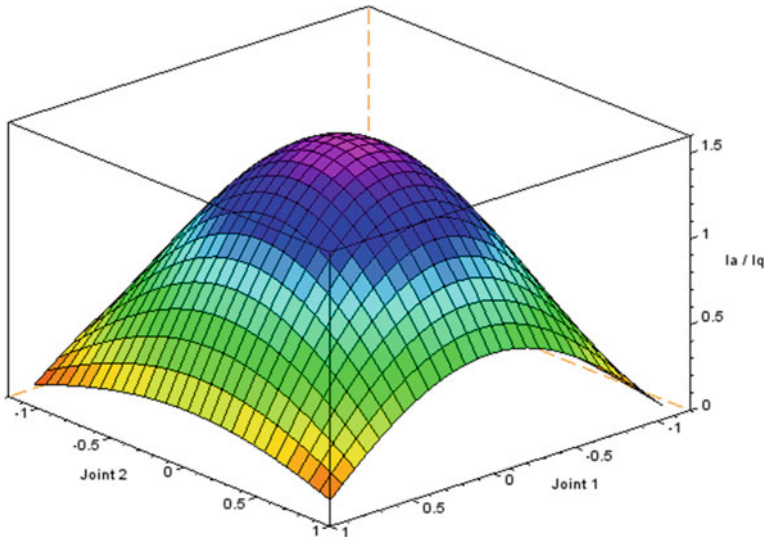


Fig. 4 Variation of moments of inertia respect to joint angle changes

$$\Delta Q = \frac{m_A(A_{CM} - Q_{CM})}{m_Q + m_A} \tag{10}$$

$$\Delta A = \frac{m_Q(Q_{CM} - A_{CM})}{m_Q + m_A} \tag{11}$$

The target coordinates of center of mass are obtained with trigonometry rules from vector system. Due to complicity of analytic form of the solution, we plotted result on the diagram. The result of calculation of CM changes relative to the quadcopter moment of inertia I_Q presented in Fig. 4.

The image is plotted relatively to quadrotor moment of inertia Iq. The shown effect in the Fig. 4 is similar for both joins, but Joint 1 has greater shift speed. Quadrotor Changes of the moment of inertia caused by a manipulating process usually processed as disturbances to the quadcopter model. In this paper, we tried to use gained rotation effects, caused by moving of center of mass, as an additional instrument of the control system.

3 Adaptive Controller for Center of Mass Displacement Manipulation

The most popular way to organize flight controller is to create and tune PID-controller to level control error oscillations [7–9]. In this experiment, we choose adapted PID controller. Applying of observed in Sect. 2 effects needs to add

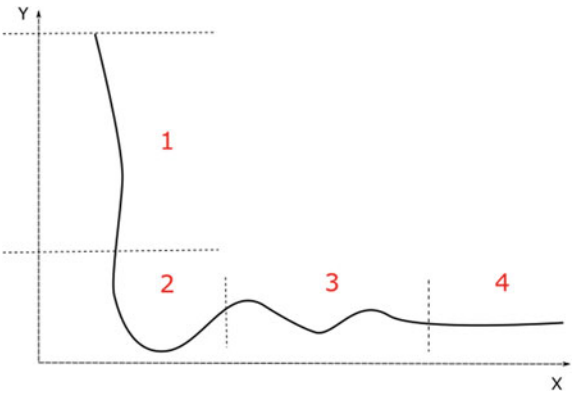


Fig. 5 The structure of quadcopter maneuver

additional input signal—the information about the kind of current maneuver. We used digital input signal, which indicates current movement state.

The model of maneuvers contains the chain of states:

1. Preparing state
2. Active maneuver state
3. Stabilization state
4. Post-maneuver state

The illustration of the states presented in Fig. 5. It shows rotation maneuver around Z-axis to 90 degrees.

The dynamics of moving in interval #2 includes centripetal force. In this case, we need to create an additional moment of rotation in roll direction to create an additional force in opposite to centripetal force direction. Implementing of this feature needs to the synthesis of control for servos model of the manipulator. To simulate this schema we added manipulator regulation to the controller layout scheme. The layout scheme of control system is presented in the Fig. 6.

The result matrix of F-τ values transfers to quadcopter model through transfer function. We choose the transfer function introduced in [1]. The tail model block uses current joints angles, maneuver state information and current dynamic characteristics from quadcopter to produce target transfer function for servomotors. Tail servo-dynamics block capturing gained transform to move manipulator to target configuration. Movement of the tail produces additional changes to the force and thrust matrix to be applied in the quadrotor model.

The focus of this paper is to show the ability to control the dynamic of quadrotor via manipulating of the "tail". Full dynamic model of the quadcopter and its rotors are not included into the scope of this paper. We implemented an experiment in Scilab environment to simulate quadcopter rotation maneuver. The trajectories of the flight are plotted in Fig. 7.

As it is shown on the image, applying of tail-assisted control gives faster maneuver state flight and lower amplitude of stabilization state. However, second active interval is followed by altitude reduction.

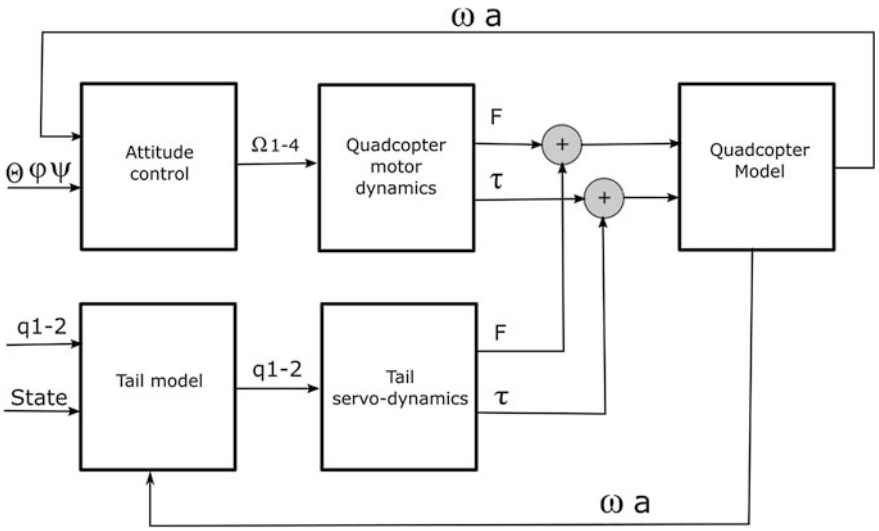


Fig. 6 Adapted controller layout scheme

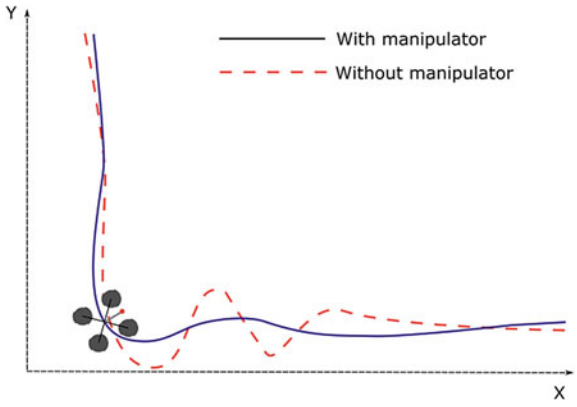


Fig. 7 The resulting trajectory of maneuver simulation

4 Conclusions

In this paper we presented a single-arm manipulating quadcopter model. The main goal of this research is to show the possibility of tail-assisted active flight correction. The introduced experiment illustrates kinematic and dynamic link of the manipulator model and quadrotor model. Introduced controller scheme can be extended by new maneuvers to support more flight situations. Simulation results confirmed the model and controller system.

In the future, we plan to test proposed results on read quadcopter model and create high-level control model for planning and recognition of maneuvers. In addition, we need to make some investigations in the field of control stability analysis.

Acknowledgments This work is supported in part by grant of the Ministry of Education and Science of The Russian Federation (project number #2.1917.2014K_2014) and grant of The Russian Foundation for Basic Research (grant number #16-07-00635A).

References

1. Korpela, C., Orsag, M., Pekala, M., Oh, P.: Dynamic stability of a mobile manipulating unmanned aerial vehicle. In: Proceedings—IEEE International Conference on Robotics and Automation, pp. 4922–4927 (2013)
2. Khalil, W.: Dynamic modeling of robots using recursive newton-euler techniques. In: 7th International Conference on Informatics in Control, Automation and Robotics, vol. 1, pp. 1–13 (2010)
3. McMillan, S., Orin, D.E., McGhee, R.B.: Efficient dynamic simulation of an underwater vehicle with a robotic manipulator. IEEE Trans. Syst. Man Cybern. **25**(8), 1194–1206 (1995)
4. Merchant, M.P., Miller, L.S.: Propeller Performance Measurement for Low Reynolds Number UAV Applications (2006)
5. Miskovic, N., Vukic, Z., Bibuli, M.: Marine vehicles' line following controller tuning through self-oscillation experiments. In: 17th Mediterranean Conference on Control and Automation, MED, pp. 916–921 (2009)
6. Ng, A.Y., Coates, A., Diel, M., Ganapathi, V., Schulte, J., Tse, B., Liang, E.: Autonomous inverted helicopter flight via reinforcement learning. In: Proceedings of International Symposium on Experimental Robotics, pp. 363–372 (2004)
7. Orsag, M., Korpela, C., Danko, T., Kobe, B., McNeil, C., Pisch, R., Oh, P.: Flight stability in aerial redundant manipulators. In: Proceedings—IEEE International Conference on Robotics and Automation, pp. 3529–3530 (2012)
8. Orsag, M., Korpela, C., Oh, P.: Modeling and control of mm-uav: Mobile manipulating unmanned aerial vehicle. J. Intell. Robot. Syst. Theory Appl. **69**(1–4), 227–240 (2013)
9. Trautmann, S., Albers, A., Howard, T., Nguyen, T.A., Frietsch, M., Sauter, C.: Semi-autonomous flying robot for physical interaction with environment. In: 2010 IEEE Conference on Robotics, Automation and Mechatronics, RAM 2010, pp. 441–446 (2010)

Intellectualization of Industrial Systems Based on the Synthesis of a Robotic Manipulator Control Using a Combined-Maximum Principle Method

Andrey Kostoglotov, Sergey Lazarenko, Zoya Lyashchenko and Igor Derabkin

Abstract Modern processes of intellectualization and industrial automation of systems based on the application of algorithms for controlling dynamic mechanical systems, which include robotic manipulators usually described as a two-section mathematical pendulum with sufficient accuracy. An important issue encountered during the development and implementation of algorithms for dynamic objects control is constituted by a challenge of ensuring the motion of such mechanical system according to a given law in a position of unstable equilibrium. This paper presents an option for a synthesis of optimal control in a position of unstable equilibrium with the use of a new approach known as a combined-maximum principle, which is expressed in a maximum value reached by a generalized power function during optimal controls. The dependencies that determine the hypersurfaces for switching control are established. Analytical solutions for the laws of control are obtained.

Keywords Synthesis · Optimal control · Pendulum · Manipulator · Combined-maximum principle

1 Introduction

Intellectualization of industrial robots is based on the use of an expert system that provides for selection of a control algorithm in the form of solving a problem of synthesis, which, in accordance with the set criterion, is most effective in the current situation. This requires the solution of famous problem of a need to choose

A. Kostoglotov (✉) · S. Lazarenko · Z. Lyashchenko · I. Derabkin
Rostov State Transport University, Rostov-on-Don, Russian Federation
e-mail: kostoglotov@icloud.com

Z. Lyashchenko
e-mail: hrustyaz@mail.ru

I. Derabkin
e-mail: i.deryabkin@jint.biz

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_37

feedback as dependence on the phase coordinates. Traditionally, for these purposes, Pontryagin's maximum principle is used to make it possible to obtain optimality conditions in the form of a boundary value problem [1]; at the same time, the control law synthesis requires a phase space structure to be deeply analyzed, what does not always lead to efficient solutions with regard to multijoint manipulators when using the approximation. The existing approaches to solving the problem of control synthesis, that are based on decomposition methods, use of the prescribed program trajectory and the construction of Lyapunov's function [2–4], require assumptions that can hardly be always tenable.

The use of the Hamilton-Ostrogradskii principle [5]—when searching for a minimum of objective functional-allows to get a direct solution to the problem of optimal control synthesis. The optimality conditions are presented in the form of a generalized power maximum principle [6–9]. Synthesis of optimal control is based on the analysis of the phase space structure. At the same time, analytical dependence of controls on the phase coordinates and communication with objective functional are established.

The object of the research held is a two-section mathematical pendulum. Equation of state of such dynamical system is a simplified non-linear model of mechanical two-section robotic manipulator with gearless drives and absolutely rigid assembly components. Such object is controlled through changing the torques in the device hinges.

The research is aimed to synthesize control of robotic manipulator based on the combined-maximum principle and to analyze the solution efficiency.

2 Formulation of the Problem

Motion of the system under control is subordinated to the principle of Hamilton-Ostrogradskii at a finite time interval $t \in [t_0, t_k]$ [5]

$$\begin{aligned} &\delta' R = \int_{t_0}^{t_k} (\delta T + \delta' A) dt = 0, \\ &t = t_0, \quad q(t_0) = q_0; \quad \dot{q}(t_0) = \dot{q}_0; \quad t = t_k, \quad q(t_k) = q_k; \quad \dot{q}(t_k) = \dot{q}_k; \end{aligned} \tag{1}$$

The objective functional is selected as a measure of quality of the process under control

$$J_1 = \int_{t_0}^{t_k} F(q, u, t) dt \to \min. \tag{2}$$

Herein $q \in [q_1, \ldots, q_n]$; $\dot{q} \in [\dot{q}_1, \ldots, \dot{q}_n]$—generalized coordinates and velocities; $F(q, u, t)$—positive-definite function; $T = \frac{1}{2} \sum_{s,k=1}^{n} a_{sk} \dot{q}_s \dot{q}_k$—kinetic energy; $\delta' A = \sum_{s=1}^{n} Q_s \delta q_s$—elementary work of generalized forces depending on the controls, $Q_s = Q_s(q, \dot{q}, u, t) = \sum_{m=1}^{M} u_{sm} \phi_{ms}(q, \dot{q}, t)$; $u = \{u_{sm}\} \in \bar{G}_u$, $s = \overline{1, n}$—controls selected from certain closed field $\bar{G}_u$, M—number of required parameters.

Lagrange equations of the second kind follow from principle (1)

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_s} - \frac{\partial T}{\partial q_s} = Q_s, \quad s = \overline{1, n}. \tag{3}$$

3 Combined Maximum Principle

Form the extended functional

$$J = \int_{t_0}^{t_k} [\lambda(T + A) + F] dt. \tag{4}$$

where λ—undefined Lagrange multiplier.

Theorem 1 *In order that generalized force $Q(q, \dot{q}, u, t) \subset \bar{G}_Q$ and trajectory corresponding to it $(q, \dot{q}) \in R^{2n}$ afford a minimum to the extended functional (4), it is necessary to meet the maximum conditions for generalized power [6–8]*

$$\Phi(q, \dot{q}, Q, \lambda) = \max_{Q \subset \bar{G}_Q} \sum_{s=1}^{n} [\lambda Q_s(q, \dot{q}, u) + V_s] \dot{q}_s \tag{5}$$

furthermore $\lambda = const > 0$, and at the ends of the trajectory $t = t_0$, $t = t_k$ transversality conditions are met

$$\lambda(A - T) + F = 0; \tag{6}$$

herein $V_s = \frac{\delta' F}{\delta q_s}$—fictitious force, which depends on the form of objective functional representation.

Theorem allows to determine the required generalized forces with an accuracy of negative sign synthesizing function $\mu_s(q_s, \dot{q}_s)$ [9]

$$Q_s = \lambda^{-1}[\mu_s(q_s, \dot{q}_s)\dot{q}_s - V_s], \qquad s = \overline{1,n}. \tag{7}$$

In the phase space of control switch hypersurface, generalized force is equal to zero [7], therefore the transversality conditions (6) are converted into a condition of constancy of generalized kinetic potential at a given time

$$L(q, \dot{q}, t) = \lambda T(q, \dot{q}) - F(q, \dot{q}) = l = const. \tag{8}$$

This equation is a hyperbolic paraboloid surface in the Lagrange variables phase space $q_s, \dot{q}_s (s = \overline{1,n})$.

Legendre transformation [5] of function $L(q, \dot{q}, t)$ via variables $\dot{q}_s (s = \overline{1,n})$ is a Hamilton function representing the ellipsoid surface in the variables of Hamilton $q_s, p_s (s = \overline{1,n})$

$$H(q, p, t) = \lambda T + F = \frac{1}{2}\lambda \sum_{s,k=1}^{n} \frac{A_{sk}}{D} p_s p_k + F = h = const; \tag{9}$$

at the same time, when transforming the value q, t play the role of parameters. Herein A_{sk}—cofactor of the element a_{sk} of the kinetic potential Hessian matrix.

The problem of constructing the synthesizing function $\mu_s(q, \dot{q})$ is to find the switching lines and the angular coefficients for it. The synthesizing function can be constructed based on research of Lagrange equations of the second kind at the switching lines by analogy with [8]. It is determined by formula

$$\hat{\mu}_s = -\lambda \left| \frac{dp_s}{dq_s} \right| = -\lambda \left| -\frac{\frac{\partial H}{\partial q_s}}{\frac{\partial H}{\partial p_s}} \right| = \frac{|V_s|}{\left| \sum\limits_{k=1}^{n} \frac{A_{sk}}{D} p_k \right|} = \frac{|V_s|}{|\dot{q}_s|}, \tag{10}$$

where function $\hat{\mu}_s$ is a modulus of the tangent to the phase trajectories $q_s(t)$ at the surface of switching $H(q, p, t)$ or $L(q, p, t)$ with a deformation factor λ. Depending on the initial conditions, constant-energy surfaces (8), (9) form a family which degenerates into point in case of $l_s \to 0, \quad h_s \to 0$. That's why any switching line which intersects said constant-energy surfaces must have its direction towards the focus. However, it is possible, if the modulus of tangent to the switching line μ_s the modulus of tangent to the constant-energy surface $\hat{\mu}_s$ are in the ratio

$$\hat{\mu}_s \mu_s = -\frac{1}{L_s}, \quad L_s > 0. \tag{11}$$

Then, according to the property of N. Luzin's measurable functions, the law of changing the optimal generalized force on the phase trajectory of the true motion shall look as follows:

$$Q_s = \lambda^{-1}\left[\frac{-|\dot{q}_s|\dot{q}_s}{L_s|V_s| + \varepsilon_s} - V_s\right], \quad s = \overline{1,n}, \tag{12}$$

where $\lambda, L_s, \varepsilon_s$—the constants determined from solution of control boundary value problem.

According to (11) the respective family of tangents to the switching lines

$$\frac{dp_s}{dq_s} = \frac{\dot{q}_s}{L_s V_s} = \frac{1}{L_s V_s}\sum_{k=1}^{n}\frac{A_{sk}}{D}p_k, \quad s = \overline{1,n}, \tag{13}$$

where the values p_s, q_s, V_s and derivatives are calculated along the switching line. In the simplest cases, the Eq. (13) is integrated. In case of $L = 2, \quad C = 0$—this is equation of parabola passing through the coordinate origin and corresponds to the problem of transferring the phase point to the coordinate origin. There are two possible solutions to the equations:

1. If the trajectory of the point coincides with the switching line

$$u = \lambda^{-1}\left[-\frac{|\dot{q}|\dot{q}}{\lambda^{-1}L|q| + \varepsilon} - q\right], \quad u \in \overline{G_u}; \tag{14}$$

2. If the trajectory of the point crosses the switching line

$$u = \lambda^{-1}\left[-\frac{k|\dot{q}|^{L-1}\dot{q}}{2\lambda^{-1}} - q\right], \quad u \in \overline{G_u}. \tag{15}$$

Equation (15) matches the solutions of L.S. Pontryagin and A.T. Fuller [1, 6].

4 Laws of Optimal Control in a Position of Unstable Equilibrium

Flat two-rod system is reviewed [5]. Angles φ_1, φ_2 of the OA and AB rods formed with the Ox axis are taken as generalized coordinates. Moments of inertia of the OA and AB bars in relation to the axes perpendicular to the Oxy plane and passing through the O and A hinges are designated as I_1 and I_2. Weights of the rods m_1 and m_2. Length of the rods l_1 and l_2. Distance from hinges to the center of gravity of the rods $s_1 = 0.5l_1$, $s_2 = 0.5l_2$.

Position of the system in case of $\varphi_1 = \varphi_2 = 0$—stable equilibrium; in case of $\varphi_1 = \pi + \alpha; \quad \varphi_2 = \pi + \beta, \quad \alpha > 0, \quad \beta > 0$—unstable equilibrium, if the angles $\alpha, \ \beta$ meet the condition of static stability of the system relative to the pole O

$$\sin\,\beta = -\frac{l_1}{l_2}\left(\frac{m_1}{m_2}+2\right)\sin\,\alpha. \tag{16}$$

Upon arbitrary angles $\alpha,\ \beta$ the state of the system is non-equilibrium.

The O hinge is attached the torque $u_1 = M_0$, the A hinge—the torque $u_2 = M_A$. Via said torques, the equilibrium and motion of the rod system is controlled.

The research objective: synthesize the laws of bounded controls $(u_1, u_2) \in \bar{G}_u$ enabling the system being subject to external uncontrollable exposures $y_1(t),\ y_2(t)$ to be transferred from an arbitrary initial position $t = 0, \quad \varphi_s(0), \quad \dot{\varphi}_s(0) \quad (s = \overline{1,2})$ into a position of unstable equilibrium $\varphi_1(t_k) = \pi + \alpha$; $\varphi_2(t_k) = \pi + \beta$, or any non-equilibrium position and be held in a state of static equilibrium for a prescribed time t_k, or manage the prescribed motion $x_1(t),\ x_2(t)$ relatively to such new equilibrium position. At the same time, the functional of assessing the control quality

$$J = 0.5\int\limits_0^{t_k}\left[\left[\varphi_1 - \pi - \alpha - x_1(t)\right]^2 + \left[\varphi_2 - \pi - \beta - x_2(t)\right]^2\right]dt \to \min \tag{17}$$

takes a minimum value.

For the solution of problem, the theorem of combined-maximum principle is applied.

Kinetic energy of the system

$$T = \frac{1}{2}\left[\left(I_1 + m_2 l_1^2\right)\dot{\varphi}_1^2 + l_1 l_2 m_2 \cos\left(\varphi_2 - \varphi_1\right)\dot{\varphi}_1\dot{\varphi}_2 + I_2\dot{\varphi}_2^2\right]. \tag{18}$$

Potential energy of the system

$$\Pi = g[l_1(0.5m_1 + m_2)(1 - \cos\,\varphi_1) + 0.5l_2 m_2(1 - \cos\,\varphi_2)]. \tag{19}$$

Elementary work of the controlling and external forces

$$\delta' A = u_1\delta\varphi_1 + u_2(t) + y_1(t)\delta\varphi_1 + y_2(t)\delta\varphi_2. \tag{20}$$

Expanded form of Lagrange equations of the second kind

$$\begin{aligned}
&\left(I_1 + m_2 l_1^2\right)\ddot{\varphi}_1 + 0.5m_2 l_1 l_2\cos(\varphi_2 - \varphi_1)\ddot{\varphi}_2 \\
&\quad - 0.5m_2 l_1 l_2\sin(\varphi_2 - \varphi_1)\dot{\varphi}_2^2 + g l_1(0.5m_1 + m_2)\sin\,\varphi_1 = u_1 - u_2 + y_1(t); \\
&0.5m_2 l_1 l_2\cos(\varphi_2 - \varphi_1)\ddot{\varphi}_1 + I_2\ddot{\varphi}_2 \\
&\quad + 0.5m_2 l_1 l_2\sin(\varphi_2 - \varphi_1)\ddot{\varphi}_1^2 + 0.5g l_2 m_2\sin\,\varphi_2 = u_2 + y_2(t).
\end{aligned} \tag{21}$$

In accordance with theorem of combined-maximum principle, the formula regarding the function of generalized power (5) shall be recorded as the maximum condition

$$\Phi(\varphi, \dot{\varphi}, u, \lambda) = \max_{u \in \bar{G}_u} \sum_{s=1}^{2} [[\lambda(u_1 - u_2) + (\varphi_1 - \pi - \alpha - x_1(t))]\dot{\varphi}_1 + [\lambda u_2 + (\varphi_2 - \pi - \beta - x_2(t))]\dot{\varphi}_2]. \tag{22}$$

Based on this condition the laws of controlling exposure have been obtained

$$U_1 = u_1 - u_2 = \lambda^{-1}\left[-\frac{|\dot{\varphi}_1|\dot{\varphi}_1}{L_1|\varphi_1 - \pi - \alpha - x_1(t)| + \varepsilon_1} - (\varphi_1 - \pi - \alpha - x_1(t))\right] = \lambda^{-1}a;$$
$$u_2 = \lambda^{-1}\left[-\frac{|\dot{\varphi}_2|\dot{\varphi}_2}{L_2|\varphi_2 - \pi - \beta - x_2(t)| + \varepsilon_2} - (\varphi_2 - \pi - \beta - x_2(t))\right] = \lambda^{-1}b. \tag{23}$$

The torque values are equal

$$M_0 = U_1 + u_2; \quad M_A = u_2. \tag{24}$$

The controls are selected from the range of piecewise continuous functions.

$$\bar{G}_u: \quad -[\lambda^{-1}a] \le (u_1 - u_2) \le [\lambda^{-1}a]; \quad -[\lambda^{-1}b] \le u_2 \le [\lambda^{-1}b], \tag{25}$$

where $[\lambda^{-1}a], [\lambda^{-1}b]$—admissible values of controls in the range of saturation.

In the internal points, the ranges of $\bar{G}_u$ control are computed according to formulas (23), and in the range of saturation they are taken to be equal to admissible values.

For the purposes of numerical modeling of the control process, the following system parameters are selected: $m_1 = 2$; $m_2 = 9$; $l_1 = 1$; $l_2 = 2$. Initial conditions: $t = 0$; $\varphi_1(0) = 0.75$; $\dot{\varphi}_1(0) = 0$; $\varphi_2(0) = -1.75$; $\dot{\varphi}_2(0) = 0$. Prescribed motion is modeled under harmonic law: $x_1(t) = X_1 \sin t$; $x_2(t) = X_2 \sin t$. External uncontrollable exposure is modeled by rectangular pulses $y_1(t) = Y_1 h[(t-5)(7-t)]$; $y_2(t) = Y_2 h[(t-15)(17-t)]$, where $h(t)$—Heaviside step function. The angles that determine a broken system configuration are determined by the condition (23) $\alpha = 0.785$; $\beta = -0.904$. The modeling results are presented in Figs. 1 and 2. The figures show the structure of optimal control and phase portrait of the system's motion.

Figure 1 gives solutions of the problem of the system transfer into a position of unstable equilibrium $\varphi_1(t_k) = \pi + \alpha$; $\varphi_2(t_k) = \pi + \beta$. Prescribed motion $x_1(t) = x_2(t) = 0$; external exposure $y_1(t) = y_2(t) = 0$. Admissible controls $[\lambda^{-1}a] = 650$; $[\lambda^{-1}b] = 650$.

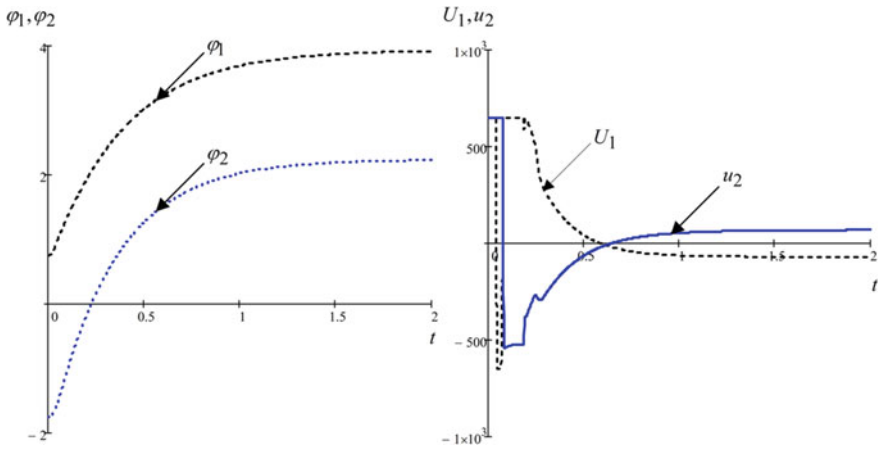


Fig. 1 The modeling results

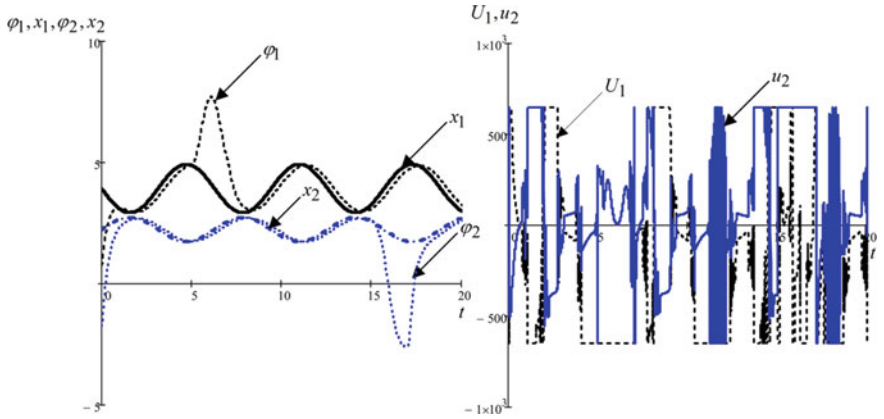


Fig. 2 The modeling results

Figure 2 gives solutions of the problem of the system transfer into a position of unstable equilibrium $\varphi_1(t_k) = \pi + \alpha$; $\varphi_2(t_k) = \pi + \beta$, of the controlling motion under the prescribed law $x_1(t) = -\sin t$; $x_2(t) = 0.5 \sin t$ and in case of external load exposure $y_1(t) = 650h[(t-5)(7-t)]$; $y_2(t) = -650h[(t-15)(17-t)]$. Admissible controls $[\lambda^{-1}a] = 650$; $[\lambda^{-1}b] = 650$.

The results analysis suggests: structure of control laws is different and in a state of unstable equilibrium the shape of control pulses is close to rectangular. The non-equilibrium state triggers the emergence of a series of pulses of different duration; the control is stable even in the presence of perturbations of large amplitude.

5 Conclusion

The method of optimal control synthesis—combined-maximum principle—has the versatility and simplicity what demonstrates the effectiveness of the practical application of the algorithms synthesized on its basis in the problems of automation and intellectualization of technological processes. Synthesized controls do not require a lot of computational costs, have differences as compared with traditional ones that are expressed in the simple nonlinear structure of feedback, which is that in particular cases lead to certain solutions.

Acknowledgments The paper has been supported of RFBR grants No. 15-08-03798, 15-38-20835, 16-37-60034, 16-38-00665.

References

1. Pontryagin, L.S.: Matematicheskaya Teoriya Optimal'nykh Protsessov (Mathematical Theory of Optimal Processes). Nauka, Moscow (1976)
2. Naumov, G.E.: Postroenie krivoj perekljuchenija dlja zadach optimal'nogo upravlenija s uchashhajushhimisja perekljuchenijami (switching curve construction for the tasks of optimal control with switching becoming more frequent). Izv. RAN. Teorija i sistemy upravlenija. **3**, 46–51 (2003)
3. Pyatnitskiy, E.S.: Princip dekompozicii v upravlenii mehanicheskimi sistemami (decomposition principle in managing the mechanical systems). Dokl. Akad. Nauk SSSR **300**, 300–303 (1988)
4. Ananievskiy, I.M.: Nepreryvnoe upravlenie po obratnoj svjazi vozmushhennymi mehanicheskimi sistemami (A continuous control on a reverse connection by disturbed mechanical systems). Prikladnaja matematika i mehanika. **67**, 163–178 (2003)
5. Lur'e, A.I.: Analiticheskaya Mekhanika (Analytical mechanics). Gos. Izd. Fiz.-Matem. Liter., Moscow (1961)
6. Kostoglotov, A.A.: Solution of fuller's problem on the basis of the joint pontryagin-hamilton-ostrogradskii principle. Autom. Control Comput. Sci. **41**, 179–187 (2007)
7. Kostoglotov, A.A., Kostoglotov, A.I., Lazarenko, S.V.: Joint maximum principle in the problem of synthesizing an optimal control of nonlinear systems. Autom. Control Compt. Sci. **41**, 274–281 (2007)
8. Kostoglotov, A.A.: Method for synthesis of optimal attitude stabilization algorithm based on joint maximum principle. Autom. Control Comput. Sci. **36**, 21–28 (2002)
9. Kostoglotov, A.A., Kostoglotov, A.I., Lazarenko, S.V.: The combined-maximum principle in problems of estimating the motion parameters of a maneuvering aircraft. J. Commun. Technol. Electron. **54**, 431–438 (2009)

System Reconfiguration Using Multiagent Cooperative Principles

Eduard Melnik, Vladimir Korobkin and Anna Klimenko

Abstract The paper is devoted to the problem of fault-tolerant informational and control systems (ICSs) design. The class of ICSs for the hazardous objects with the criticality of error is considered. An approach to the fault-tolerant ICS design is based on performance redundancy and decentralized monitoring and control. It increases the fault-tolerant possibilities of the system, but requires the design and development of new models, methods and algorithms of system recovery. A new model of the reconfiguration problem and the problem solving algorithm based on cooperative principles of multiagent systems are given. Also some simulation results are represented and discussed briefly.

Keywords Reconfiguration · Transport · Power plant · Hazardous objects · Informational and control system · Simulated annealing · Cooperation · Cooperative search

1 Introduction

Informational and control systems (ICSs) are used in various complex objects.

Special ICS class, operating on most hazardous systems can be defined. Failure of such an ICS can lead to environmental disaster or numerous casualties.

Hazardous systems primarily include power systems (power plants, including nuclear power plants), transport (civil and military aviation), objects of space industry [1]. These systems have high requirements of reliability and fault tolerance. Since ICS is an integral component of these systems, the same requirements are applied to the ICS.

E. Melnik
OOO "Neirosetevye Technologii", Taganrog, Russia
e-mail: Evm_17@mail.ru

V. Korobkin · A. Klimenko (✉)
SFedU Acad. Kalyaev Scientific Research Institute of Multiprocessor Computer Systems, Taganrog, Russia
e-mail: Anna_klimenko@mail.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_38

Building resilient ICS is an actual contemporary scientific problem.

A new approach to the design of fault-tolerant ICS is described in [1–3]. This approach is based on the principles of performance redundancy instead of the structural one and decentralization of ICS architecture.

The system failure is compensated by the reconfiguration, which, broadly speaking, is the process of re-executing the failed control tasks on the operational computational units (CUs) instead of failed ones. As the system is decentralized, the monitoring and control are implemented by the equal software agents. Every agent controls its own CU exchanging messages with other agents. When one agent does not send the "presence" sign to the communication network (CN), other agents begin the system reconfiguration, making decision about the new configuration in a cooperative manner. An agent is represented like an additional control task running on the CU.

This paper contains:

- a new model of the reconfiguration problem for the ICS with the performance redundancy and decentralized monitoring and control;
- a new multiagent cooperative algorithm, which obtains the problem solution in an acceptable time frame;
- some simulation results representing the effect of the algorithm developed on the solution quality.

The Sect. 2 of the paper contains a brief review of the redundancy and monitoring and control methods. Section 3 is devoted to the problem formalization with the respect to the reconfiguration issues. The cooperative algorithm of the problem solving is given in Sect. 4. Section 5 contains simulation results discussed shortly.

2 Performance Redundancy and Decentralized Monitoring and Control

The basis of the approach to fault tolerant ICS design is the performance redundancy and decentralized architecture. The CUs are represented by the software agents, which are the equal elements of the community. This approach is free of the disadvantages of structural redundancy, since there are no unused nodes in the system and it is possible to balance the load of CUs [3, 4].

The load balancing, in turn, increases the no-failure operation possibility [2–4].

Classifying ICSs on the basis of system monitoring and control, the following basic classes can be distinguished:

- Centralized ICSs, which uses a central controller (manager) to manage the operation of ICS;
- Hierarchical ICSs, where the system performs like a tree structure, and a dedicated controller manages "his" personal branch;
- Decentralized ICSs, where the system consists of equal CUs.

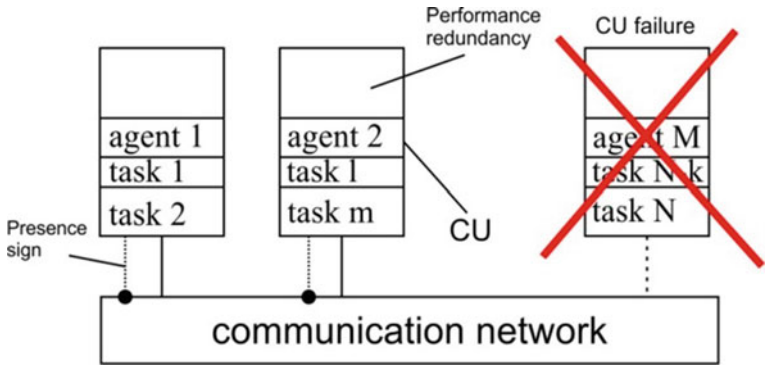


Fig. 1 Performance redundancy and representing the CUs by software agents

Comparing the ICS classes with dedicated monitoring and control elements and decentralized one, ICS with decentralized architecture provides the best performance in terms of reliability and fault tolerance [1], introducing, at the same time, the need to design and develop new methods and tools for equal elements collaboration.

Decentralized ICS is represented by the multiagent system [1]. It can be said, that ICS is managed by the community of equal, peer-to-peer software agents (Fig. 1).

While performance redundancy assumes no dedicated reserve elements, subtasks are performed on the failed node can be relocated to the operational nodes by the reconfiguration procedure.

Subtasks from the failed node can be performed by more than one operational CU, depending on the performance reserve availability.

Management subtasks are conditionally divided to critical (to ensure the basic functionality of the system) and non-critical. Let's assume that some of the non-critical subtasks can be eliminated from the system during the reconfiguration procedure.

3 Reconfiguration Problem Formalizing

In the context of this paper the first objective function of the problem is load balancing of CUs. The importance of this criterion is proved in [2–4].

Let the input data be the following:

1. A set of control subtasks $G = \{x_i\}$, $i = 1...N$, where x_i—the size of subtask i, N—the number of subtasks.
   Let $G = G_c \cup C_{nc}$, $G_c \cap G_{nc} = \emptyset$,where G_c—a subset of critical subtasks, G_{nc}—a subset of non-critical subtasks, which can be eliminated from the system. The number of critical tasks is N_c, and the number of non-critical tasks N_{nc};

2. A planned completion time for the set G T_{plan}.
3. Number of CUs is M with the performance p.

Let the resource allocated by CU j for the subtask i be λ_{ij}.
Subtasks distribution will be described by the matrix R.

$$R = \begin{vmatrix} r_{11} & r_{12} & r_{1M} \\ \dots & \dots & \dots \\ r_{N1} & \dots & r_{NM} \end{vmatrix} \tag{1}$$

where

$$r_{ij} = f(\frac{x_i}{\lambda_{ij}p})g(x_i); \tag{2}$$

$$f\left(\frac{x_i}{\lambda_{ij}p}\right) = \begin{cases} \frac{x_i}{\lambda_{ij}p}, & \textit{if } x_i \textit{ is running on CU } j, \\ 0, & \textit{otherwise}. \end{cases} \tag{3}$$

$$g(x_i) = \begin{cases} 0, & \textit{if } x_i \in G_{nc} \textit{ and eliminated from the system}, \\ 1, & \textit{otherwise}. \end{cases} \tag{4}$$

Herewith the constraint described below must be satisfied:

$$\forall j : \sum_{i=1}^{N} r_{ij} \leq T_{plan}, j \in [1 \dots M] \tag{5}$$

Eventually, the first objective function will be as following with the constraint (5):

$$\forall k, l : \underset{R}{MIN}(\sum_{i=1}^{N} \lambda_{ik} - \sum_{i=1}^{N} \lambda_{il}) \tag{6}$$

In other words, we need to find subtasks allocation with respect to load balancing objective function.

One more objective function must be considered. It is related to the desirable option to keep running as much non-critical tasks as possible. It is interconnected with the system vitality term: a number of non-critical subtasks can be eliminated, but obviously, it will be better to keep them running.

While the desirable option is to keep running as much subtasks as possible, the second criteria should be formed. The maximum number of non-critical tasks running equals to the maximum summa of all g(xi) in the matrix R.

The second criteria will be as following:

$$\underset{R}{MAX} \sum_{i}^{N} g(x_i), \tag{7}$$

Let's formalize the reconfiguration problem objective functions and constraint.

$$\forall k, l : \underset{R}{MIN}(\sum_{i=1}^{N} \lambda_{ik} - \sum_{i=1}^{N} \lambda_{il}) \tag{8}$$

$$\underset{R}{MAX} \sum_{i}^{N} g(x_i),$$

$$\forall j : \sum_{i=1}^{N} r_{ij} \leq T_{plan}, j \in [1 \ldots M].$$

So, we have a problem with two criteria to solve in the circumstances of deficiency of time. The next section of this paper contains the multiagent method of the problem solving on the parallel simulated annealing basis.

4 Multiagent Reconfiguration Algorithm

Let's form the general scalar objective function. We assume both objective functions equivalent in the term of importance. The Eq. (6) is transformed as following:

$$\forall k, l : \underset{R}{MAX}(-(\sum_{i=1}^{N} \lambda_{ik} - \sum_{i=1}^{N} \lambda_{il})) \tag{9}$$

Θ the number is big enough.

After scalarization procedure the scalar objective function will be as following:

$$F = 0,5(\sum_{i,}^{N} x_i) + 0,5(\Theta - (\sum_{i=1}^{N} \lambda_{ik} - \sum_{i=1}^{N} \lambda_{il})) \rightarrow \underset{R}{MAX}. \tag{10}$$

With respect to the satisfaction of constraint (5).

In the scope of this research the simulated annealing (SA) with the "quenching" temperature was chosen as a method of solution obtaining [5, 6].

SA method is iterative and serial, and difficult for parallelizing. Nevertheless, independent parallel launches with different points of initialization, various laws of cooling, etc. are well-known ways to speed-up the algorithm convergence [7–9]. Within such strategies of speed-up the global optimum can be never reached, but the solution is obtained in the reasonable time.

With the shortage of time and the decentralized ICS architecture using, it is appropriate to initiate a search for a new configuration at all operational nodes (which are represented the agents).

The idea of multiagent cooperation begins with works [10–12]. The results of studies mentioned are show the advantages of cooperative problem solving and cooperative constraint satisfaction problem solving too.

So, the following scenario makes sense: as soon as one of the agents finds allowable configuration, solving, in fact, a constraint satisfaction problem, it notifies other agents, which take the solution found as a new configuration proposed to perform.

Generalized multiagent algorithm based on a parallel simulated annealing will be as following:

1. *Set the initial parameters: temperature, quenching ratio, etc*
2. *Generate solution R. g(xi) is generated in a random manner*
3. *beginning of the cycle*
   - 3.1 *If there is a signal of the end of the calculation*
     - 3.1.1 *send to the communicational network the current solution*
     - 3.1.2 *Range the solutions in the communicational network and choose the best one from the point of view of the value F*
     - 3.1.3 *Go to step 4*
   - 3.2 *Generation of new solutions: R*
   - 3.3 *Calculate the value of F*
   - 3.4 *Check to limit the admissibility of F*
   - 3.5 *If the current solution is acceptable, put into the communication environment the solution and a signal about the end of the calculation*
     - 3.5.1 *Choose the best solution from the communication environment*
     - 3.5.2 *Go to step 4*
   - 3.6 *Temperature correction. Go to step 3*
4. *Adopt the selected configuration to run*
5. *End*

5 Experimental Results

We conducted a simulation for the following initial data: | G | = 50 (subtasks) with random xi. The values of the number of CUs = 5, 10, 15. Assessment opportunities for local minima of the objective function was produced in the first pilot study for a

variety of quenching ratios: 0.6; 0.7; 0.8; 0.9. The experimental results suggest that the average use of simulated annealing with low values of the quenching ratio makes possible to obtain a local minimum for a time of about 20–30 iterations (Fig. 2). It should be noted that the results improvement is expected with the usage of multiple peer agents. The first agent which is found a permissible solution interrupts the calculations of other agents.

For the further simulation we choose combinations of the 5, 10, and 15 software agents. Number of runs of the algorithm for each situation = 20. The results are shown in the Figs. 3, 4, 5.

The results of the first experimental study is on the Fig. 3. The lower border of the simulated annealing iterations number is approximately 18, the best objective function value is approximately 100.

With the software agents number growth (Fig. 4) the best objective function value becomes better (88), and the number of objective function calls for the "fastest" agent reduces to 12.

With the further agent number growth (Fig. 5) number of iterations of the algorithm keeps the tendency to reduce with the objective function value freezing (because minimas found can't reduce more).

The simulation results confirm the trend of improving both the quality and speed of the solutions obtained. It should be noted that this trend is expected, because with the start points increasing, initiating a random search increases the likelihood of getting into the most "successful" point.

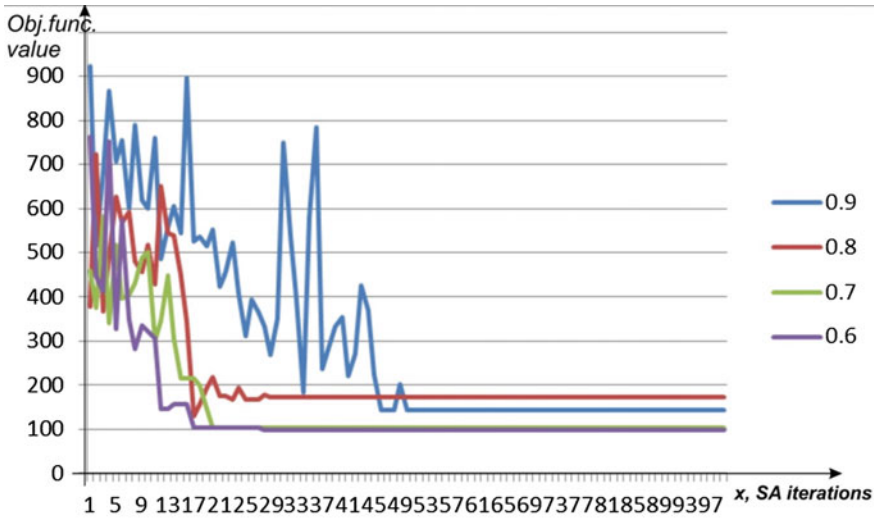


Fig. 2 Simulated annealing with quenching ratios: 0.6, 0.7, 0.8, 0.9

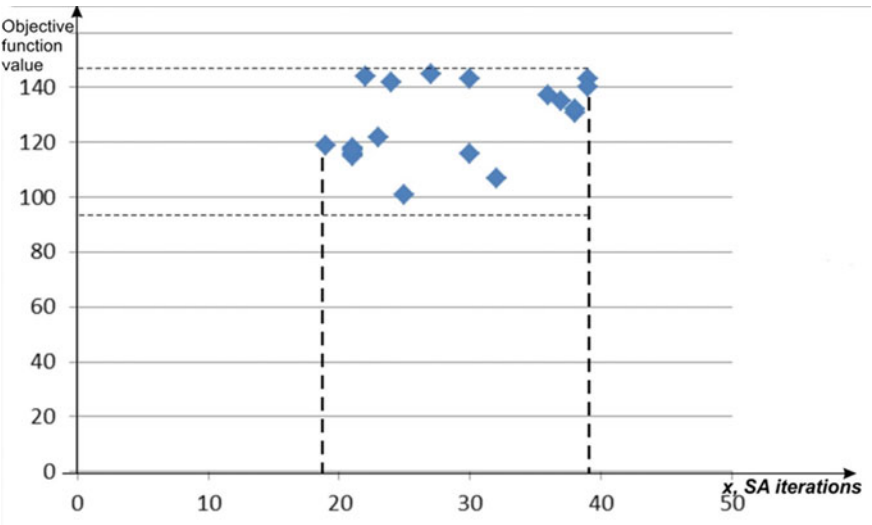


Fig. 3 Simulation results for 5 software agents

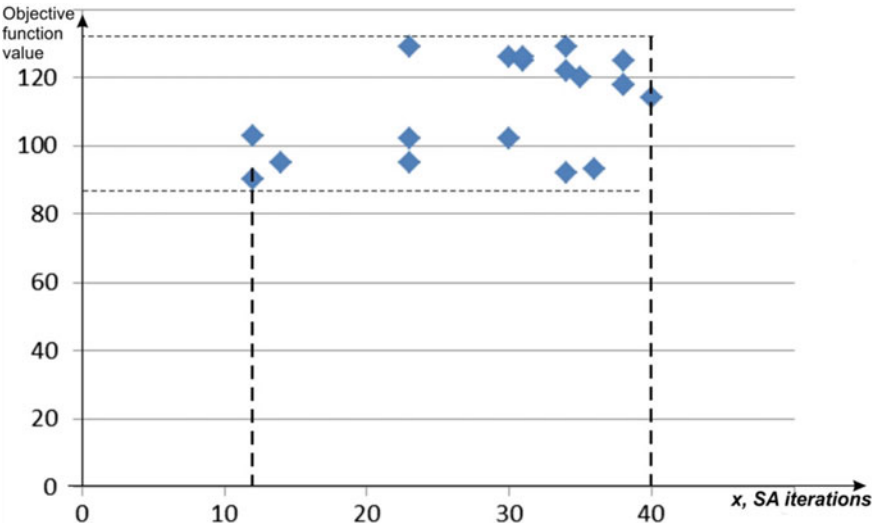


Fig. 4 Simulation results for 10 software agents

It should also be noted that the application of the cooperative principles of the multiagent system during reconfiguration, promotes a significant reduction of search time (in the best cases—up to 10 iterations). With the increasing of the agent

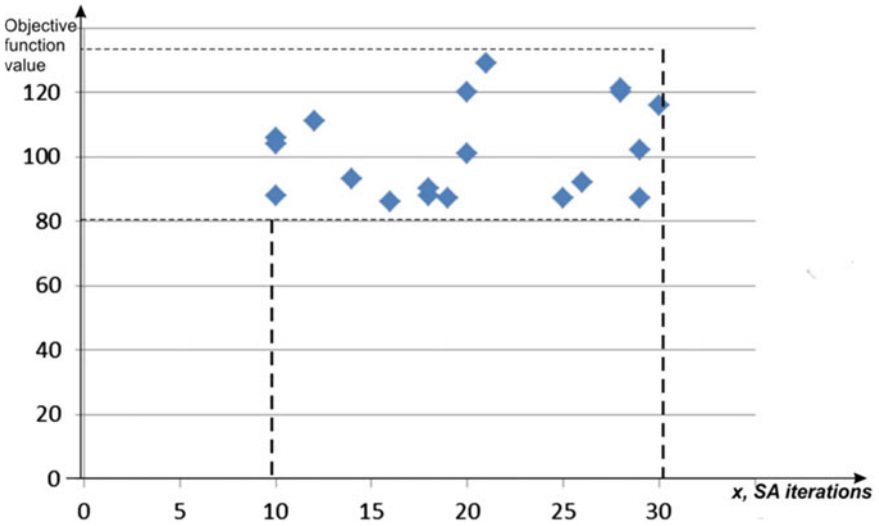


Fig. 5 Simulation results for 15 software agents

number the improvements of the results are expected, at least in terms of rate of solution obtaining. However, it does not taken into account the communication overhead in the context of this article. For this reason, one of the future research directions is to obtain estimates of the optimal number of agents with respect to the communication overhead.

6 Conclusions and Future Work

ICSs with the performance redundancy and decentralized monitoring and control are the prospective branch of the area. Such systems are considered to be fault-tolerant because of system reconfiguration possibilities, but the reconfiguration problem iscomplex and requires additional methods, models and algorithms research and development.

In the scope of this paper the effect of co-operative principles of multi-agent environment in relation to the problem of ICS with a performance redundancy and decentralized monitoring and control reconfiguration was investigated. Within this paper a new problem model was designed, with respect to the chosen ICS class problem. Also the multiagent algorithm was synthesized.

The simulation of some quantitative compositions of multi-agent systems is done. The simulation results confirm the trend of improving the speed of solution obtaining with a usage of cooperative principles of multiagent society. The

cooperative problem solving allows to improve speed and quality of the solutions obtained. As a further areas of work it is planned to assess the optimal number of agents, taking into account the communication costs.

Acknowledgments This work was supported by the RFBR.

References

1. Kalyaev, I.A. Melnik, E.V. Decentralized Systems of Computer Control. Rostov n/D: Publishing SSC RAS, p. 196 (2011)
2. Melnik, E.V.: Principles of organization of the decentralized network-centric information management systems. In: Melnik, E.V., Ivanov, D.Ya. (eds.) Herald of Computer and Information Technologies, №4, pp. 25–30(2013)
3. Melnik, E.V.: Effect processor computational load balancing devices in highly distributed information management system. In: Gorelova, G.V., Melnik, E.V. (eds.). Mechatronics, Automation, Control, p. 29–35 (2012)
4. Melnik, E.V.: Simulation options for redundancy in distributed information and control systems with a decentralized organization. In: Melnik, E.V., Gorelova, G.V. (eds.) Proceedings of SFU. SER Technical science, №3, pp. 184–193 (2013)
5. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220** (4598), 671–680 (1983)
6. Ingber, L.: Simulated annealing: practice versus theory (1993). http://citeseer.uark.edu:8080/citeseerx/viewdoc/summary?doi=10.1.1.15.1046
7. Azencott, R. (ed.): Simulated Annealing: Parallelization Techniques, pp. 47–79. Wiley, New York (1992)
8. Durand, M.D., White, S.R.: Trading accuracy for speed in parallel simulated annealing with simultaneous moves. Parallel Comput. **26**, 135–150 (2000)
9. Crainic, T.G., Toulouse, M.: Parallel metaheuristics. In: Crainic, T.G., Laporte, G. (eds), Fleet Management and Logistics, pp. 205–251, Kluwer Academic Publishers (1998)
10. Crainic T.G., Toulouse M., Gendreau, M.: Communication issues in designing cooperative multi thread parallel searches. In: Osman, I.H., Kelly, J.P. (eds.) Meta-Heuristics: Theory & Applications, pp. 501–522, Kluwer Academic Publishers (1996)
11. Clearwater, S.H., Hogg, T., Hubermann, B.A.: Cooperative problem solving. In: Hubermann, B.A. (ed) Computation: The Micro and The Macro View, pp. 33–70. World Scientific, (1992)
12. Clearwater, S.H., Hubermann, B.A., Hogg, T.: Cooperative solution of constraint satisfaction problems. Science **254**:1181–1183 (1991)

Intelligent Agent System to Analysis Manufacturing Process Models

Alexander Afanasyev and Nikolay Voit

Abstract Business process and workflows represent diagrams as Business Process Model Language (BPML), Unified Model Language (UML), which may contain errors made by a designer. In order to check that errors, we developed a new grammar (called RV-grammar) that allows us to increase quality of business process management and workflow systems by 20 % oppositely with a reserved graph grammar, a positional grammar, and a relational grammar. Thus, we propose an intelligent agent system based on RV-grammar to analysis business-process models for manufacturing. The system has linear time in order to facilitate the automatically identification of a variety of syntax and semantical errors in workflow models, which enables the analysis of 4 additional types of errors.

Keywords Automaton grammars · Visual languages · Agent approach

1 Introduction

Manufacturing sector interests application areas since 1980, and at that time the research area was summarized by the term Distributed Artificial Intelligence (DAI). The first DAI book by Huhns [1], the work by Parunak [2], Fischer and colleagues [3] and witnessing commercialization driven by companies such as Whitestein Technologies and Magenta Technologies describe manufacturing as a major application domain. Indeed, references collected for this paper over the whole period which we used as a cut of year for collecting data. Manufacturing covers the supply chain of production. Therefore, aspects such as virtual enterprises, business process management and workflow systems, and information systems/agents are prominent in this vertical sector.

A. Afanasyev (✉) · N. Voit
Ulyanovsk State Technical University, Ulyanovsk, Russia
e-mail: a.afanasev@ulstu.ru

N. Voit
e-mail: n.voit@ulstu.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_39

Whitestein Technologies and Magenta Technologies note that first-generation static workflow product lifecycle management systems no longer fit the needs of many companies, the approach of standardizing process with first-generation static workflow automation tools has reached its limits, and the consequences are ill-fitting processes and soaring process development and improvement costs [4].

Industrial deployment of agent-based systems covers business process management of manufacturing as well as workflow systems. We use the agent architecture, called Distributed Control Architecture (DCA), for instance in [5], but we replace its components with agent RV-grammar, agent diagram model, agent planner, agent execution control, and agent diagnostic. Agent RV-grammar creates a grammar to describe a diagram. Agent diagram model contains a XML description of a diagram. Agent planner is a major agent to manage the intelligent agent system. Agent execution control performs actions getting from agent planner. Agent diagnostic checks a diagram for errors. We use the Foundation for Intelligent Physical Agents (FIPA) to develop the standard implementation of the open, heterogeneous and interoperable agents, which consists of Announcement Protocols, Ask-about Protocols, Task/Action Agreement Protocols that are described in the paper [6].

There is the modern theory of graphical and visual languages to represent the diagrams, that consists of syntax models as spatial and logical, including the attributes of graphical objects (for example, a rectangle, a circle), and link types.

The spatial model has relative or absolute coordinates of a graphical object. The use of the spatial model is difficulty to control and analysis a structure, addressing to topology (syntax), and the attributes of diagrams.

As a rule, the logical model is used to describe the syntax of diagrams, which has the basis of the graphical grammar.

John L. Pfaltz and Azriel Rosenfeld proposed a web-grammar [7]. Zhang and Costagliola [8] developed the positional graphical grammar [7], relating to the context-free grammar. Wittenberg and Weitzman [9] developed a relational graphical grammar. Zhang and Orgun [10, 11] described preserving graphical grammar in their papers.

The mentioned graphical grammar has the following shortcomings.

Most tools, supporting computer-aided design (IBM Rational Unified Process, ARIS [12]), has a direct method of an analysis that requires a lot of iterations to check diagrams for errors. However, they cannot detect errors, called separate context, related to the use of logical links as ‘AND’, ‘OR’, ‘XOR’, etc. The context-sensitive semantic errors may be into a text (a notation) of any diagram, which have not been fund with the tools, and become “expensive” errors in a design.

Business process and workflows can be presented in IBM Rational Software Architecture [12], MS Visio [13], Dia [14], ARIS with the help BPML [4], UML, and represented by diagrams, including errors that made by a designer. Thus, to check the diagrams for the errors we create the intelligent agent system that consists of the agents mentioned above.

In section Grammars models, we describe grammars, and a new grammar RV-grammar [15]. For instance, we take UML diagram to represent business process, although RV-grammar can check and fix errors into BMPL, IDEF0 [16],

IDEF3, Petri net diagrams. Section Checking diagrams consists of authors methods to analysis the diagram for errors. Intelligent agent system is suggested by authors in section Intelligent agent system. In Results, we assess efficiency of applying the intelligent agent system.

2 Grammar Models

2.1 Review of Grammars

We review a reserved graph grammar, a positional grammar, a relational grammar, and a web grammar to represent business process and workflow as diagrams with the help the tools.

1. A positional grammar: uses a plex-structure, which has not attaching points, and does not consider dynamically links of inputs\outputs, therefore, it cannot be used for graphical languages by a designer.
2. A relational grammar: generates a non-exhaustive list of errors that cannot be defined to analysis.
3. A reserved graph grammar, and a web grammar: save sentential form of a diagram, that takes a lot of time to analysis it.
4. For reviewed grammar, we summarize problems as:
5. Increasing rules to describe the grammars takes much time for an analysis, and has exponential or polynomial requires of time.
6. Reviewed grammars has a sentential form to represent a diagram.
7. There are not tools for the semantic analysis of diagram attributes.

2.2 RV-Grammar

The authors have developed automaton grammar, called RV-grammar, to analysis and check (control) diagrams for the tools, described in the papers [17–19].

RV-grammar has a basis of the L (R) language grammar [19], which can be written as:

$$G = (V, \Sigma, \widetilde{\Sigma}, \overline{\Sigma}, R, r_0) \tag{1}$$

where $V = \{v_l, l = \overline{1, L}\}$ is an auxiliary alphabet (the alphabetical operations with the internal memory); $\Sigma = \{a_t, t = \overline{1, T}\}$ is a terminal alphabet, which is the union of its graphic objects and links (the set of primitives); $\widetilde{\Sigma} = \{\tilde{a}_t, t = \overline{1, \widetilde{T}}\}$ is a quasi-terminal alphabet, extending the terminal alphabet. The alphabet includes:

- quasi-terms of graphic objects,
- quasi-terms of graphic objects with more than one input,
- quasi-terms of links, marked with the specific semantic,
- quasi-term for the end of an analysis,
- $R = \{r_i, i = \overline{1,I}\}$ is the scheme of the grammar (a set of rules, where each complex r_i consists a subset P_{ij} of rules, where $r_i = \{P_{ij}, j = \overline{1,J}\}$),
- $r_0 \in R$ is an axiom of RV-grammar (the initial complex of rules), $r_k \in R$ is a final complex of rules.

The complex of rules $P_{ij} \in r_i$ is given as:

$$\tilde{a}_t \overset{W_\gamma(\gamma_1,\ldots,\gamma_n)}{\longrightarrow} r_m \tag{2}$$

where $W_\gamma(\gamma_1, \ldots, \gamma_n)$—n-ary relation, defining an operation with the internal memory depending on $\gamma \in \{0, 1, 2, 3\}$;

$\Omega\mu$ is a modification operator, changing the type of the operation with the memory, and $\mu\{0, 1, 2\}$;

$r_m \in R$ is the receiver of rules.

The internal memory is presented by a stack for processing the graphic objects that have more than one output to save the information of link-marks, and elastic tapes for processing the graphic objects that have more than one input to mark the number of returns to a given vertex, and hence the number of incoming links. It should be noted that the elastic tape reads data from cells of the internal memory without a content destruction, and the cells of elastic tapes operates on data as a counter defined on positive integers.

The chain of $\varphi = \alpha_{t1}, \alpha_{t2}, \ldots, \alpha_{t\lambda}$ is called RV-output $\alpha_{t\lambda}$ of α_{t1} and it is denoted $\alpha_{t1} \overset{RV}{\Rightarrow} \alpha_{t\lambda}$ if for any $\xi < \lambda$ and $r_e \in R$ are as $\alpha_{t\xi+1} \in r_e \left(a_t \overset{\Omega_\mu[\lambda_1,\ldots,\lambda_n]}{\longrightarrow} r_e \right) \in r_i$.

RV-output is considered to be complete (it is denoted as $\alpha_{t1} \overset{RV}{\Rightarrow} \alpha_{t\lambda}$), if $a_{t\lambda}$ is generated by rules with r_k on the right-hand side.

RV-grammar is effective both for generating and recognizing.

The application of any complex of rules r_0 (RV-grammar axiom) generates some chain of language L on its RV-grammar. The complex of rules determines both the initial symbol of generated chain, the operation on the internal memory, and also the name of receiver of rules. The generation is completed using a complex of rules with r_k on the right-hand side.

The recognition of the chain runs verifying the first symbol using rules r_0, while next symbol appearing, and the last symbol of the chain must belong to a complex of rules with r_k on the left-hand side.

The use of rules is accompanied with appropriate operations on the internal memory. The internal memory is empty at the start, and at the end of these processes, the memory contains operations of rules with r_k on the right-hand side.

3 Intelligent Agent System

In the previous section we describe mathematical models RV-grammar that is a core of our intelligent agent system to check for errors of diagrams. In this section, we review papers about industrial deployment of multi agent system (MAS) [20] to build our intelligent agent system. MAS is used in following domains usually: Manufacturing, Logistics, Air Traffic, Space Exploration, Training, Automotive, Supply-chains, and Networking. We focus on Manufacturing for business process models and workflow, and also review concepts for building MAS as: BDI, negotiation, simulation, interoperability, coordination, distributed control and diagnostics [3], therefore, we take a concept for the distributed control and diagnostics. Manufacturing agent systems have following functional: control and diagnostics. For instance, in [20], we review DCA architecture as an example, and create our architecture to control and diagnose diagrams, that depicts in Fig. 1.

Design and reengineering of application systems: develop business process model and workflow to describe process for making manufacturing (industrial) equipment.

MS Visio, IBM Rational Software Architecture, Dia, ARIS: help to represent business process model and workflow as diagrams.

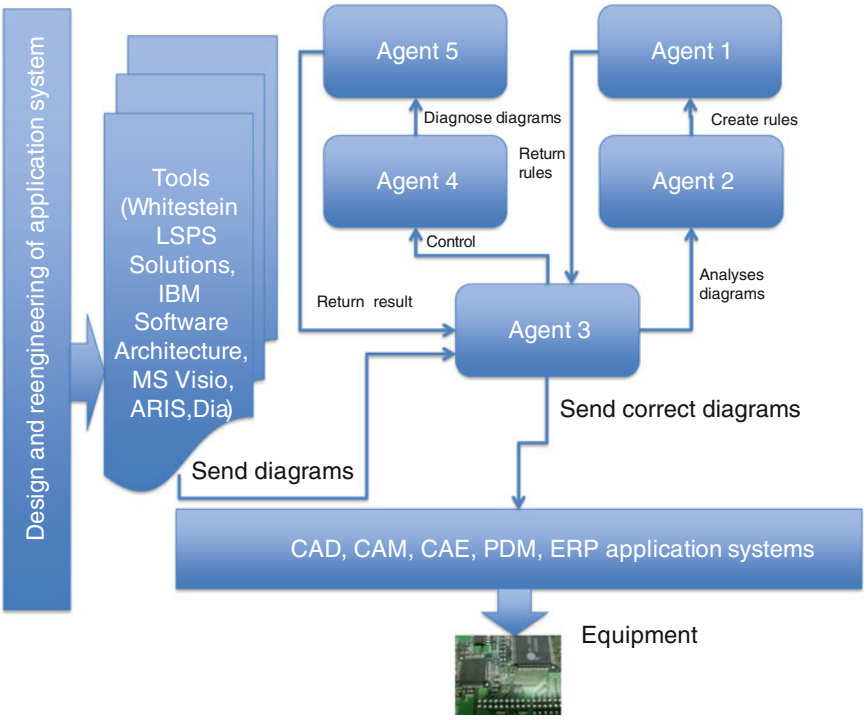


Fig. 1 Our DCA architecture of the intelligent agent system

Agent diagram model (1): contains a description of a diagram as a table of rules.

Agent RV-grammar (2): creates a grammar to describe a diagram of a business process.

Agent planner (3): manages agents.

Agent execution control (4): performs actions getting from agent planner.

Agent diagnostic (5): checks a diagram for errors.

Application systems as CAD, CAE, CAM, PDM, ERP: use the diagrams to plan and develop manufacturing equipment.

Any of the agents can proactively start to plan an action. The planning process have three phases: Creation, Commitment, and Execution. During the Creation phase, an agent recognizes the current situation, creates a plan template (using Announcement Protocols, and Ask-about Protocols) with other agents to cover the whole plan template. In the Commitment phase, the agents commit their resources to fulfill the task. In the last, Execution phase, the plan is executed by all participating agents (using ask/Action Agreement Protocols).

Announcement Protocols: the agent with the task sends out an announcement of the task as a specification that must encode a description of the task, and any constraints. Ask-about Protocols: agents that receive the announcement decide what to do to solve the task. Ask/Action Agreement Protocols: agents execute the plan.

The agents communicate to each other using a Job Description Language (JDL) [21] as a content language for FIPA-ACL. Intelligent agent system supports Directory Facilitator (DF) that provides matchmaking (to locate agents by capabilities) and the Agent Management Services (AMS) functionality (to locate agents by addresses). Communication between agents depicts in Fig. 2.

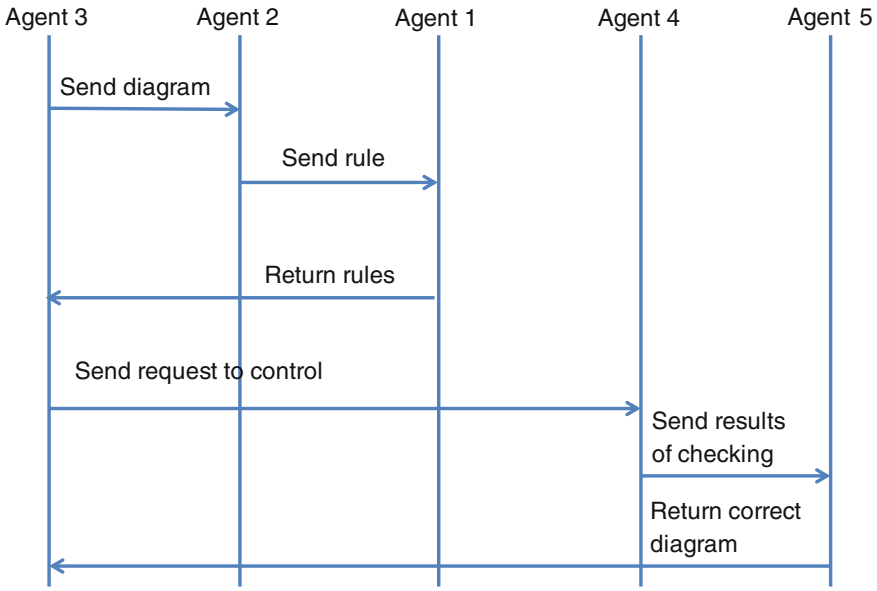


Fig. 2 Sequence diagram for communication between agents

We use WADE platform [22] to create the intelligent agent system that may integrate with MS Visio, Dia.

The agent allows to perform follows:

- analyses any diagrams (BPML, UML, IDEF0, IDEF3, Petri net) using RV-grammar;
- be integrated as plug-ins into MS Visio, Dia;
- checks 20 types of errors, and 4 additional errors;
- gives recommendations to a designer for improve a diagram;
- supports a distributed design as a part of DAI.

4 Results

We review industrial areas that has an interest in application tools to control and diagnose business process and workflow as diagrams. So, we define BPML, UML, IDEF0, IDEF3, Petri net, and extract 20 types of errors that may occur into them. Modern tools can only check 16 types of errors. Thus, we develop a new RV-grammar to check and fix all type of errors, which has a linear require of time to analysis oppositely with other grammars with exponential and polynomial requires of time. Figure 3 shows the efficiency. Graphic objects are graphic figures as a circle, a rectangle, a rhombus, a square, a line and etc. We use a formula [17–19] for calculation of the efficiency that cab be written as:

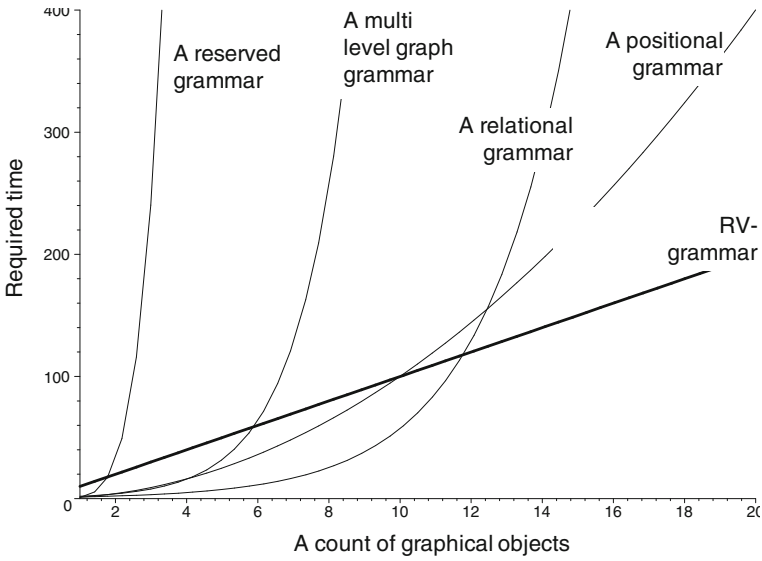


Fig. 3 Efficiency of checking and fixing errors with the help RV-grammar into diagrams

$$Required\ time = \mathrm{c} \cdot L_s, \tag{3}$$

$$L_s = \sum_{i=1}^{m}\left(\sum_{j=1}^{V_i} v_{in_{ij}} + \sum_{j=1}^{V_i} v_{out_{ij}}\right) + \sum_{i=m+1}^{t}\left(V_i + \sum_{j=1}^{V_i} v\_out_{ij}\right) + no\_label \tag{4}$$

where

c	the constant of realization of algorithm, which determines a quantity of time (operators) that are spent to analysis one graphic object;
L_s	the number of graphic objects;
V_i	the number of graphic objects of i-type;
v_in_{ij}	the number of inputs to j-graphic objects of i-type;
v_out_{ij}	the number of outputs from j-graphic objects of i-type;
t	a total of object types;
m	a quantity of object types that have more than one output.

We review a lot of paper and tools, and find that there are no tools to check and fix semantic errors in diagrams of workflow and business process. We investigate and develop intelligent agent system, supporting DCA architecture, FIPA protocols, and using WADE platform, that is distributed and helps to control and diagnose workflow diagrams of distributed industrial corporations. It offers to remove a lot of errors, including semantic, when designed equipment.

5 Conclusion

We found interesting that is specific sectors of industry, especially those that are motivated to streamline their business process and cut the costs, require solutions to be delivered in the short amount of time [23]. Many business process cab be written with diagrams. When we may control the diagrams for errors and fix them, we increase a quality of the diagrams. Any diagram has a grammar that can be used to check and fix errors. So, we develop a new RV-grammar to analyses and control of the diagrams, which offers to check and fix 20 types of errors (4 additional errors). Therefore, RV-grammar check and fix 20 % more than a reserved graph grammar, a positional grammar, and a relational grammar.

A lot of concerns, consisting of corporations and plants, want to have distributed tools to control and diagnose workflow as diagrams. We thought that RV-grammar can be written with the help MAS to create distributed systems for manufacturing, that can efficiency check, and fix errors, and improve a quality of business process and workflow for industrial corporation. Thus, we create intelligent agent systems, based on RV-grammar, for manufacturing, that contains our DCA architecture, supporting FIPA protocols, 5 new agents, communication protocols, and the use of WADE platform.

The intelligent agent system will deploy on industrial corporation to remove ill-fitting processes, and soaring process development, and improvement costs.

Acknowledgments The reported study was partially supported by RFBR, research project №15-07-08268a.

References

1. Huhns, M.N.: Distributed Artificial Intelligence. Pitman/Morgan Kaufmann, SanMateo, CA (1987)
2. Parunak, H.V.D.: Manufacturing experience with the contract net. In: Huhns, M.N. (ed.) Distributed Artificial Intelligence, pp. 285–310. Morgan Kaufmann Publishers, San Mateo, CA (1987)
3. M'uller, J.P., Fischer, K.: Application Impact of Multiagent Systems and Technologies: A Survey. http://www.researchgate.net (2014)
4. A global Swiss company offering advanced intelligent application software for multiple business sectors. http://whitestein.com/
5. Pechoucek, M., Marik, V.: Review of Industrial Deployment of MultiAgent Systems. http://citeseerx.ist.psu.edu/ (2008)
6. Alkhateeb, F., Maghayreh, E.A., Doush, I.A. (eds.): Multi-Agent Systems—Modeling, Control, Programming, Simulations and Applications (2011)
7. Fu, K.: Structural Methods of Pattern Recognition, p. 319. Mir, Moscow (1977)
8. Costagliola, G., Lucia, A.D., Orece, S., Tortora, G.: A parsing methodology for the implementation of visual systems. http://www.dmi.unisa.it/people/costagliola/www/home/papers/method.ps.gz
9. Wittenburg, K., Weitzman, L.: Relational grammars: Theory and practice in a visual language interface for process modeling. http://citeseer.ist.psu.edu/wittenburg96relational.html (1996)
10. Zhang, D.Q., Zhang, K.: Reserved graph grammar: A specification tool for diagrammatic VPLs. In: Proceedings. 1997 IEEE Symposium on Visual Languages, pp. 284–291. IEEE (1997)
11. Zhang, K.B., Zhang, K., Orgun, M.A.: Using Graph Grammar to Implement Global Layout for A Visual Programming Language Generation System (2002)
12. ARIS. http://www.ariscommunity.com/aris-express
13. MS Visio: http://www.microsoft.com/ru-ru/office/vip/visio.aspx
14. Dia. https://wiki.gnome.org/Apps/Dia/
15. Aho, A.V., Johnson, S.C.: LR Parsing (1974)
16. IDEF. http://www.idef.com/idef0.htm
17. Sharov, O.G., Afanas'ev, A.N.: Syntax-directed implementation of visual languages based on automaton graphical grammars. Program. Comput. Softw. **6**, 56–66 (2005)
18. Sharov, O.G., Afanas'ev, A.N.: Neutralization of syntax errors in the graphic languages. Program. Comput. Softw. **1** 61–66 (2008)
19. Sharov, O.G., Afanas'ev, A.N.: Methods and tools for translation of graphical diagrams. Program. Comput. Softw. **3**, 65–76 (2011)
20. Pechoucek, M., Marik, V.: Review of Industrial Deployment of MultiAgent Systems
21. Tich, P., Lechta, P., Maturana, F., Balasubramanian, S.: Industrial MAS for Planning and Control. In: LNAI, No. 2322, pp. 280–295. Springer, Heidelberg (2002)
22. A software platform based on JADE that provides support for the execution of tasks defined according to the workflow metaphor. http://jade.tilab.com/wadeproject/
23. Belecheanu, R.A., Munroe, S., Luck, M., Miller, T., McBurney, P., Pechoucek, M.: Commercial applications of agents: lessons, experiences and challenges. In: Proceedings of AAMAS-06—Industry Track, pp. 1549–1555. ACM Press (2006)

Part VII
Applied Systems

Use of Numerical Methods in the Analysis of Traction Energy Systems—An Overview of the Practical Examples

Mikołaj Bartłomiejczyk

Abstract A characteristic feature of trolleybus transport is the random nature of traffic caused by congestion. It predestinates statistical and numerical methods for the analysis of trolleybus energy system. There are presented 3 methods of trolleybus traction system analysis: simulation of supply system based on Monte Carlo method, analysis of energy recovery potential based on statistical data analysis and benchmark of trolleybus supply system by one of Multiple Criteria Data Analysis—Data Envelopment Analysis (DEA) method.

Keywords Electric traction · Supply system · Data analysis · MCDA · DEA

1 Introduction

Trolleybus, as opposed to rail public transport vehicles, in most situations uses public roads shared with other vehicles. The result of this is exposed to the congestion, which is the problem associated with heavy traffic of road vehicles. Results of this the movement has an irregular nature, even if the timetables are based on the constant intervals, the real traffic of vehicles is irregular. The impact of other road users on trolleybus movement is very significant. As a result, the power load of the trolleybus power system also has a stochastic nature and a trolleybus power system must be treated as a stochastic object [1].

The influence of road congestion on the load of supply system is presented on Fig. 1. Load of an exemplary supply system with two vehicles is presented. The same figure depicts the situation, when the vehicles run with constant intervals and the load is regular. But may occur in traffic delays the imposition of currents of both vehicles. As that result, the instantaneous value of total load can increase up to two times. Random movement disorders may significantly affect the load supply.

M. Bartłomiejczyk (✉)
Faculty of Electrical and Control Engineering, Gdansk University of Technology, G. Narutowicza 11/12, 80-233 Gdansk, Poland
e-mail: mikolaj.bartlomiejczyk@pg.gda.pl

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_40

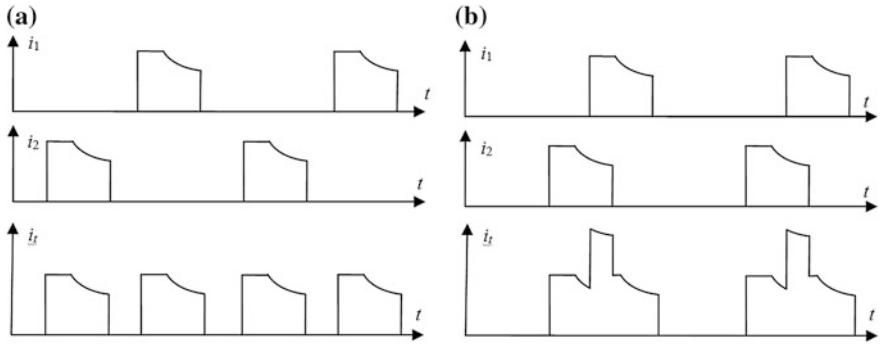


Fig. 1 An example of the possible impact of irregular movement on load substation current: **a** vehicles move according to a timetable in fixed intervals; **b** irregular movement; i_1, i_2—current vehicles, i_t—current power supply

This situations have random nature and they are difficult for simulation modelling. Popular methods of supply system analysis are based on analytic methods or time-simulation and differential equations. Due to, it is very difficult to take into account the random nature of their supply. Alternative methods of analysis are based on statistical methods, which are presented in the paper.

2 Trolleybus Supply System Calculations

Trolleybuses are supplied with direct power from overhead lines—catenary. The energy used to supply the municipal overhead lines is drawn from the public HV AC grid and is converted to DC in traction substations, where voltage reduction occurs. After transformation and rectification, the energy from traction substations is supplied into power supply points of the overhead lines using power cables (feeders). The overhead lines are divided into traction power supply sections so that it is possible to turn off a part of the line while maintaining supply to the remaining part [2].

2.1 The Monte Carlo Method

Solution for many computational problems bases on an algorithm (list of actions), that leads to finding the value f, either precisely or with a specified error. Shall we label with $f_1, f_2, \ldots, f_n$ results of subsequent accumulations of an algorithm, then

$$f = \lim_{n \to \infty} f_n \tag{1}$$

Process performs a limited number of iterations before it is stopped. This process is strictly determined: every algorithm results in an identical solution. There are problems for which it is complicated to define such an algorithm. In such situation the task is modified, basing on the law of large numbers (LLN) from the probability theory. Utilising the stochastic analysis related to multiple random samples evaluations $f_1, f_2, \ldots, f_n$ of the searched variable are obtained. This requires the random variable f_n to be stochastically convergent to the searched value f. Therefore it is valid for any $\varepsilon > 0$:

$$\lim_{n\to\infty} P(|f - f_n| < \varepsilon) = 1 \tag{2}$$

where P stands for probability of the given event. Choice of the variable f depends on each problem's specifics. Frequently the searched variable is considered as an occurrence of a particular event. Such computational process is not deterministic as it is defined by set of random samples' outcome [1].

Methods using algorithms that rely on repeated random sampling to obtain are called Monte Carlo methods. In this method, precise assessment of various relationships between the input data and searched values is not necessary to precisely. That states for a great advantage of such approach thus enabling multiple factors' impact analysis.

2.2 Simulation of Supply System

The simulation model [1] is based on the following initial data: trolleybus timetables, deviations in timetables, the trolleybus speed profile (a relationship between expected vehicles speed and their location), traction characteristics of trolleybuses. On the basis of timetables and deviations from their realization there was indicated a probability histogram of the number of vehicles operating simultaneously on the power supply section. A number of vehicles being on the power supply section at the very moment is indicated on its basis. The speed profile (Fig. 2) is the basis for indicating the layout of specific vehicles location probability along the power supply section (Fig. 3).

Probability $P(s_1, s_2)$ to find the vehicle between points with coordinates s_1 and s_2 is proportional to the time of travel between these two points, which can be written:

$$P(s_1, s_2) = k \cdot \frac{s_2 - s_1}{v_{sr}} \tag{3}$$

where: $v_{śr}$—average speed on the road between points s_1 and s_2; k—coefficient of proportionality. This probability is equal to the integral of the probability density $p(s)$:

$$P(s_1, s_2) = \int_{s_1}^{s_2} p(s)ds \tag{4}$$

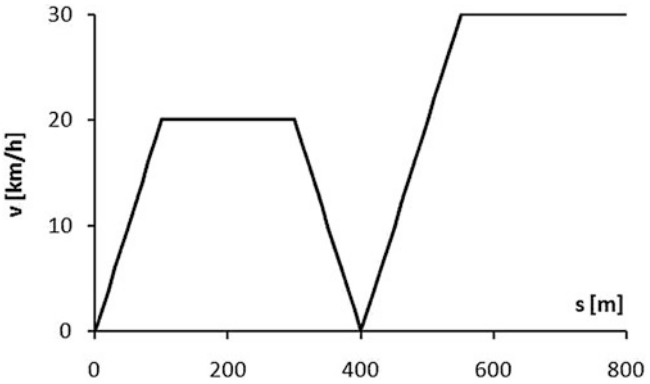


Fig. 2 An exemplary speed profile; s—the vehicle location, v—speed

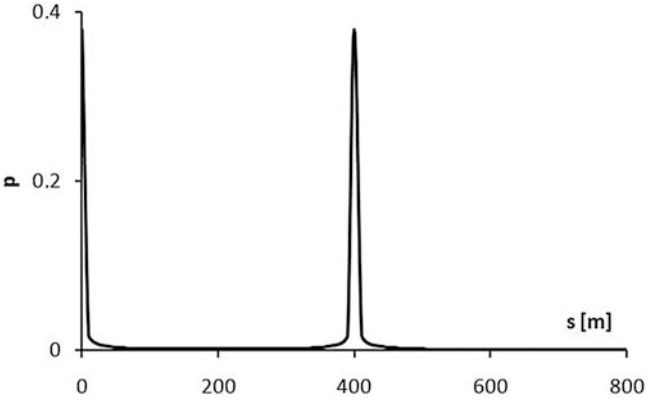


Fig. 3 A layout of the vehicles (s) location probability (p)

marking the difference $s_2 - s_1$ as Δs we can write:

$$P(0, \Delta s) = k \cdot \frac{\Delta s}{v'_{sr}} = \int_0^{\Delta s} p(s)ds \tag{5}$$

on $\Delta s \rightarrow 0$ this equation takes the form:

$$p(s) = k \cdot \frac{1}{v(s)'} \tag{6}$$

which means that the density of the probability of finding a vehicle in a given point is inversely proportional to its speer [1].

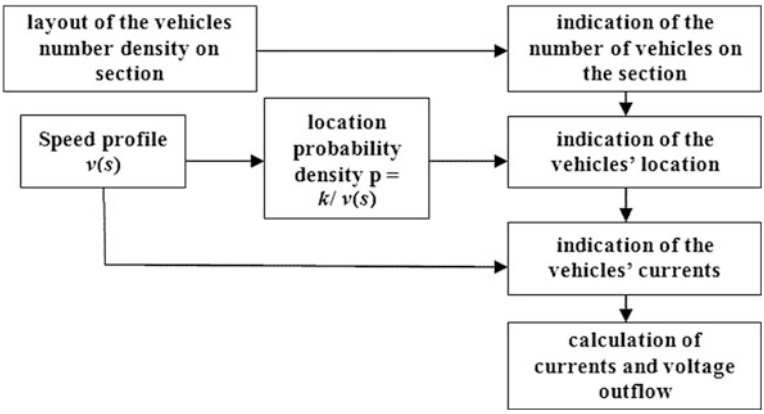


Fig. 4 A basic diagram of a simulation model

A basic simulation cycle (Fig. 4) includes the following phases:

1. indicating the number of trolleybuses being on the power supply section on the basis of timetables and deviations from their realization,
2. determining the location of vehicles on the basis of the vehicles location probability layout,
3. determining the currents absorbed from the traction network by vehicles,
4. calculating currents and voltage outflow in the power supply system.

This cycle is numerously repeated and as a result one acquires currents and voltage probability layout in the power supply system. The current of specific trolleybuses is determined on the basis of the trolleybus speed and its derivative dv/dt in a defined place, which determines the phase: accelerating, breaking, or inertial ride.

2.3 Verification of the Monte Carlo Method

The simulation research was carried out in Scicos/Scilab program. The Fig. 5 shows the layout histograms of currents value density probability resulting from a simulation run for scheduled and irregular traffic on the power supply section. In both cases these histograms related to identical measurements data recorded in realistic conditions, namely irregular traffic.

The Fig. 5 and Table 1 depicts an influence of the traffic punctuality on the feeder load current value visible in the increase of the maximum feeder current value while taking into account in irregularity of the trolleybus rides. The effect of irregular trolleybus running is simultaneous occurrence of a large amount of vehicles on the power supply section in relation to the regular traffic. As a result of it, the probability of simultaneous start-ups of several trolleybuses is on the

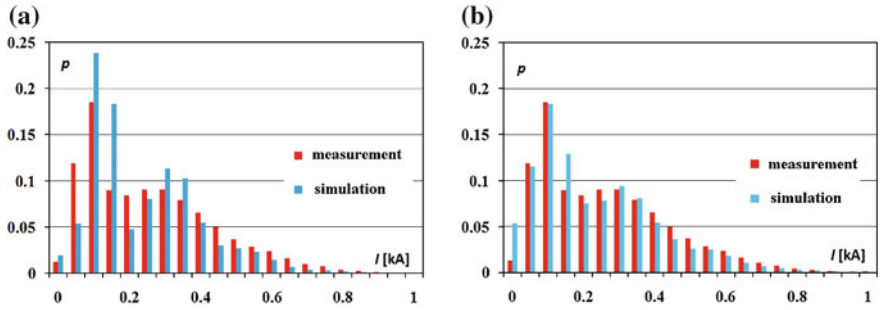


Fig. 5 The histograms of the feeder current, a simulation for: **a** the traffic consistent with the timetable; **b** irregular traffic

Table 1 Current load measurements and results of simulation of Kołłątaja supply section

	Measurement	Regular traffic		Traffic with disturbations	
		Simulation	δ (%)	Simulation	δ (%)
Average current	236 A	209 A	11	208 A	11
Maximal current	1075 A	911 A	15	1090 A	1

increase. It is evident from the data mentioned above, that taking into consideration the irregularity of traffic in simulation has significant influence calculation results, especially on the maximal load current value. For this reason the Monte Carlo method of simulation is better than standard analytic method [3].

3 Analysis of Recuperation Increase Potential

Every electrical machine is characterized by its capability of operating with a bidirectional energy flow. For traction motors it means that braking enables regenerative operation, which consists in converting the vehicle's kinetic energy into electrical energy, which in turn generates braking torque [4].

In a classic supply system, where substations are not equipped with energy-storing devices, the recuperative braking energy can be absorbed only by the second vehicle in the power supply section, i.e. the vehicle is in motion. However, in common situations in the power supply areas there are no vehicles present ability to absorb the energy. When this situation occurs, the recouped energy is dissipated in the braking resistors. It results is under-utilization of the potential for recuperation. In order to avoid such situation, the storage energy systems can be installed. Moreover, installation of systems of that type connected with the significant financial investments and it is necessary to perform analysis of potential of not-used recuperation energy. It can be realized by theoretical

simulation, but due to numerous random parameters which affect traffic movement this simulation is fraught with significant error [2]. Another method is to provide the measurement of energy dissipated in breaking resistors in vehicles [2], but it requires vehicle to be equipment with measurement system. The simplest method is to analyze the load of traction substations.

The analysis is based on the example of two traction substation in Gdynia trolleybus system: Traction Substation Grabówek and Traction Substation Sopot. These two substations are significantly different in terms of traffic intensity. TS Grabówek supplies large supply area with high density of traffic (every 3 min), TS Sopot supplies sub-urban line with a minor traffic intensity (30 min).

3.1 Statistical Substation Load Analysis

In order to recover braking energy it is necessary to ensure the receivers for the generated energy. In classical supply system without storage energy systems with one-side supply of catenary and without taking into account transmission losses this means that substation load P_{TS} must be larger than the power P_{rec} generated during recuperation by vehicles in supply area of substation. It can be written by equation:

$$P_{TS} \geq P_{rec} \tag{7}$$

If this condition is fulfilled, all generated energy is absorbed in supply system. Substation load has randomly character as a result of the influence of congestion, Fig. 6 shows the histogram of TS Grabówek and TS Sopot load. The different character of the trolleybus traffic in the areas of both power substation is reflected in

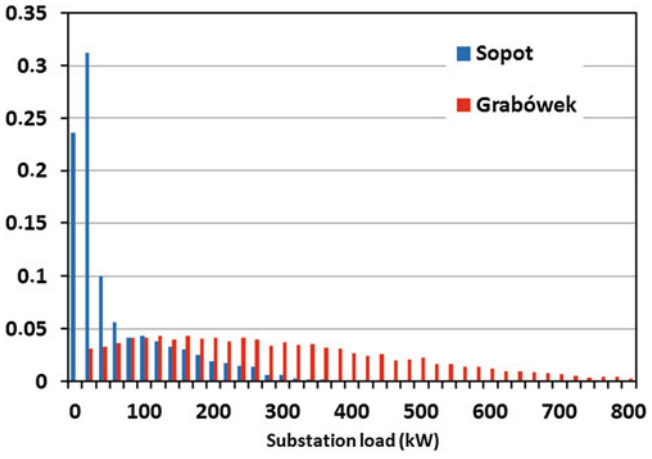


Fig. 6 The histograms of load of substations TS Grabówek and TS Sopot

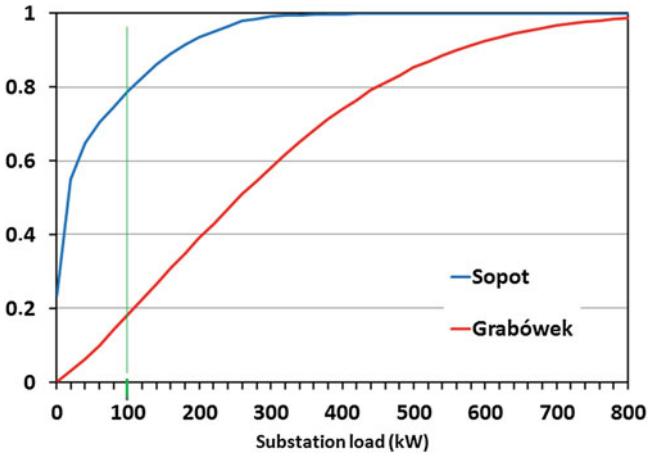


Fig. 7 The cumulative distribution of load of substations TS Grabówek and TS Sopot. The average value of recovery breaking power, equals 100 kW, is marked with *green*

histograms. In case of TS Sopot most of the values is cumulated in the left side, what is caused by the much lower load of TS Sopot than TS Grabówek.

Fulfilling of the condition (7) can be analyzed basing on the cumulative distribution on the traction substation load (Fig. 7). Value d of cumulative distribution corresponding of the particular value P of load means, that substation load will be smaller than P with the probability d. Basing on the (7) can be said, that recovery breaking power of the value P will be fully consumed in supply system with the probability d. The average power generating during breaking equals 100 kW, so in case of TS Grabówek the probability that the power generated during regenerative breaking will be completely consumed is 0.8, in case of TS Sopot it is 0.2 (Fig. 7).

If condition (7) is not fulfilled, only a part of recuperated power, equals the load P_{TS}, can be re-used. The remaining energy is lost in breaking resistors. Therefore, due to limited load power P_{TS} of substation is possible to use only a part generated energy.

It was analyzed how much energy from the source with power P_{gen} is able to absorb the supply system with changeable value of load $P_{\mathrm{TS}}(\mathrm{t})$. Source P_{gen} is an equivalent of vehicle during regenerative breaking. The calculations were based on registrations of loads $P_{\mathrm{TS}}(\mathrm{t})$ of traction substations in Gdynia.

For this purpose the function $e(P_{\mathrm{gen}})$, which expresses the relative amount of generated energy from the source P_{gen} in time T has been defined:

$$e\left(P_{gen}\right) = \frac{\int_0^T P_{Ch}\left(P_{gen}, t\right) dt}{P_{gen} \cdot T} \tag{8}$$

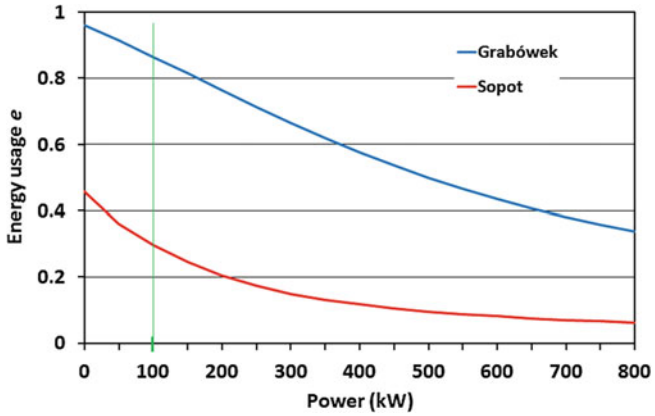


Fig. 8 Relative usage of recuperated energy in function of value of recuperation power for two substations: TS Grabówek and TS Sopot. The average value of recovery breaking power, equals 100 kW, is marked with *green line*

$P_{\mathrm{Ch}}(P_{\mathrm{gen}}, t)$ expresses the limit of absorption of generated energy:

$$P_{Ch} = \begin{cases} P_{gen} < P_{TS} \rightarrow P_{Ch} = P_{gen} \\ P_{gen} \geq P_{TS} \rightarrow P_{Ch} = P_{TS} \end{cases} \tag{9}$$

On the Fig. 8 the values of relative usage of generated energy is presented. As it is observed, in case of TS Grabówek substation 80–90 % of generated energy during recuperation can be absorbed by supply system. In case of TS Sopot substation the level of recuperation usage is only 20–30 %. It can be explained by the fact, that with the bigger number of vehicles in substation supply area the probability of founding receiver for recuperated energy increases and as a result of this the efficiency of breaking recovery increases too.

4 Benchmark of the Trolleybus Supply System

In various situations it is necessity to make a benchmark of existing technical state of the supply system. It can be useful e.g. for feasibility studies of modernization of trolleybus infrastructure. The main element is the technical evaluation of individual elements of the supply system [3].

The main electrical and motion parameters, which describes the load and usage of supply sectors are: average value of feeder current I_{av}, maximal value of feeder current I_{max}, length of the supply section l, trolleybus traffic interval: Δt.

In order to perform benchmarking and comparison of analyzed supply sections there is need to create one complex index, which splits these four independent parameters into one value. It can be performed by application of Multi Criteria Decision Analysis. One of these methods, which is used, is Data Envelopment Analysis (DEA).

4.1 Data Envelopment Analysis (DEA)

The DEA model is used to estimate and analyze efficiency of multiple units and multiple outputs, which can have very different units of measurement [3]. A common measure for relative efficiency is:

$$Efficiency = \frac{weighted - sum - of - outputs}{weighted - sum - of - inputs} \tag{10}$$

For each analyzed option the efficiency is defined as a ratio of the weighted sum of outputs to the weighted sum of inputs [3]. In the DEA model for a particular option in the data set of n options (j = 1, 2, …, n) a h_o value is determined using a set of the best indicator weights u_r *and* v_i which h_o value of each option satisfies the restrictions. Each option is described by r values, $v_1 - v_r$ are the inputs and $u_1 - u_r$ are the outputs:

$$\max h_0(u, v) = \frac{\sum_r u_r y_{r0}}{\sum_i v_i x_{i0}} \tag{11}$$

$$s.t. \frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}} \leq 1, j = 1, 2, \ldots, n \tag{12}$$

$$u_r, v_i \geq 0, \text{for all } i \text{ and } r \tag{13}$$

To make it more clear, DEA is a linear programming model, where each entity selects a set of weights which are most favorable for itself to give a standardized efficiency score (between zero and one). The DEA model (8–10) allows to choose "the best" set of weights to use.

4.2 The Analysis of the Load Flow by DEA

In the subject of analysis the inputs will be the motion parameters of the network: Δt and l. The values of load flow I_{av} and $I_{\max}$ will be taken as the output parameters.

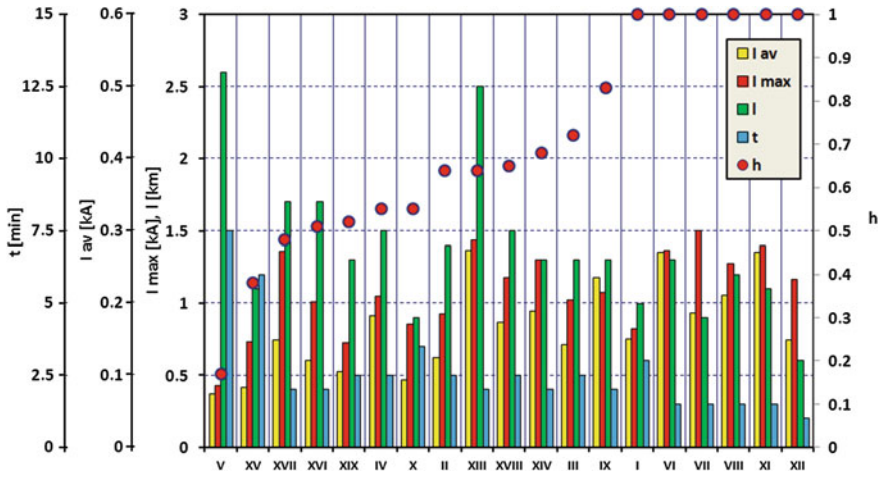


Fig. 9 The inputs, outputs and the values of supply sections efficiency

The efficiency estimated by DEA implies how big is the load flow of each supply section in relation to the motion parameters. A bigger value of efficiency means bigger values of the supply section currents compared to motion parameters, consequently it means worse traffic conditions (stronger congestion). In the Fig. 9 there are presented the results of the calculations, *h* is the values of efficiency.

The performed DEA analysis proves the relationship between the intensity of the road traffic and the value of calculated efficiency. The highest values of efficiency *h* are visible for sections I, VI, VII, VIII, XI, XII. This result can be easily explained on the basis of traffic observation. The sections VI, VII, VIII, are located on the Morska street, which is one of the most important and most crowded streets in Gdynia. The sections XI, XII ale placed in the very center of city with many stops and crossings. Also section I can be characterized as a section with high intensity of the trolleybus traffic as there is localized one of the biggest trolleybus stations. On the other hand the lowest value of efficiency is reached in section V which is located in the suburbs of the Gdynia.

By means of DEA can be created a numerous benchmark of the traffic congestion in the electrical traction, which provides a quantitative assessment of a complicated phenomenon. Multi-attribute decision methods allow for the analysis of various not connected with each other and difficult indicate elements.

5 Conclusions

The presented methods are results of practical experiences of the paper author gained during his work in the trolleybus transport system of Gdynia. Numerical methods allow you to perform analysis of the trolleybus power supply system

including the impact of random factors [5–9]. Main contribution of paper is practical aspect of presented research. Tasks shown in the paper were base on technical problems which have appeared in the practical exploitation of trolleybus supply system and the sophisticated numerous tools were used to solve them.

Acknowledgments This paper has been elaborated in the framework of the project ELIPTIC co-financed by the European Union's Horizon 2020 research and innovation programme under the grant agreement No 636012.

References

1. Hamacek, S., Bartlomiejczyk, M., Hrbac, R., Misak, S., Styskala, V.: Energy recovery effectiveness in trolleybus transport. Electr. Power Syst. Res. **112**, 1–11 (2014)
2. Bartłomiejczyk, M., Połom, M..: Spatial aspects of tram and trolleybus supply system. In: 8th International Scientific Symposium on Electrical Power Engineering (ELEKTROENERGETIKA), pp. 223–227, Technical University of Kosice, Faculty of Electrical Engineering and Informatics, Department of Electrical Power Engineering (2015)
3. Bartłomiejczyk, M., Połom, M., Styskala, V.: Benchmark of the traffic congestion in electrical transport by means of multi criteria decision analysis. Przegląd Elektrotechniczny 11, 248–252 (2013)
4. Jarzebowicz, L.: Indirect measurement of motor current derivatives in PMSM sensorless drives. Elektronikair Elektrotechnika **20**(7), 23–26 (2014)
5. Korenčiak, D., Gutten, M.: Opportunities for integration of modern systems into control processes in intelligent buildings. Przegląd Elektrotechniczny **88**(2) (2012)
6. Šebök, M., Gutten, M., Kučera, M.: Diagnostics of electric equipments by means of thermovision. Przegląd Elektrotechniczny **87**(10) (2011)
7. Judek, S., Skibicki, J.: Wyznaczanie parametrów elektrycznych trakcyjnego układu zasilania dla złożonych warunków ruchu przy wykorzystaniu programu pspice. Przegląd Elektrotechniczny 12, 270–273 (2009)
8. Figlus, T.: The application of a continuous wavelet transform for diagnosing damage to the timing chain tensioner in a motorcycle engine. J. Vibroeng. **17**(3), 1286–1294 (2015)
9. Kolář, V., Hrbáč, R.: Measurement of ground currents leaking from DC electric traction. In: 15th International Scientific Conference on Electric Power Engineering (EPE), pp. 613–617. Brno, Czech Republic (2014)

Application of Vector Control Technology for Linear Reactive Reluctance-Flux Reciprocating Generator

Pavel G. Kolpakhchyan, Alexey R. Shaikhiev and Alexander E. Kochin

Abstract The article deals with application of vector control technology for linear reactive reluctance-flux reciprocating generator. The applying of Park's equations to describe the processes in the electrical machine at issue is justified. The results of mathematical modeling of processes in the system composed of the free piston combustion engine and the linear reactive reluctance-flux generator when using method of regulation at issue are given.

Keywords Reluctance-flux electrical machine · Linear electrical generator · Vector control

1 Introduction

One of the ways to improve the autonomous electrical generators based on internal combustion engines is the use of linear reciprocating electrical machine. The common use of such electric machines with the free-piston engine allows avoiding the use of a crank mechanism. It makes possible to eliminate the heavily loaded hinges, to adjust the degree of compression and to change the stroke time of engine cycle. As a consequence, the engine can be adapted to any liquid or gaseous fuel and its work is optimized under different loading [1].

The free-piston engines systems have a number of advantages compared with crank mechanism engines. They have a simple structure with a small number of moving parts. It reduces friction losses and simplifies system maintenance. High speed of the piston when it passes the top dead center and a high rate of expansion

P.G. Kolpakhchyan (✉) · A.R. Shaikhiev
Science and Production Association "Don Technology", Novocherkassk
Russian Federation
e-mail: kolpahchyan@mail.ru

A.E. Kochin
Rostov State Transport University, Rostov-on-Don, Russian Federation

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_41

during the power stroke improve formation of fuel-air mixture, reduce heat loss and reduce the harmful substances emissions that envy of the temperature [2, 3].

In this article we will describe the new construction of the linear electrical machine working as electrical generator in conjunction with a free-piston combustion engine.

The paper organized as follows. The analysis of existing developments in the area of application of linear electrical generators intended to the work in conjunction with a free-piston combustion engine is presented in Sect. 2. The construction of linear reactive electrical machine is describes in Sect. 3. Section 4 deals with the modeling of electrical machine under consideration with the use of Park equations. In Sect. 5 the control problem formulation is accomplished. In Sect. 6, an illustrative example is presented, which clearly demonstrates the correct apply of the electrical machine control methods. Section 7 contains conclusions.

2 Known Research in the Field of Electric Generators on the Basis of Free-Piston Engines

Various research teams and industrial companies are engaged in the problem of creating the linear electric machines to be used as power generators in conjunction with the free-piston engines. České vysoké učení technické v Praze (Fakulta elektrotechnická) [4, 5]; Newcastle University, UK [2, 3, 6–8]; Sandia National Laboratories. Livermore, USA [9]; Toyota Central R&D Labs Inc., Japan [10, 11] present the most interesting workings.

Experience in development of linear reciprocating electrical machines allows revealing a number of factors that impair their effectiveness. Their account is required in the design. There are following: edge effect, the limited mass of the movable parts of the construction, pulsating and discontinuous velocity dependence on the time. The magnetic circuit construction imposes high requirements.

Synchronous electrical machine with permanent magnets on the movable element is the most common type of linear electrical machines used as generators. With high efficiency, good output characteristics, they are expensive and require complex control systems. All this considerably affects the cost and manufacturability of electricity autonomous source system in the whole. Therefore, the use of reactive (reluctance-flux) electrical machines without the windings and the permanent magnets on the rotor is promising.

There are different types of linear generator conjunction with free-piston engine: with free end, with gas spring, with opposed cylinders [12]. One cylinder and the gas spring system is the most used and the most researched system. It significantly reduces the impact of inertial forces of moving parts and is a compromise between the desire to make a mechanically balanced system and the simplicity of control of the combustion process in the cylinder [6–8].

3 Design of the Linear Electric Machine and Control Problem Formulation

Toyota Central R&D Labs Inc project is one of the most interesting workings in the field of electric generators based on a free-piston internal combustion engines. W-shape piston is the main feature of the proposed design (Fig. 1) [10, 11]. Location of the active parts of the generator on the outer side of the device and reduction of its length by the arrangement inside the cylinder-piston group is an advantage of this design. A large diameter of the piston of the gas spring is another advantage. It allows improving the gas-dynamic processes, reducing energy losses and heating.

This construction is the most rational to create electric generator based on the free-piston engine, so this design was the prototype for a linear generator described further [6–8, 10, 11]. Permanent magnets are located on the movable element in close proximity to the combustion chamber in zone of the large mechanical loads that are disadvantages of the described system. Therefore, we decided to use the reactive electrical machine without winding and permanent magnets on the movable element as electrical generator. Application of three-phase distributed winding on stator allows a uniform force throughout the range of movement of the movable element.

The preliminary engineering of the linear reactive electrical machine with three-phase distributed winding was performed to assess the possibility of its use as an electric generator operating in conjunction with a free-piston engine. This electrical machine was designed to work with a single cylinder and the gas spring.

In the design, the following parameters of the internal combustion engine have been taken:

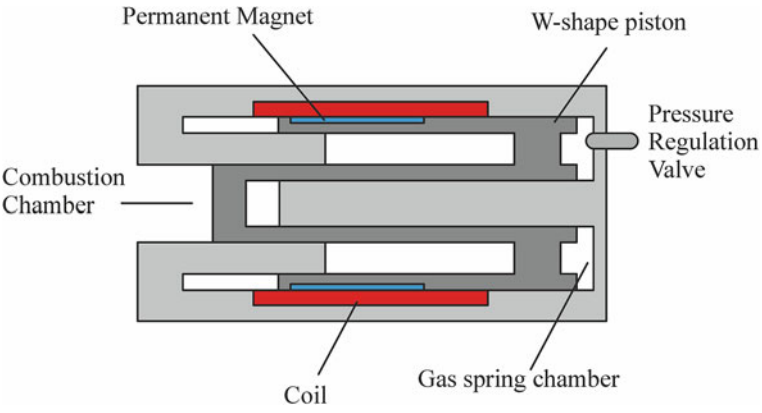


Fig. 1 Sketch of Toyota Central R&D Labs Inc. linear electric generator based on the free-piston engine

The Work cycle	Otto, two stroke
The Cylinder diameter	76 mm
The piston stroke	90 mm
Compression ratio	8.8
The pressure in the cylinder at the beginning of the expansion stroke	8.4 MPa
The operating frequency	50 Hz

The basic parameters of the electrical machine of the considered type were obtained as a result of preliminary calculation:

The number of phases of the stator winding, m	3
The number of poles, $2p$	8
The inner diameter of the stator, d_1	140 mm
The outer diameter of the stator, D_1	210 mm
The air gap above the middle tooth of the movable element	0.25 mm
The inner diameter of the rotor, d_2	90 mm
The pole pitch, τ	90 mm
The number of turns per pole and phase, q	1

The gap above the rotor teeth is made uneven; increasing towards the edges to 1.0 mm. Figure 2 shows a sketch of the active area of the described electrical machine.

Continuous work in the transient regime is one of the features of the reciprocating linear electrical machine work. The moving element performs sinusoidal oscillations; its speed also varies according to the harmonic law. The electric generator work on the basis of the free-piston internal combustion engine is largely determined by the ability to maintain oscillation of the moving element with the given parameters: amplitude and frequency. When this condition breaks down, the work efficiency of electric generator drastically reduces or becomes impossible. Maintaining waveforms are possible when the electrical machine can respond with

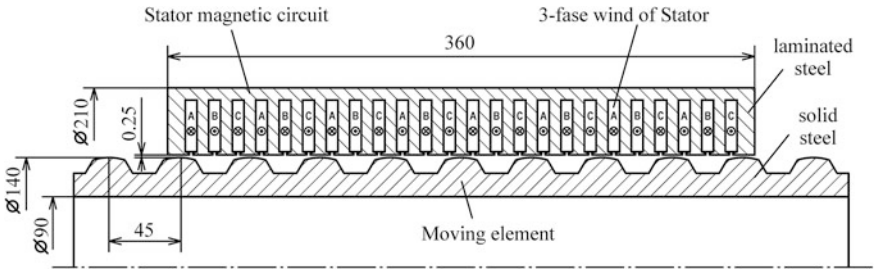


Fig. 2 Sketch of the active area of the reciprocating reactive reluctance-flux electrical machine

the sufficient speed to the change of the force acting on the piston of the internal combustion engine during the operating cycle. Therefore, the force control system of the electromechanical transducer must be fast and accurate. The constant of the electromagnetic force time control should not exceed 1.5–2.0 ms to maintain the oscillation with an amplitude value corresponding to the pole pitch (in this case 45 mm) and an operating frequency of 50–60 Hz.

A large value of windings inductance is a feature of this linear switched-reluctance electrical machine. The lack of the excitation winding increases the influence of cross coupling between the currents and voltages in the longitudinal and transverse axes and it is the cause of the large variations in the electromagnetic force in transient mods. Therefore, it is necessary to apply the principle of vector control for controlling electrical machine under consideration with a high precision and a required speed.

The use of vector control for the reactive reluctance-flux electrical machine has its own characteristics. The winding or permanent magnets on the rotor are absent and only the reactive component of the electromagnetic force is used. However, the system structure of the force control has no significant difference from similar control systems of electrical AC machines (asynchronous or synchronous with permanent magnets on the rotor) [12, 13].

The main idea of vector control is to divide flux control channels and electromagnetic force control channels. For this purpose, electromagnetic processes in the electrical machine are represented by Park's equations. Its use makes possible the selection of the stator current components on the longitudinal and transverse axes and the formation of control actions. Therefore, the application of vector control principles requires that process in a controlled electric machine were described with sufficient accuracy based on Park's equations.

4 Mathematical Model of the Linear Reluctance Electric Machine and the Free-Piston Engine

Park's equations represent the processes in the reactive reluctance-flux electrical machine under consideration [14, 15].

The processes are described in the $d - q$ coordinate system rotating in electric space synchronously with the magnetic flux vector. The three-phase stator winding is presented by two diametrical disposed windings concentrated orthogonally along the axes of the coordinate system. The transition from the natural three-phase coordinate system to the $d - q$ coordinate system is realized with the Clarke and Park transforms [14, 15].

The equations for the currents of windings that are arranged along the axes d and q can be written as:

$$\begin{aligned}\frac{di_d}{dt} &= \frac{1}{L_d}\left(U_d - R_s i_d + \omega_{0el} L_q i_q\right);\\ \frac{di_q}{dt} &= \frac{1}{L_q}\left(U_q - R_s i_q - \omega_{0el} L_d i_d\right),\end{aligned} \tag{1}$$

where U_d, U_q, i_d, i_q—voltages and currents of the windings placed along the d and q axis; L_d, L_q, R_s—inductance and resistance of windings placed along the d and q axes; ω_{0el}—radian frequency of rotation of the $d - q$ coordinate system.

The rotation angular velocity of the $d - q$ coordinate system in electric space is related to the linear displacement rate of the moving element by the relation:

$$\frac{d\gamma}{dt} = \omega_{0el} = \frac{\pi}{\tau} V, \tag{2}$$

where γ—the rotation angle of the $d - q$ coordinate system; V—linear speed of moving element;

The electromagnetic force is calculated by the expression:

$$F_{em} = \frac{3}{2} \cdot \frac{\pi}{\tau}\left(L_d - L_q\right) i_d i_q \tag{3}$$

An approach based on the idea of the simulated device in the form of a multibody system is used to describe the mechanical processes. The moving represents the body having a mass m and a single degree of freedom (the linear motion). The electromagnetic force, the force of the engine piston, the force of the gas spring, resistance force act on the body. The Moving of the moving element is described by the equation:

$$m\ddot{x} = F_p + F_{em} - d \cdot \dot{x} - C_s \cdot x, \tag{4}$$

where F_{em}, F_p—electromagnetic and engine piston force; x—displacement from the equilibrium position; m—moving element mass; C_s—spring stiffness; d—damping coefficient;

Considering the complicated configuration of the active layer of the stator and the rotor, the magnetic system saturation of the linear electric machine under consideration, to determine the parameters included in the system of Eq. (1) it is necessary to use the field theory methods. Design features a linear reactive reluctance-flux electrical machines allow considering the problem of calculating the magnetic field in axisymmetric statement.

The calculations were performed using the program FEMM (Finite Element Method Magnetic © David Meeker). The value of the electromagnetic force and the flux linkage of the stator windings according to the position the moving element and winding current values are determined on the basis of information about the distribution of magnetic fields in the computational domain. Figure 3 shows the results of calculating the magnetic field distribution in the electrical machine

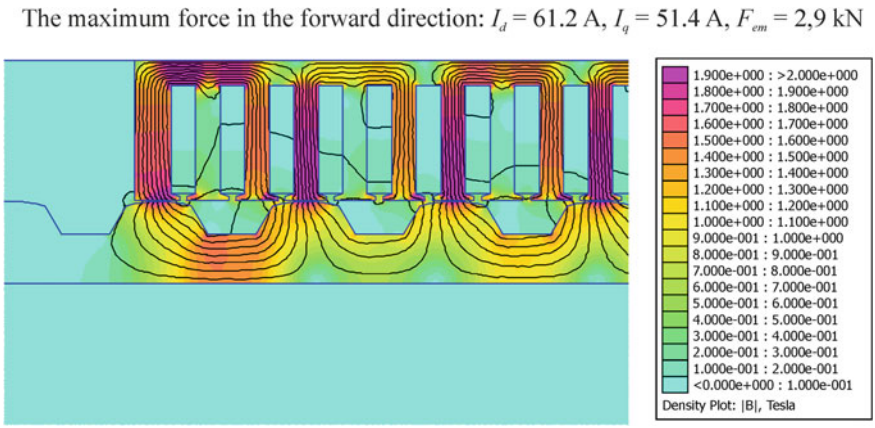


Fig. 3 Magnetic field distribution in a linear reactive reluctance-flux electrical machine

(fragment) under consideration at the position of the moving element and the currents ratio in the stator windings giving the maximum electromagnetic force.

Figure 4 shows the dependences of the flux linkage of the equivalent windings located along the d and q axes, and their inductance determined on the basis of the results of calculation of the magnetic field distribution.

Based on these parameters using Eqs. (1)–(3) the calculation of the dependence of the electromagnetic force acting on the moving element in dependence of its position and the comparison with the calculation results obtained by using the field theory were performed (Fig. 5).

Comparison of the results shows that the mathematical model based on the use of Park equations adequately describes the electromagnetic processes in this linear reactive reluctance-flux electrical machine.

On this basis, the use of vector control methods for controlling the electric machine under consideration can be considered as valid.

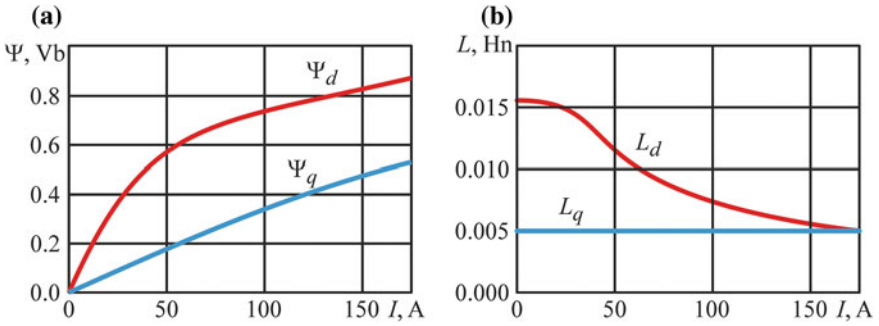


Fig. 4 Dependence of flux linkage (**a**) and windings induction (**b**) located along the $d - q$ axes

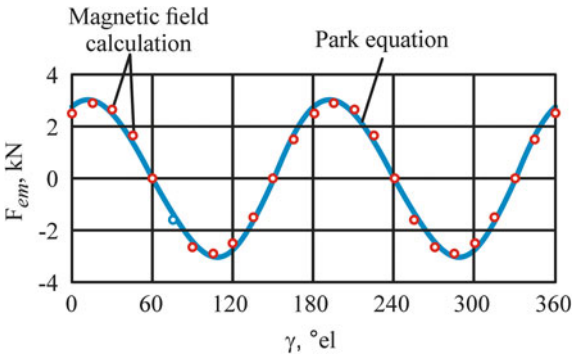


Fig. 5 The calculation results of the electromagnetic force acting on the moving element

5 Control Problem Formulation

The efforts adjustment system could agree the following principles. The stator current control on *d*-axis is difficult, since the inductance of the electrical machine along this axis has a significant value. The control with high speed requires considerable reserve of voltage. Therefore, the method of regulation when the stator current is stabilized on the *d*-axis at a predetermined level, and force control performed by changing the current on the *q*-axis is rational. *q*-axis inductance is substantially low, stator current and efforts regulation along this axis can be performed with high speed.

On the basis of accepted approach to power control system structure, we can distinguish two parallel operating channels. One of them is a current control channel on the *d* axes. The other is the channel of the power control by changing the *q*-axis current. To compensate the currents cross impact on *d* and *q* axes the cross-link's compensation unit is entered into the system. Formation of the output voltage of the three-phase inverter is carried out by using pulse-width modulation [16].

Two external control circuits are incorporated in the control system structure for regulating the speed and position of the moving element.

Figure 6 shows a block diagram of the control system of a linear reluctance-flux electrical machine created with the outlined provisions.

6 Simulation Results and Analysis

The studies of electromechanical processes in the "linear reactive electrical machine —a free-piston engine" system were carried out using the outlined mathematical model. Figure 7 show graphs of changing the current and voltage components along the *d* and *q* axes, an electromagnetic force and the engine piston power, the linear movement of the moving element.

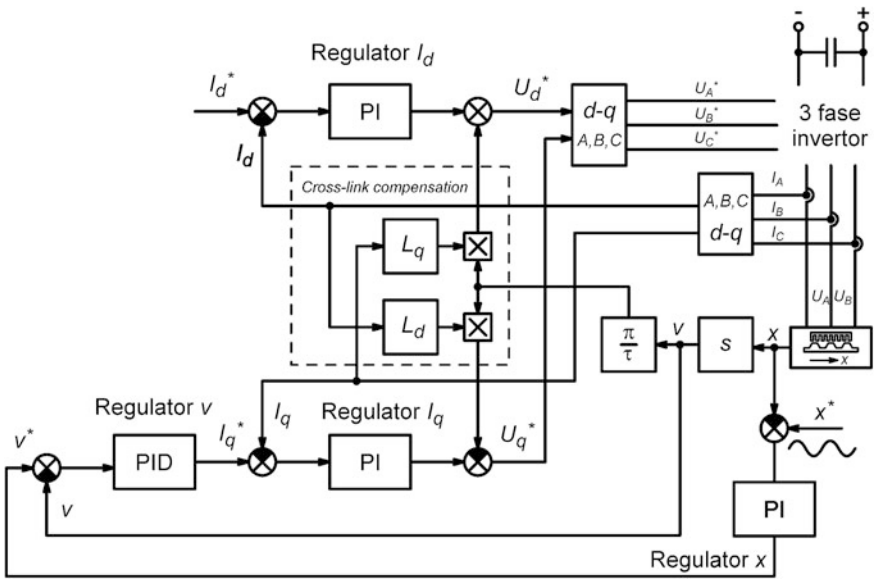


Fig. 6 Block diagram of the automatic control system

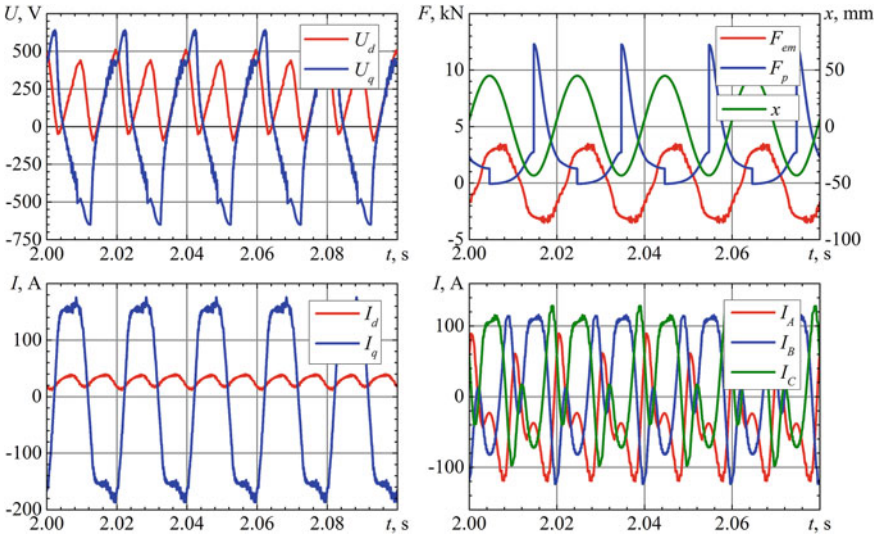


Fig. 7 The process modeling results in the "linear reactive electrical machine—a free-piston engine" system: U_d, U_q, I_d, I_q—windings voltage and current (components arranged on $d-q$ axes); I_A, I_B, I_C—false current; F_{em}, F_p—electromagnetic force and the force acting on the piston; x—the moving element displacement

The calculations were performed for the frequency of 50 Hz. The electric generator output power is 14 kW in this mode. The electrical machine makes full use of current and voltage in this mode. The results show that the use of vector control methods allows controlling with good dynamics the reactive reluctance-flux electrical machine. The use of proposed principles of the automatic control system design allows to implement the control of the "linear reactive electrical machine—a free-piston engine" system with high-energy performance.

In this example reactive reluctance-flux electrical machine with voltage limitation [16]. Therefore, the stator current control channel on *d*-axis can not stabilize the current at a predetermined level in the time intervals when displacement velocity reaches the maximum. Analysis of the simulation results showed that this effect does not lead to a breach of the system functionality in this example. It is achieved by increasing the current component along the *d*-axis [14, 16, 17]. By increasing the operating frequency, the stator current component on the *d*-axis is reduced and electrical machine goes into weak field mode. The limit operating frequency of the electrical machine with permanent magnets is limited by the EMF generated by the magnets. The limit frequency can be much increased in reactive electrical machine, which allows changing in a wide range the parameters of the internal combustion engine working cycle.

7 Conclusions

Analysis of publications let confirm that the use of free-piston engines and generators on their base has advantages and prospects of use for the mobile systems of the autonomous power supply. The conjugation of linear electrical machine and free-piston engines with a gas spring (one cylinder) or an opposite scheme (two cylinders) is rational. The single-piston schemes are most often used.

The reactive electric machine with three-phase distributed winding is an alternative to electric machine with permanent magnets on the moving element. The processes in the selected type of the electrical machine are described by Park equations with sufficient accuracy; it makes possible the use of vector control principles to control the force. The stabilization of the stator current in a *d*-axis and the power regulation by changing the stator current in *q*-axis is a rational method of the linear reactive electrical machine control. The obtained simulation results show that the used method of efforts regulation of electrical machine under investigation allows organizing reliable operating of the "linear reactive electrical machine—free-piston engine" system.

Acknowledgments The work is done by the authors as part of the agreement No 14.579.21.0064 about subsidizing dated 20.10.2014. The topic is "Development of experimental model of reversible electric machine reciprocating power 10–20 kW heavy-duty" by order of the Ministry of Education and Science of the Russian Federation, Federal Targeted Programme (FTP) "Researches and developments in accordance with priority areas of Russian science and technology sector evolution for 2014–2020 years". The unique identity code of applied researches (the project) is RFMEFI57914X0064.

References

1. van Blarigan, P.: Advanced internal combustion electrical generator. In: Proceeding 2002 U.S. DOE Hydrogen Program Review, NREL/CP-610-32405, pp. 1–16 (2002)
2. Mikalsen, R., Roskilly, A.P.: The design and simulation of a two-stroke free-piston compression ignition engine for electrical power generation. Appl. Therm. Eng. **28**(5–6), 589–600 (2008)
3. Mikalsen, R., Roskilly, A.P.: A computational study of free-piston diesel engine combustion. Appl. Energy **86**(7–8), 1136–1143 (2009)
4. Němeček, P., Šindelka, M., Vysoký, O.: Control of two-stroke free-piston generator. In: The 6th Asian Control Conference. Bali, Indonesia (2006). ISBN: 979-15017-0
5. Němeček, P., Šindelka, M., Vysoký, O.: Modeling and control of linear combustion engine. In: IFAC Symposium on Advances in Automotive Control, pp. 320–325. Elsevier Science, Oxford (2004)
6. Mikalsen, R., Roskilly, A.P.: The fuel efficiency and exhaust gas emissions of a low heat rejection free-piston diesel engine. Proc. IMechE Part A: J. Power Energy **223**, 379–384 (2009)
7. Mikalsen, R., Roskilly, A.P.: The control of a free-piston engine generator. Part 1: fundamental analyses. Appl. Energy **87**, 1273–1280 (2010)
8. Mikalsen, R., Roskilly, A.P.: The control of a free-piston engine generator. Part 2: engine dynamics and piston motion control. Appl. Energy **87**, 1281–1287 (2010)
9. van Blarigan, P.: Advanced internal combustion electric generator. In: Proceedings of the 2001 DOE Hydrogen Program Review. NREL/CP-570-30535
10. Kosaka, H., Akita, T., Moriya, K., Goto, S. et al.: Development of Free Piston Engine Linear Generator System Part 1—Investigation of Fundamental Characteristics, SAE Technical Paper 2014-01-1203 (2014). doi:10.4271/2014-01-1203
11. Goto, S., Moriya, K., Kosaka, H., Akita, T. et al.: Development of Free Piston Engine Linear Generator System Part 2—Investigation of Control System for Generator, SAE Technical Paper 2014-01-1193 (2014). doi:10.4271/2014-01-1193
12. Mikalsen, R., Roskilly, A.P.: A review of free-piston engine history and applications. Appl. Therm. Eng. **27**(14–15), 2339–2352 (2007)
13. De Doncker, R.W.A.A., Pulle, D.W.J., Veltman, A., SpringerLink (Online service): Advanced Electrical Drives: Analysis, Modeling, Control. Springer, Dordrecht (2011)
14. John, C.: Modeling and High Performance Control of Electric Machines. IEEE Press Series on Power Engineering, vol. 26, p. 736. Wiley (2005)
15. Lipo, T.A.: Analysis of Synchronous Machines, p. 590. CRC Press, Hoboken (2012)
16. Holmes, G., Lipo, T.A.: Pulse Width Modulation for Power Converters: Principles and Practice. Chichester, New York (2002)
17. Glumineau, A., de Leon Morales, J.: Sensorless AC Electric Motor Control: Robust Advanced Design Techniques and Applications, p. 244. Springer (2015)
18. http://www.femm.info/wiki/HomePage

One Model of Two-Sided Price Coordination on the Transport Market

Nikolay Goryaev, Konstantin Kudryavtsev and Sergey Tsiulin

Abstract A mathematical model of determining the tariff systems of information exchange between transport market participants is offered. The algorithm called "the method of minimal concessions" allows to determinate the tariff automatically. This article presents a comparison of simulated transportation tariffs and prices for real applications; it shows the adequacy of the proposed algorithm and the process of harmonization of prices between customers and transport carriers. Application of game theory allows to automate the process of concluding carriage contracts, and to improve significantly the efficiency of information exchange systems of the transport market participants.

Keywords Mathematical model · Transport · Information system · Method of minimal concessions · Mechanism · Matching contract

1 Introduction

During the organization process of long-distance haulage the determination of cargo is always important for making complete routes. The way of making this research belongs to information systems, which are used in cooperation between stakeholders on the transport market. Recent studies by Goryaev [1] have shown that using these systems on the Russian market allows completing more than 100 thousand daily freight requests. However, these systems are unable to make the process automatic due to big variance in pricing. An example (July 3, 2014) is

N. Goryaev · K. Kudryavtsev (✉) · S. Tsiulin
South Ural State University, Chelyabinsk, Russia
e-mail: kudriavtcevkn@susu.ru; kudrkn@gmail.com

N. Goryaev
e-mail: goriaevnk@susu.ru

S. Tsiulin
e-mail: ciulinane@susu.ru

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_42

Table 1 Example of pricing facilities on the Moscow-Chelyabinsk direction

Client	Client's offer, rubles	Carrier	Carrier's offer, rubles
Chel-dostavka, LTD	72000 (41,2 r/km)	Energiya, LTD	63000
Bogomolov Oleg, SP	77000 (43,2 r/km)	Aspekt, LTD	78000
VITAR, LTD	72000 (41,2 r/km)	Gruzovue perevozki, LTD	73000
Bogomolov Yan, SP	69000 (39,5 r/km)	Moyi malenkiyi mir, LTD	72000
Yanus-Expediciya, LTD	43000 (24,6 r/km)	VIS, LTD	90000
Kostrizin Aleksandr, SP	74000 (42,3 r/km)	National Logistics Group, LTD	87000
Agat, LTD	72000 (41,2 r/km)	Auto-411, LTD	79000
SZSK №7, LTD	74000 (42,3 r/km)	TK7, LTD	87000

shown in the Table 1. Overall, this tendency is decreasing the intercity freight efficiency in the Russian Federation (Goryaev, [2]).

According to the data represented below (Table 1), the average pricing value of haulage from Moscow to Chelyabinsk for clients is 69,125 rubles, for carriers—78,625 rubles. In addition, the minimum client's offer was more than 2 times less than the maximum carrier's offer.

In order to automate the process of price determination we present a method of minimal concessions which is described below.

2 Requirements to the Contracting Process

How the automation process should work and be shown in conditions of close cooperation between stakeholders on the market? According to Hurwicz [3], mechanism is defined as a process of interaction between the subject and the center (arbiter) and consists of 3 stages:

1. Each subject i $(i = 1, \ldots, n)$ determines the message x_i transmitted to the arbiter;
2. The center calculates the expected result $Y = f(x_1, \ldots, x_n)$ after acquiring all n messages;
3. The center states the result Y and realizes it, if necessary.

Thereby, the mechanism of automated contracts between stakeholders can be formalized as a non-cooperative strategy game. There are two types of participants in this game: carriers and clients. The participants' strategy presents offering preliminary contract prices. The game goes as follows: each player makes an offer—preliminary price of the contract, without discussing it with anyone, whereby a situation is formed (profile of strategies)—a vector composed by the proposed prices. Using this profile of strategies, the center defines a number of matching (pairs between which the contract will be concluded), where a customer is matched to a carrier and, according to some algorithm, generates the contract price for each

pair. The goal of each participant (carrier) is the selection of preliminary price where the price can be as minimum (maximum) as possible.

The algorithm controlling the center must be absolutely fair. In detail, none of the participants should be subjected to disproportional (from his personal point of view) deviation from the assigned initial price. If the condition fails, "unsatisfied" players will simply refuse to participate in the proposed system. On the other hand, the center has to eliminate (or at least minimize) the ability of individual players to manipulate the price (we call this property stability). It is necessary to mention once again, participants are unable to cooperate with each other, therefore, they are unable to manipulate the system.

3 Sustainable Matchings

The first work made by Gale and Shapley in 1962 [4], which started a fundamentally new task concerning the optimal choice of contracts, was published by David Gale and Lloyd Shapley [4]. In that article the concept of sustainable allocation of contracts, "matching", was introduced. The idea was developed in works published by Alvin Roth et al. [5], Sotomayor and Roth [6], Gale and Sotomayor [7], Hatfield and Milgrom [8], Perez-Casttrillo and Sotomayor [9], Sotomayor and Ozak [10] and many others. However, the majority of these studies represent only the mechanism of sustainable allocation of contracts. The procedure of price definition was proposed by Perez-Casttrillo et al. [9]. Another procedure was proposed in Faratin [11]. This article offers the alternative way of price definition (on the logistics market).

Let A and B be two set of n elements. Elements of set A are clients $A_1, A_2, \ldots, A_n$, where set B includes all carriers $B_1, B_2, \ldots, B_n$. A match is a bijection from A into B, i.e. set of n contracts [12].

Let us assume that each client has a list of carriers ordered by descending their "attractiveness" for him. Similarly, each carrier has a list of customers, ordered by the same principle. Thus, the preference of both sides is represented by two matrices with the size $n \times n$.

Matching can be called unsustainable if a carrier and a client, while making the contract, choose another partners. Otherwise, in case of reciprocal arrangement from both sides, matching can be called sustainable [12].

Let each participant in tender A_i $(i = 1, \ldots, n)$ bid his own provisional price for a contract a_i, and, at the same time, each carrier b_j makes his own provisional price B_j $(j = 1, \ldots, n)$. Furthermore, we assume that the conditions can be confirmed as:

$$\begin{cases} [i \neq j] \quad \Rightarrow \quad [a_i \neq a_j] \wedge [b_i \neq b_j] \quad \forall i, j = 1, \ldots, n, \\ a_i < b_j \quad \forall i, j = 1, \ldots, n. \end{cases} \tag{1}$$

The meaning of the assumption (1) is that all clients and carriers bid different prices where carriers' prices are always higher than the clients' ones.

To maintain generality, we assume that the condition is

$$\begin{array}{c} a_1 > a_2 > \cdots > a_{n-1} > a_n, \\ b_1 < b_2 < \cdots < b_{n-1} < b_n, \end{array} \tag{2}$$

otherwise, bidders can always be renumbered.

As a measure of "attractiveness" between a client A_i and a carrier B_j we can use the difference between the initial contract prices

$$w_{ij} = a_i - b_j.$$

In fact, this means that each client A_i $(i = 1, \ldots, n)$ will prefer to deal with a carrier whose price is the lowest. At the same time, each carrier B_j $(j = 1, \ldots, n)$ will prefer to deal with a client, whose price is higher.

Giving the fact of conditions (1) and (2), sets A and B are strictly separated by preference. Hence, according to Gale's and Shapley's theory, there is a sustainable matching for them.

Proposition 1 *If conditions (1) and (2) for carriers and clients are satisfied, then the sustainable matching would be:*

$$(A_1B_1, A_2B_2, A_3B_3, \ldots, A_nB_n) \tag{3}$$

As it follows from the Proposition 1, the contracts signed in pairs A_iB_i $(i = 1, \ldots, n)$ can be called sustainable. For these pairs there is a need to define the price of the signed contract. Obviously, the average calculated price will be more "fair" in the contract between a client A_i and a carrier B_i:

$$\varepsilon_i = \frac{a_i + b_i}{2},$$

As far as each side of the contract goes with the same step as another one, which is equal to half of the difference $b_i - a_i = -w_{ii}$, so both parties share the calculated price equally.

However, there is one disadvantage in the approach presented above—instability of the price formation. It can be avoided by using algorithm for pricing contracts presented below. This method is called "the method of minimal concessions", which is based on the following rule: *on each stage of the bidding the price should be corrected in half of the difference of the price as far as the price of the most preferable partner is reached.*

4 The Method of Minimal Concessions

We can assume below that sets of clients $A = \{A_1, A_2, \ldots, A_n\}$ and carriers $B = \{B_1, B_2, \ldots, B_n\}$ are set. Each carrier and client at the same time announce their preliminary price of contract independently from each other, where we get lots of sets $\{a_1, a_2, \ldots, a_n\}$ and $\{b_1, b_2, \ldots, b_n\}$ of preliminary prices from carriers and clients as a result. Also, we suppose that conditions (1) and (2) are satisfied.

For formation of sustainable matching (3) and defining the price of signed contracts the following algorithm can be used.

The algorithm of minimal concessions

Step 1. For each client A_i $(i = 1, \ldots, n)$ the contract price is set

$$a_i(1) = a_i - \frac{w_{i1}}{2} = a_i + \frac{b_1 - a_i}{2} = \frac{a_i + b_1}{2}.$$

For each carrier B_j $(j = 1, \ldots, n)$ the price of contract is appointed

$$b_j(1) = b_j + \frac{w_{1j}}{2} = b_j - \frac{b_j - a_1}{2} = \frac{a_1 + b_j}{2}.$$

Herewith the prices of client A_1 and carrier B_1 are matched, hence, the new contract is going to be signed in pair A_1B_1 with the following price:

$$\gamma_1 = a_1(1) = b_1(1) = \frac{a_1 + b_1}{2}.$$

Step 2. For each client A_i $(i = 2, \ldots, n)$ it is more preferable to sign the contract with B_2 therefore, the price for them sets as:

$$a_i(2) = a_i(1) + \frac{b_2(1) - a_i(1)}{2} = \frac{a_i(1) + b_2(1)}{2}.$$

Similarly, for each carrier B_j $(j = 2, \ldots, n)$ client is more preferable, therefore, for their contract the price is appointed as below:

$$b_j(2) = b_j(1) - \frac{b_j(1) - a_2(1)}{2} = \frac{a_2(1) + b_j(1)}{2}.$$

As a result of the second stage, the conceded price is $a_2(2)$, assigned to the client A_2, and a certain price for carrier B_2 the price of the contract $b_2(2)$. Hence, we have the second contract in pair A_2B_2 with the price:

$$\gamma_2 = a_2(2) = b_2(2) = \frac{a_2(1) + b_2(1)}{2}.$$

All the next steps follow the scheme which has been used for the first and the second stages, and repeat $n-2$ more times.

Step k (k = 3, …, n). For each client A_i $(i = k, \ldots, n)$ the price of contract is set:

$$a_i(k) = a_i(k-1) + \frac{b_k(k-1) - a_i(k)}{2} = \frac{a_i(k-1) + b_k(k-1)}{2}.$$

For each carrier B_j $(j = k, \ldots, n)$ the price of contract is appointed

$$b_j(k) = b_j(k-1) - \frac{b_j(k-1) - a_k(k-1)}{2} = \frac{a_k(k-1) + b_j(k-1)}{2}$$

In pair A_kB_k the contract is signed with the price of

$$\gamma_k = a_k(k) = b_k(k) = \frac{a_k(k-1) + b_k(k-1)}{2}.$$

This way during the n stages the algorithm forms a sustainable matching (3) and defines the price for each contract from (3).

Remark 1 The proposed algorithm, as shown in the Example 1, can be applied in case if the 3rd inequality from the (1) condition becomes wrong:

$$a_i < b_j \quad \forall i, j = 1, \ldots, n,$$

If preliminary prices of some clients will be higher than prices announced by certain carriers.

Remark 2 If the first two implications from the condition (1) are wrong

$$[i \neq j] \quad \Rightarrow \quad [a_i \neq a_j] \quad \forall i, j = 1, \ldots, n,$$
$$[i \neq j] \quad \Rightarrow \quad [b_i \neq b_j] \quad \forall i, j = 1, \ldots, n,$$

thus, if there are equal offers among preliminary prices from carriers and clients, it acts on sets of preferences in the lax order. In this case we assume that among 2 equal offers we will accept that one which was received earlier.

As an example of the algorithm we define contracts by the data presented in table below.

Example 1 After making the price list from the Table 1 we can define sets of clients A and carriers B (Table 2).

Next, we apply the algorithm of minimal concessions (Table 3).

Table 2 Defining clients A and carriers B

i	A_i	a_i	B_i	b_i
1	Bogomolov Oleg, SP	77000	Energiya, LTD	63000
2	Kostrizin Aleksandr, SP	74000	Moyi malenkiyi mir, LTD	72000
3	SZSK №7, LTD	74000	Gruzovue perevozki, LTD	73000
4	Chel-dostavka, LTD	72000	Aspekt, LTD	78000
5	VITAR, LTD	72000	Auto-411, LTD	79000
6	Agat, LTD	72000	National Logistics Group, LTD	87000
7	Bogomolov Yan, SP	69000	TK7, LTD	87000
8	Yanus-Expediciya, LTD	43000	VIS, LTD	90000

The comparison of our results with actual contracts shows closeness between the price made by the algorithm and prices on the market. The average price set in the company LTD "YUTEP" is about 70,222 rubles considering standard deviation of 878 rubles. Also, minimal price is 69000 rubles, maximum—72000 rub. (in total, by the 3rd of July, 2014 LTD "YUTEP" had registered 18 contracts).

The next example shows the sustainability of the method of minimal concessions to price manipulation.

Example 2 Let carrier B_8 from the example 1, in order to influence the final contract (contract A_8B_8), offer the price two times higher than his initial offer $b_8 = 180000$ rubles. The price of signed contract will be $\gamma_8 = 72207,03$ rubles.

Remark 3 In a real working system the quantity of applications from carriers and clients is not the same. The sustainable matching followed by the algorithm, step by step forms applications similar to each other. Thus, every attempt of manipulation on the market will be threatened by high risks of not signing the contract (kicking out of any matching).

Remark 4 In case of interruption into 3rd inequality from the condition (1)

$$a_i < b_j \quad \forall i,j = 1,\ldots,n \tag{4}$$

(preliminary prices of some clients can be higher than prices, announced by some of the carriers), but there is a way of manipulation from the different side.

A carrier can announce a lower price in order to "punish" his rivals, it can decrease the average price for the rest of participants. Similarly, the same thing can happen with clients. The way to exclude manipulations is, for example, to ban the access (4) to the market and to apply method of minimal concessions for these participants.

The structure of assigning prices can be seen via algorithm of minimal concessions.

Table 3 Applying the algorithm of minimal concessions

Step	Client price	Carrier price	Contract	Price of contract
1	$a_1(1) = 70000$ $a_2(1) = 68500$ $a_3(1) = 68500$ $a_4(1) = 67500$ $a_5(1) = 67500$ $a_6(1) = 67500$ $a_7(1) = 66000$ $a_8(1) = 53000$	$b_1(1) = 70000$ $b_2(1) = 74500$ $b_3(1) = 75000$ $b_4(1) = 77500$ $b_5(1) = 78000$ $b_6(1) = 82000$ $b_7(1) = 82000$ $b_8(1) = 83500$	A_1B_1	$\gamma_1 = 70000$
2	$a_2(2) = 71500$ $a_3(2) = 71500$ $a_4(2) = 71000$ $a_5(2) = 71000$ $a_6(2) = 71000$ $a_7(2) = 70250$ $a_8(2) = 63750$	$b_2(2) = 71500$ $b_3(2) = 71750$ $b_4(2) = 73000$ $b_5(2) = 73250$ $b_6(2) = 75250$ $b_7(2) = 75250$ $b_8(2) = 76000$	A_2B_2	$\gamma_2 = 71500$
3	$a_3(3) = 71625$ $a_4(3) = 71375$ $a_5(3) = 71375$ $a_6(3) = 71375$ $a_7(3) = 71000$ $a_8(3) = 67750$	$b_3(3) = 71625$ $b_4(3) = 72250$ $b_5(3) = 72375$ $b_6(3) = 73375$ $b_7(3) = 73375$ $b_8(3) = 73750$	A_3B_3	$\gamma_3 = 71625$
4	$a_4(4) = 71812,5$ $a_5(4) = 71812,5$ $a_6(4) = 71812,5$ $a_7(4) = 71625$ $a_8(4) = 70000$	$b_4(4) = 71812,5$ $b_5(4) = 71875$ $b_6(4) = 72375$ $b_7(4) = 72375$ $b_8(4) = 72562,5$	A_4B_4	$\gamma_4 = 71812,5$
5	$a_5(5) = 71843,75$ $a_6(5) = 71843,75$ $a_7(5) = 71750$ $a_8(5) = 70937,5$	$b_5(5) = 71843,75$ $b_6(5) = 72093,75$ $b_7(5) = 72093,75$ $b_8(5) = 72187,5$	A_5B_5	$\gamma_5 = 71843,75$
6	$a_6(6) = 71968,75$ $a_7(6) = 71921,88$ $a_8(6) = 71515,63$	$b_6(6) = 71968,75$ $b_7(6) = 71968,75$ $b_8(6) = 72015,63$	A_6B_6	$\gamma_6 = 71968,75$
7	$a_7(7) = 71945,31$ $a_8(7) = 71742,19$	$b_7(7) = 71945,31$ $b_8(7) = 71968,75$	A_7B_7	$\gamma_7 = 71945,31$
8	$a_8(8) = 71855,47$	$b_8(8) = 71855,47$	A_8B_8	$\gamma_8 = 71855,47$

Statement 2. Contract prices, which were formed after the process of minimal concessions, can be calculated by the recurrence formula:

$$\begin{aligned} \gamma_1 &= \frac{a_1 + b_1}{2}, \\ \gamma_k &= \frac{a_k + b_k}{2^k} + \sum_{m=1}^{k-1} \frac{\gamma_{k-m}}{2^m} \quad (k = 2, \ldots, n). \end{aligned} \tag{5}$$

As it is shown (5), during increasing number k $(k = 1, \ldots, n)$ signed in pair contract $A_k B_k$ influence preliminary prices a_k and b_k goes down promptly. In fact, that decreases the chance of price manipulation by bidding unreasonably low/high prices to minimum.

5 Conclusion

The proposed "method of minimal concessions" allows defining sustainable "client-carrier" matching and automates the order of price determination. In future, we are planning to study the Nash equilibrium strategies of carriers and customers.

References

1. Goryaev, N.K.: Acquisition's efficiency of transport services on the basis of the operative tender. Trans. Urala. **3**(26), 17–19 (2010)
2. Goryaev, N.K.: The effectiveness of long-distance haulage in the context of market reforms in Russia. In: Procedia—Social and Behavioral Sciences, vol. 54, pp. 286–293 (2012). doi:10.1016/j.sbspro.2012.09.747
3. Hurwicz, L.: Optimality and informational efficiency in resource allocation processes. In: Arrow, K.J., Karlin, S., Suppes, P. (eds.) Mathematical Methods in the Social Scienes, pp. 27–46. Stanford University Press, Stanford (1960)
4. Gale, D., Shapley, L.: College Admissions and the Stability of Marriage. Am. Math. Mon. **69**, 9–15 (1962)
5. Roth, A.E.: The college admissions problem is not equivalent to the marriage problem. J. Econ. Theory **36**(2), 277–288 (1985). doi:10.1016/0022-0531(85)90106-1
6. Roth, A.E., Sotomayor, M.A.O.: Chapter 16. Two-sided matching. In: Handbook of Game Theory with Economic Application, vol. 1, pp. 485–541 (1992). doi:10.1016/S1574-0005(05)80019-0
7. Gale, D., Sotomayor, M.: Some remarks on the stable matching problem. Discret. Appl. Math. **11**(3), 223–232 (1985)
8. Hatfield, J.W., Milgrom, P.R.: Matching with contracts. Am. Econ. Rev. **95**(4), 913–935 (2005). doi:10.1257/0002828054825466
9. Perez-Casttrillo, D., Sotomayor, M.: A simple selling and buying procedure. J. Econ. Theory **103**(2), 461–474 (2002). doi:10.1006/jeth.2000.2783
10. Sotomayor, M., Ozak, O.: Two-Sided Matching Models. In: Meyers, R.A. (ed.) Computational Complexity, pp. 3234–3257. Springer New York (2012)
11. Faratin, P.: Economics of overlay networks: an industrial organization perspective on network economics. In: Proceedings of the Joint Workshop on The Economics of Networked Systems and Incentive-Based Computing (NetEcon+IBC), in conjunction with ACM Conference on Electronic Commerce (EC'07), San Diego, California, 11 June 2007 (2007)
12. Knut, D.E.: Sustainable Matching and others Combinatory Tasks. Introduction into the Mathematical Analysis of Algorithms. MCNO, Moscow (2014)

Correlated Colour Temperature Changes of the LED Sources Depending on the Angle of Their Radiation

Tomáš Novák, Petr Bos, Richard Baleja and Karel Sokanský

Abstract The aim of the report is a description of an issue connected with subjective perception of colours from the light of the lighting systems stepped by LED. Subjective observation of the colour change on the light of the lighting systems stepped by LED has been mostly related to extensive lighting systems where it was possible to observe identical light source (LED) within various angles. This phenomenon was manifested even more distinctively within lighting systems that made use of optical systems located in luminaries (reflector, diffusers) in a minimal way. Additional optical systems redistribute primary luminous flux coming from light source (LED) and thus there is a blending of luminous flux radiated to a certain angle of the illuminated space from various directions of the luminous flux radiation itself. In case it is not like this and the luminaries are equipped only with LED with a suitable luminous distribution curve, there is no blending of the luminous flux. Thanks to this the observer can also perceive various colours of light within individual luminaries perceived under various angles.

Keywords LED sources · Changes in correlated colour temperature · Luminophor · The radiating angle

1 Introduction

The article is aimed on the issue of dependence of temperature changes within the correlated colour temperature of the LED light sources depending on the angle of their radiation. Each LED source is of various technical characteristics and

T. Novák (✉) · P. Bos · R. Baleja · K. Sokanský
Faculty of Electrical Engineering and Computer Science,
VŠB - Technical University of Ostrava, Ostrava, Czech Republic
e-mail: tomas.novak1@vsb.cz

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_43

construction and due to radiation angle some LED sources can have smaller or more extensive changes within correlated colour temperature. These changes cause a change in light colour which can lead to a collision with catalogue of the given LED source and of course induce negative subjective perception of light from lighting systems stepped by these LED sources.

Colour of the light of LED sources depends not only on a type of a luminescent material but also on a length that has to be overcome by a ray coming through the luminophor. The length can be defined by means of an angle of the normal, i.e. from a ray coming through the luminophor upright and having the shortest track. The longer the length of the ray coming through the luminophor the wider differences in it´s colour. This phenomenon can influence sensation visual in a negative way. As it results from the article, the size of colour change is also influenced by a luminophor type. If the luminaires with LED diodes have diffusers, there is a partial blending of light and the colour change is smaller.

1.1 LED Construction

LED sources belong among electroluminescent light sources. They are semiconductor components containing PN junction. Through excitation by electric current, on PN junction there is an emission of optical radiation. LED notation comes from the term Light Emitting Diode. These light sources can be very small however they reach relatively high performance.

Nowadays materials with high cleanness are used for creation of semiconductor PN junction. These materials are alloyed by a small amount of suitable ingredients. These ingredients cause either excess of electrons (materials of N type) or conversely their deficiency and thus excess of holes (materials of P type). At the place of contact of both semiconductor types there appears PN junction. By applying direct voltage of a correct polarity on this conversion there appears a recombination of electrons and holes and a certain amount of energy is released. This energy is radiated out of the semiconductor material. Electric energy is thus directly changed into light of particular colour according to the chosen type of LED. The choice of a semiconductor determines a zone of the spectrum of radiation that can be almost random, i.e. from ultraviolet up to infrared [1, 2] (Fig. 1).

LED principle implies that they cannot radiate white light directly. That is why in the past an additive blending of radiation of three coloured RGB chips was used. By means of their combination a necessary colour of light was obtained. Another (and most widely used) possibility of getting white light is making use of phosphorescence

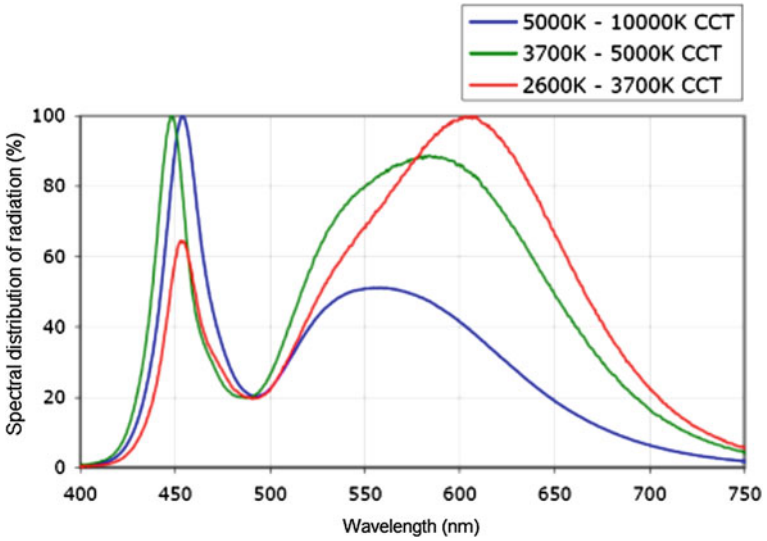
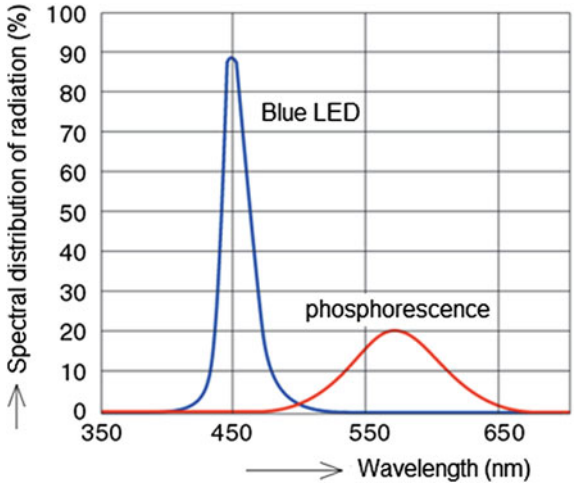


Fig. 1 Spectral radiation of white LED with various correlated colour temperature [6]

Fig. 2 Sample of emission spectrum of white LED [7]



of a luminophor applied on LED chip, see Fig. 2, which transforms a part of emitted blue light from the LED into higher wave lengths with a result of white colour of light in shades varying from cold white (6000–7000 K) up to warm white (2700–3300 K). Figure 1 presents demonstration of emission spectrum of white diodes with various values of correlated colour temperature [2–5].

Fig. 3 LED SEOUL SSC - Z5 (on the *left*) and Fortimo LED DLM Module (on the *right*) [8]

Table 1 Catalogue data for used LED sources

Parameter Type of LED	Luminous flux (lm)	Correlated colour temperature (K)	Colour rendering index	Power (W)
SEOUL SSC - Z5 SZW05A0B	124	6000	70	1
Fortimo LED DLM Module	3000	4000	80	44

2 Catalogue Data of Applied LED

Two various LED sources were used for evaluation of the change in correlated colour temperatures within LED sources depending on an angle of their radiation. The first LED labelled SEUOL SSC Z5 SZW05A0B is used for day light in cars. Second one used LED sources came from Philips company, i.e. Fortimo LED DLM for interior usage (Fig. 3 and Table 1).

3 Verification of the Prediction Model

This chapter points out a principle what changes in colour of light within LED light sources can occur. As mentioned above, there is a various length of the ray transmission through the luminophor ensuring a shift of primary blue radiation into areas of higher wavelengths. Due to increase of the length of the ray transmission through the luminophor there is a transformation of larger content of primary radiation by means of phosphorescence, see Fig. 2, and thus there is also decrease of correlated colour temperature within higher angles of radiation. The following relation for calculation of transmission lengths of light ray through the luminophor was used. Its form is:

$$l_n = \frac{l_1}{cos_{\alpha_n}} [\%] \tag{1}$$

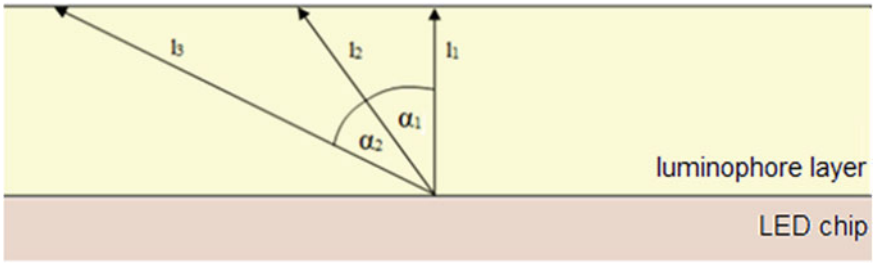


Fig. 4 Light ray transmission through the luminophor layer

l_1 … length of light ray transmission through the luminophor in vertical direction
α_n … angle of LED turning on the photogoniometer

Length l_1 is equal to thickness of the luminophor layer in the right section. As we do not know its precise value, we depict its size as a percentage of thickness of the particular layer, i.e. 100 %. We change the angle α_n with a step 5° (Fig. 4).

The above figure shows a significant relationship between the length of light ray transmission and the angular displacement of the photometric form and it causes change in values of correlated colour temperature. Ideally, there would be constant values of correlated colour temperatures within a change of angular displacement of the LED source. Theoretically, it would be possible to reach such a status if there was variable thickness of the luminophor layer. This problem could be solved for example by a preparation or an equipment that would transform the outgoing light ray in a desired way.

4 Spectral Characteristic of Individual LED

The aim of the measurement of the changes in correlated colour temperature depending on the radiating angle is to verify assumed parameters of the mentioned LED. During the measurement the following values were recorded: spectral characteristic of LED, values of luminous flux and correlated colour temperatures. The aim of the measurement is to find out if there is a change in correlated colour temperature during the change of the radiation angle of the LED source and also the scope of this temperature change.

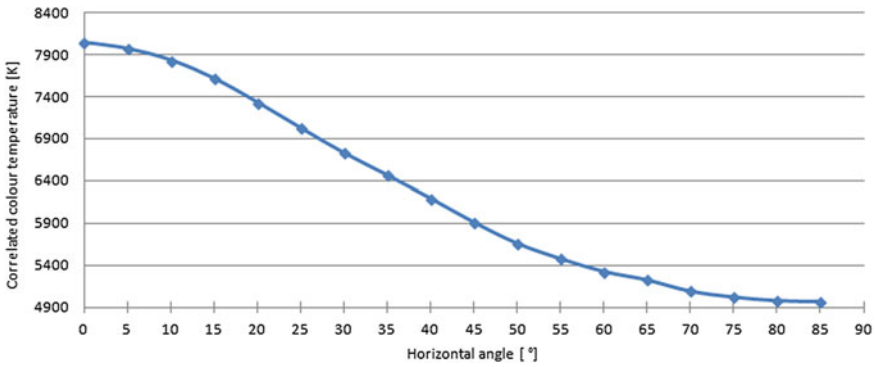


Fig. 5 The changes in correlated colour temperature with regard to the changes of the horizontal angle LED SEOUL SSC - Z5

4.1 The Process of Changes in Correlated Colour Temperatures Depending on the Change of the Horizontal Angle LED SEOUL SSC - Z5 SZW05A0B

See Fig. 5.

4.2 The Process of Changes in Correlated Colour Temperatures Depending on the Change of the Horizontal Angle Fortimo LED DLM Module 44 W/840

As a result of the measured values we can compare the prescribed value of correlated colour temperature within the catalogue data with the measured value of colour temperature of the Ulbricht sphere (Fig. 6 and Table 2).

By means of the measurement the parameters of correlated colour temperatures set by the producer for individual LED sources were verified. From the above scheme it is evident that there are higher deviations for the sources LED SEOUL SSC - Z5 SZW05A0B and Fortimo LED DLM Module.

The graphs show that measured correlated colour temperatures for the source by Philips company are comparable with correlated colour temperatures measured in a straight line at an angle of 0°.

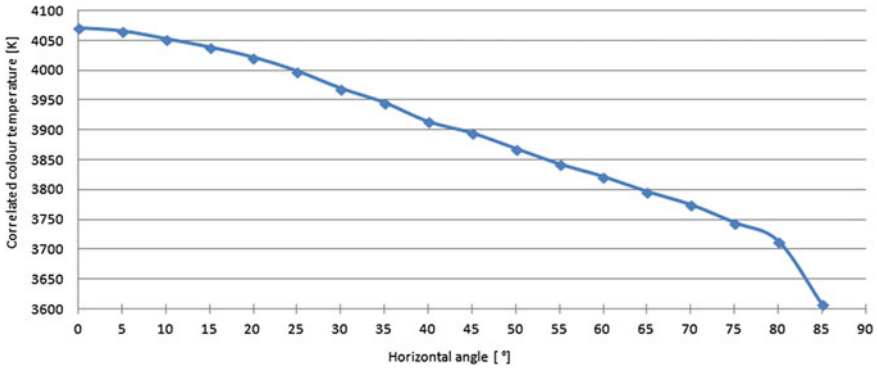


Fig. 6 The changes in correlated colour temperature with regard to the changes of the horizontal angle Fortimo LED DLM Module

Table 2 Comparison of catalogue and measured values of correlated colour temperature from the LED sources measured by the Ulbricht sphere

Parameter Type of LED	Catalogue T_{CP} (K)	Measured T_{CP} (K)
SEOUL SSC - Z5 SZW05A0B	6000	6416
Fortimo LED DLM Module	4000	3394

5 Changes in Correlated Colour Temperatures of LED Luminaries Depending on the Radiating Angle

The aim of the measurement is to point out changes in correlated colour temperatures of LED sources, i.e. by means of radiation angle change. The measurement was carried out at the photometric form and correlated colour temperatures were measured at the step 5° by means of the spectrophotometer JETI.

The figures below show the resulting comparison of dependence of correlated colour temperature on the radiating angle for individual types of LED sources. Individual graphs contain spectral characteristics at the angle of 0°, 85° and characteristics obtained by the measurement in the Ulbricht sphere with appropriate values of correlated colour temperatures.

5.1 LED SEOUL SSC - Z5 SZW05A0B

See Fig. 7.

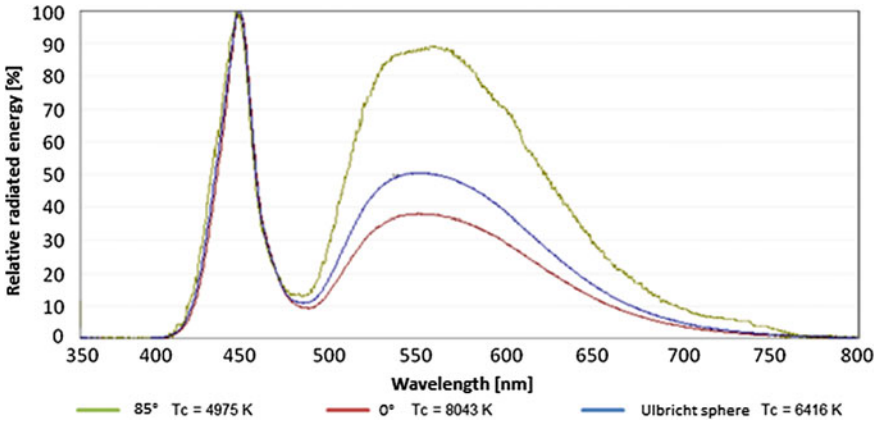


Fig. 7 Spectral characteristics for chosen types of angular displacement of LED and characteristics from the Ulbricht sphere with appropriate correlated colour temperatures

5.2 Fortimo LED DLM Module

The graphs show that due to the change of radiating angle there was the most significant change on the LED SEOUL SSC - Z5 SZW05A0B, where the correlated colour temperature changed by 3068 K and this change also caused a change in colour of light that was different from the one prescribed on the catalogue. On the second type of LED Fortimo DLM Module there were lower differences within correlated colour temperatures, i.e. by 677 K (Fig. 8).

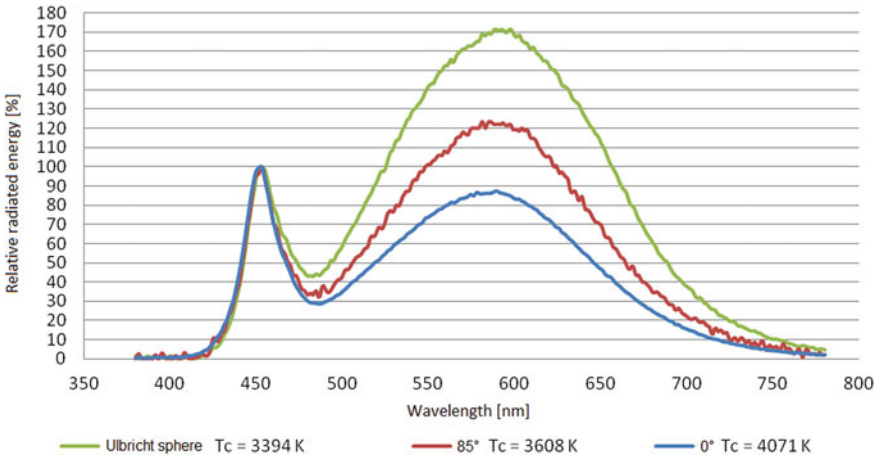


Fig. 8 Spectral characteristics for chosen types of angular displacement of LED and characteristics from the Ulbricht sphere with appropriate correlated colour temperatures

6 Conclusion

The comparison of spectral characteristics of chosen LED sources depending on angular displacement implies, based on executed measurements, a change of correlated colour temperature. In the graphs there are highest step changes of correlated colour temperatures for angular displacement of 0°, 85° and for measurement in the Ulbricht sphere. There is a conclusion for producers of luminaries—to secure a superior immixture of light coming from the luminary.

The LED SEOUL SSC - Z5 SZW05A0B experienced the most significant change in correlated colour temperature—8043 K was recorded as correlated colour temperature for radiating angle of 0° and 4975 K for the angle of 85°. Thus the difference was up to 3068 K and it caused a change of colour of radiated light. The second type of LED Fortimo DLM Module recorded lower differences in correlated colour temperature, i.e. only 677 K.

A supposed reason for correlated colour temperature is the difference in length within transmission of light ray during vertical and the largest angle of the luminophor. As a result of various lengths of transmission there can be extensive disparity of correlated colour temperature during changes of the radiating angle. This difference caused a change in correlated colour temperature on the order of 3000 K for LED SEOUL SSC - Z5 SZW05A0B.

Acknowledgments This article was prepared with the support of the project "Intelligent lighting control systems" SP2016/151, by institution of VSB-TU Ostrava.

References

1. IES Lighting Handbook, 10th edn. Illuminating Engineering, New York, Society of North America (2011)
2. Khan, T.Q., Bodrogi, P., Vinh, Q.T., Winkler, H.: LED Lighting: Technology and Perception. Wiley (2014). ISBN: 3527670173
3. Varfolomejev, L.P., Stěpanov, V.N., Rochlin, G.N.: Spravočnaja kniga po svetotěchnike. ISBN: 5-87789-051-4
4. Škoda, J., Baxant, P., Krbal, M., Sumec, S., Pavelka, T.: Photometry of LED sources. Przeglad Elektrotechniczny **89**(6), 341–344 (2013). ISSN: 00332097
5. Pavelka, T., Doležal, P., Sláma, P.: Parameters of white LEDs respecting operational conditions. Light 2015, Brno University of Technology, pp. 49–52. ISBN: 978-80-214-5244-2
6. Ensuring safety in LED lighting. http://www.electronicsweekly.com/news/products/led/ensuring-safety-in-led-lighting-2012-11/ (2012). Accessed 01 Mar 2016
7. Dvořáček, V.: Světelné zdroje - světelné diody. Časopis světlo **12**(5), 68–71 (2009). ISSN: 1212-0812
8. Seoul Semiconductor: Specification SSC-Z5 series. http://www.seoulsemicon.com/_upload/Goods_Spec/SZW05A0B.pdf (2012). Accessed 06 Jan 2016

Potential Influence of Light Sources on Human Circadian Rhythms

Karel Sokanský, Richard Baleja, Petr Bos and Tomáš Novák

Abstract The article deals with examining light sources used in households from the point of view of their potential influence on circadian rhythms. Nowadays, there are lots of different types of light sources available on the market with different light colours. An inadequate choice of the colour of the light source can result in worsening visual comfort, which can lead to disturbing circadian rhythms. When choosing light sources, it is necessary to take into account the purpose of the room for which the lights are chosen. Especially, it is a question whether LED sources with similar quantitative and qualitative characteristics have stronger influence on circadian rhythms than classical light sources or not. The reason is primary radiation even of white LED in the blue part of visual spectra. The circadian receptors are the most sensitive in the area.

Keywords Circadian sensitivity · Luminous flux · Correlated colour temperature · LED sources

1 Introduction

This article was inspired by discussions which have been appearing more and more recently. Their main theme is the question whether LED sources used in households have a greater influence on human circadian rhythms in comparison with other light sources such as compact fluorescent lamps, incandescent and halogen light bulbs. In some media, there were even commentaries that LED sources can cause cancer. So we tried to answer the following question. If we take light sources with approximately the same luminous flux and correlated colour temperature, which are used the most often to light households, and if we multiply their radiance spectra by the curve of the spectral sensitivity of human sight controlling the circadian

K. Sokanský · R. Baleja · P. Bos · T. Novák (✉)
Faculty of Electrical Engineering and Computer Science, VŠB - Technical University of Ostrava, Ostrava, Czech Republic
e-mail: tomas.novak1@vsb.cz

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_44

rhythm C, we should get relevant information about which light sources have the greatest and the least influence on circadian rhythms. This comparison was done for four types of light sources used in housing premises. It is an incandescent light bulb, a halogen bulb, an LED filament and a compact fluorescent bulb.

2 Choice of Light Sources

Human circadian rhythms are influenced by the spectral composition of the radiance of a light source, the intensity of radiance and the time during which a person is exposed to this radiance. The spectral composition of radiance is defined by the correlated colour temperature of a light source; this means that by changing the correlated colour temperature, it is possible to influence the human circadian rhythm. For example, light sources with a high correlated colour temperature are not suitable for rooms where people want to relax. Light in shades from white to blue wakes up human body and so it prevents it from a comfortable sleep. From a long-term perspective, these light sources have a bad influence on production of some substances (melatonin—a sleeping hormone), which are necessary for a good sleep and relaxation. Sources with the higher correlated colour temperature are suitable rather for spaces where an absolute concentration is needed. The choice of the colour temperature is connected with the choice of the level of illuminance. This dependence is described in Kruithof curve which is in the Fig. 1 and which describes the dependence between the intensity of illuminance and the colour temperature of the source. The Kruithof curve can be divided into three parts, down from top, there are these areas:

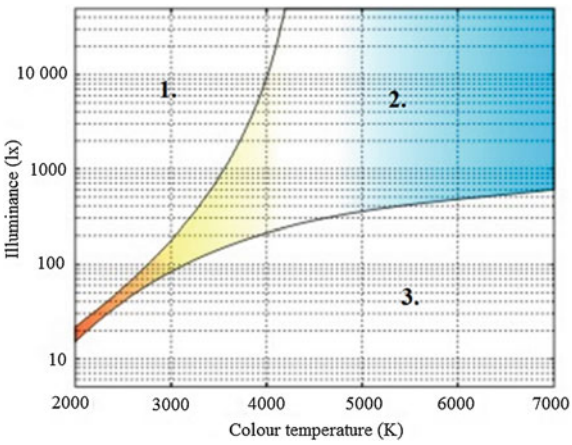


Fig. 1 Kruithof curve—the influence of colour temperature and illuminance on the feeling of visual comfort [3]

Table 1 Light and electrical parameters of light sources

Light source	LED lamp	Compact fluorescent bulb	Halogen bulb	Incandescent bulb
Luminous flux [lm]	420	430	405	420
Correlated colour temperature [K]	2800	2700	2700	2700
Colour rendering index [-]	≥80	≥80	100	100
Power input [W]	4	8	30	40
Power factor [-]	0,54	0,57	0,99	0,99

1. Upper part—The area of the feeling of flare and chromatic distortion (happens very early with low colour temperatures).
2. Middle part—The area of a pleasant feeling from the lighting (visual comfort), it is clear that especially cold light but also day light require high intensity of the light.
3. Bottom part—The area of the feeling of insufficient lighting. Because of reasons mentioned above, sources with low colour temperature are compared in this article first of all [1, 2, 3].

When choosing light sources for comparison, the light sources used in lighting systems in housing premises, where people are the most often, were preferred. They are children's rooms, living rooms and bedrooms. In these rooms, light sources called "warm" are used the most often because consumers prefer visual comfort to visual performance there when using relatively low illuminance. Therefore, light sources with correlated colour temperature from 2700 to 3000 K, colour rendering index superior to 80 and luminous flux around 400 lm were chosen for comparison. Briefly, they are warm sources with a relatively low luminous flux and high colour rendering index. For testing, light sources with following parameters were chosen. (Table 1).

3 Behaviour of Circadian Receptors

Nowadays, the question of the influence of blue light on human organism has been discussed more and more. The organ of sight not only transmits visual sensations but also influences human circadian rhythm (a biological clock) and provides an organism with information when it is a day and when it is a night. However, scientists thought for a long time that both visual sensations and the distinction of light and darkness are provided by the same retina cells (rods and cones). After a series of experiments, it was discovered that circadian rhythms are controlled by other cells—photosensitive ganglion cells. These cells are connected with hypothalamus which controls the biological clock of the organism. The series of

experiments proved that these cells are the most sensitive to blue light whereas the visual system of cones is sensitive mainly to yellow-green light.

New experiments done on blind people proved that even though these people could not see, their circadian rhythm was not influenced at all. The eyes of a patient were exposed to green and blue light of the wavelength 480 nm during the experiment. Whereas the green light did not have any influence on the production of melatonin, the blue light caused approximately 57 % fall of the level of melatonin. This experiment was done by American and British scientists at Harvard Medical School. For illustration, melatonin is a sleeping hormone which influences human day-night cycles. The production of melatonin in the pineal gland is controlled by the biological clock in the brain. The first one to discover this information was the professor Helena Illnerová. Besides that, circadian rhythm influences the cycle of physiologic processes, activity of brain, hormone production, cell regeneration and other physiologic processes [4].

The ganglion cells are the most sensitive to the blue light at the wave lengths from 450 to 482 nm with the maximum sensitivity at the wave length 455 nm. For a visual image of how individual spectra intervene in the spectral sensitivity curve of human sight controlling circadian rhythm, this curve with measured spectra of assessed light sources was drew into a graph, see Fig. 2, with highlighting the maximal sensitivity at the wavelength 455 nm [5, 6].

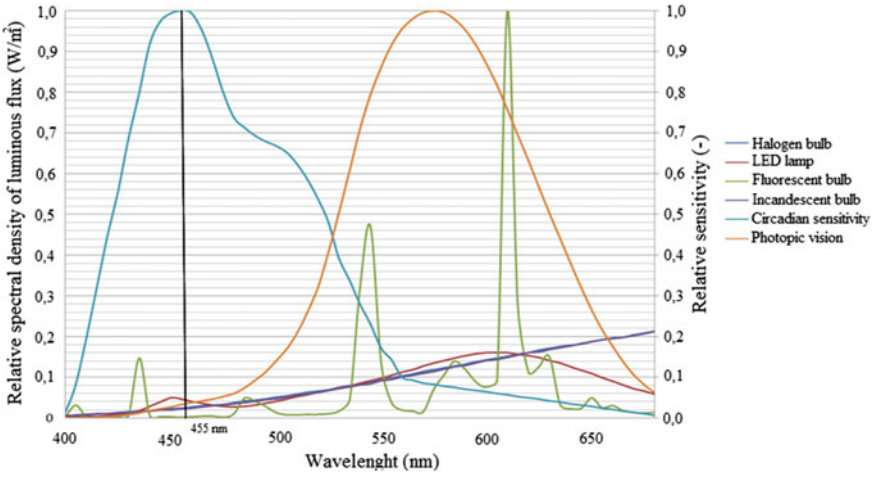


Fig. 2 Spectra of radiation of light sources, spectral sensitivity of human sight controlling circadian rhythm and the *curve* of sensitivity of photopic vision [7]

4 Description of Measurement and Calculation of the Influence of Measured Spectra on Circadian Rhythms

As it has been mentioned, "warm" light sources were chosen for the measurement, with approximately the same luminous flux and the same correlated colour temperature. To compare the influence of individual light sources on circadian sensitivity, it was necessary to measure spectral radiation of light sources so that they were comparable with each other. The spectra of individual sources were measured in an Ulbricht sphere to ensure the same conditions for all light sources. A spectrophotometer JETI Specbos 1211 was used.

With the help of the programme JETI LiVal, the measured spectra were transferred into the table editor Excel. In Excel, the spectra (curves) were depicted into one graph together with the curve of spectral sensitivity of human sight controlling the circadian rhythm C_λ and the curve of spectral sensitivity of photopic vision V_λ, see Fig. 2. These curves are displayed from 400 to 680 nm, which is the range in which circadian receptors are the most sensitive.

For comparison of the influence of individual light sources on the function of circadian receptors, individual spectra of measured light sources were transferred by the curve of spectral sensitivity controlling the circadian rhythm C_λ. By adding these values according to formula (1), we got the final integral transmission of the chosen light source transferred by circadian sensitivity. The same thing according to formula (2) was used to get the final integral transmission transferred by the sensitivity of photopic vision V_λ. Final transfer functions of chosen light sources, which are transferred by circadian receptors, are in the Fig. 3. To describe the influence on circadian rhythms, so-called relative transfer of radiant power transferred by circadian sensitivity compared with radiant power transferred by photopic sensitivity of human eye according to formula (3) was used. This assessment was based on assessing so-called S/P ratio, which is used to assess changes of influences of light sources between scotopic and photopic vision.

$$C = \sum \phi_{(\lambda)} . C_{(\lambda)} \, [W/m^2] \tag{1}$$

C final transfer function of chosen light source between wavelengths 400 and 680 nm transferred by circadian rhythms

$\phi_{(\lambda)}$ relative spectral density of luminous flux of light source at a given wavelength

$C_{(\lambda)}$ relative sensitivity of circadian receptors

$$P = \sum \phi_{(\lambda)} . V_{(\lambda)} [W/m^2] \tag{2}$$

P final transfer function of chosen light source between wavelengths 400 and 680 nm transferred by sensitivity of photopic vision

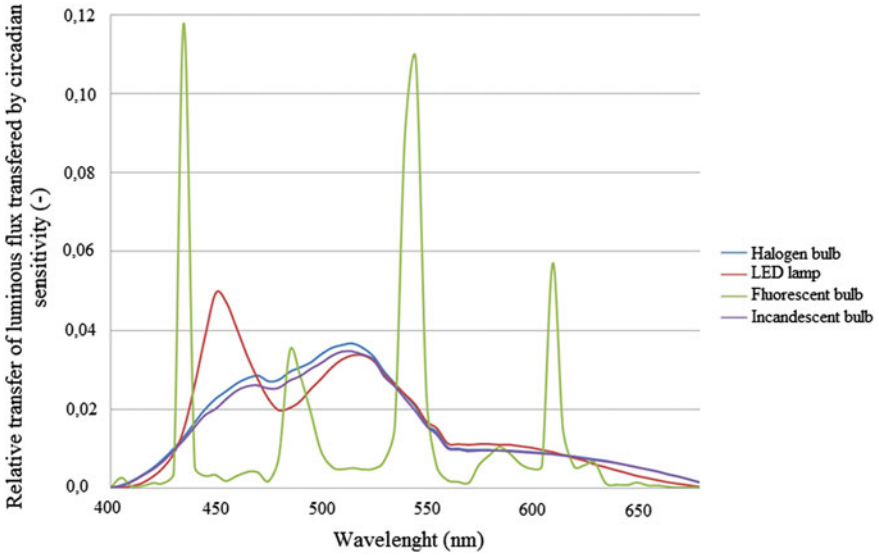


Fig. 3 Final spectral transfer of chosen light source transferred by circadian sensitivity

$\phi_{(\lambda)}$ relative spectral density of luminous flux of light source at a given wavelength

$V_{(\lambda)}$ relative sensitivity of circadian receptors

$$\frac{C}{P}ratio = \frac{C}{P} \cdot 100\,[\%] \tag{3}$$

$\frac{C}{P}$ratio Relative transfer of luminous flux transferred by circadian sensitivity to luminous flux transferred by photopic sensitivity of human eye

The Table 2 shows counted values of the final relative transfer of luminous flux transferred by circadian sensitivity to luminous flux transferred by photopic

Table 2 Final relative transfer of luminous flux transferred by circadian sensitivity to luminous flux transferred by photopic sensitivity of human eye

Light source	$\frac{C}{P}$ratio (%)
LED lamp	35,11
Halogen bulb	33,66
Incandescent bulb	32,67
Compact fluorescent bulb	24,67

Table 3 Relative spectral density of luminous flux transferred at the wave length 455 nm

Light source	Relative spectral density of luminous flux transferred at the wavelength 455 nm [W/m^2]
LED lamp	0,047
Halogen bulb	0,025
Incandescent bulb	0,023
Compact fluorescent bulb	0,002

sensitivity of human eye. Final transfer values differ only slightly although it is necessary to draw attention to the fact that if sources with different correlated colour temperature were used, final values would differ more.

As it was mentioned, ganglion cells are the most sensitive to the wavelength 455 nm. Because of that, spectral density of luminous flux at this wavelength was also compared for each light source. The values are in the Table 3.

From the point of view of evaluating the influence of the contents of blue light at the wavelength 455 nm, where the effect of light transfer on circadian trains is the highest, it is possible to state that LED diodes have the highest effect, see Table 3. Measured and counted results imply that this effect is for example in comparison with incandescent light bulbs approximately two times higher.

5 Conclusion

On the basis of above mentioned comparison, we came to the conclusion that even though the radiance spectra of light sources used in housing premises are very different, we can expect that their influence on circadian rhythms does not vary too much. The peaks in areas of maximal sensitivity of circadian sensors produced by LED lamps and compact fluorescent lamps are not so energetically important to have a significant influence on exciting circadian receptors, compared to classical warm sources. Our pilot measurement implies that an LED lamp has the greatest influence on circadian rhythms. However, it is not distinctively higher that with light bulbs. The least influence has from this point of view compact fluorescent lamp, see Table 2 [6].

Nevertheless, it is necessary to stress that it was a comparison of sources with approximately the same (low) colour temperature. In case there were light sources with higher correlated colour temperature (LED or compact fluorescent lamps), it is possible to expect that their transfer level of relative spectral luminous flux in the area of sensitivity of circadian sensors will be much higher. Further measurements will be done with light sources used in housing premises and having a higher colour temperature.

Acknowledgment This article was prepared with the support of the project "Intelligent lighting control systems" SP2016/151, by institution of VSB-TU Ostrava.

References

1. Gall, D.: Die Messung circadiener Strahlungsgrössen, Ilmenau
2. Diviš, D., Sokanský, K.: Nové poznatky v oblasti vidění. In: 10th International Scientific Conference on Electric Power Engineering, Kouty nad Desnou, pp. 34–36. (2009)
3. Davis, R.G., Ginthner, D.N.: Correlated color temperature, illuminance level, and the kruithof curce. J. Illum. Eng. Soc. **19**(1), pp. 27–38 (1990)
4. van Bommel, W.: Incandescent replacement lamps and health. Light Eng. **19**(1), 8–14 (2011)
5. Novák, T., Socha, B., Carbol, Z., Sokanský, K.: Luminance evaluation of LED indoor luminaires for workspaces lighting. In: 13th International Scientific Conference Electric Power Engineering, Brno (2012)
6. Brüning, A., Hölker, F., Franke, S., Kleiner, W., Kloas, W.: Impact of different colours of artificial light at night on melatonin rhytm and gene expression of gonadotropins in European perch. Sci. Environ. **543**, 214–222 (2016)
7. Brainard, G., et al.: Action Spectrum for Melatonin Regulat in Humans. J. Neurosci. **21**, 6405–6412 (2001)

Optimization of Antenna System for MIMO Technology

Lukáš Wežranowski, Lubomír Ivánek, Zdeněk Urban and Yahia Zakaria

Abstract This paper deals with the optimization of antenna system for MIMO technology, assembled of two rectangular microstrip patch antennas. The rectangular patch is by far the most widely used configuration. First, there is described the initial shape of the antenna, further optimization of the rectangular patch and finally is described optimization of the two antennas of this type. We are interested radiation pattern and gain of the antenna system and the leakage of signals from one antenna to another antenna. Optimization was performed using the software COMSOL Multiphysics. Its Optimization module has implemented the following optimization methods: Bound Optimization (BOBYQA), Nelder-Mead, Coordinate search, and Monte Carlo.

Keywords Optimization module · COMSOL multiphysics · Antenna system · MIMO technology patch antenna

1 Introduction

This article describes an antenna that is designed to MIMO technology. State of the art of antennas for MIMO technology is documented in the literature [1–3]. **MIMO** is an acronym that stands for **M**ultiple **I**nput **M**ultiple **O**utput. It is an antenna

L. Wežranowski · L. Ivánek · Z. Urban · Y. Zakaria (✉)
Faculty of Electrical Engineering, and Computer Science, Department of Electrotechnics, VSB–Technical University of Ostrava, 17.listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic
e-mail: yahia.salem.st@vsb.cz

L. Wežranowski
e-mail: lukas.wezranovski@vsb.cz

L. Ivánek
e-mail: lubomir.ivanek@vsb.cz

Z. Urban
e-mail: zdenek.urban@vsb.cz

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_45

technology that is used both in transmission and receiver equipment for wireless radio communication. There can be various MIMO configurations. For example, a 2 × 2 MIMO configuration is 2 antennas to transmit signals (from base station) and 2 antennas to receive signals (mobile terminal). It is basically an antenna system consisting of two patch antennas. We were interested chiefly shape optimization of antenna arrangement that allows the least leakage signal from one antenna to another antenna. Calculation and optimization of antenna were performed using COMSOL Multiphysics. Module optimization of this software includes two different optimization techniques Derivative-Free and Gradient-Based optimization. Four droughts methods are available in the Optimization Module: The Bound by Quadratic Approximation Optimization (BOBYQA) [4], Nelder-Mead [5], coordinate search [6], and Monte Carlo methods [7].

The proposed shape of the patch antennas shown in Fig. 1. Electromagnetic field distribution in the antenna shown in Fig. 2. 3D radiation pattern in Fig. 3.

2 Optimization of Antennas

2.1 Optimization of Antenna Rectangular Patch

Optimization of the dimensions of the patch antenna is performed as follows. Dimension of power port, according to parameter S11, adjusts for ensure the best possible transfer wave from the cable to the antenna. Optimization is performed with respect to the complexity on the computing power for a first frequency 5, 6 GHz—identified as f_x. After a basic overview a calculation is made in an optimization model for the given frequency range f_min to f_max.

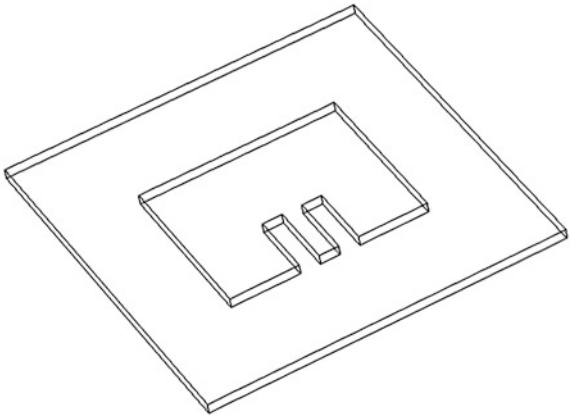


Fig. 1 Shape of the patch antennas

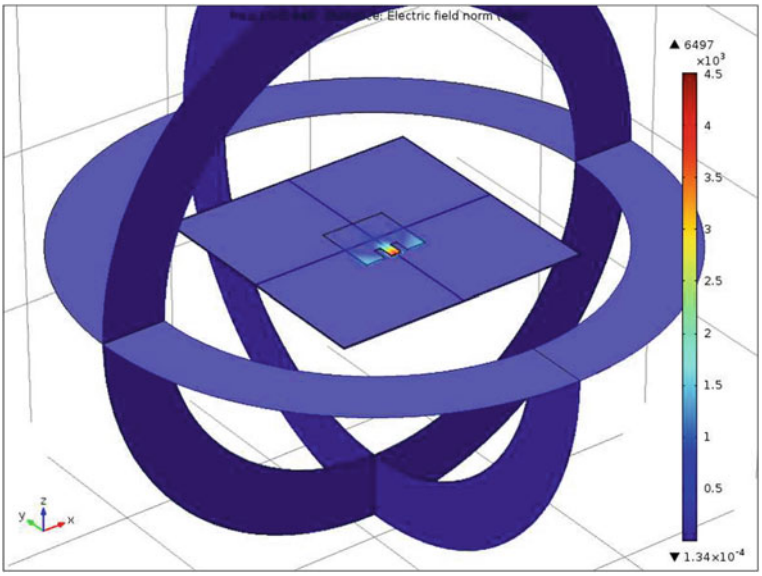


Fig. 2 Electromagnetic field distribution

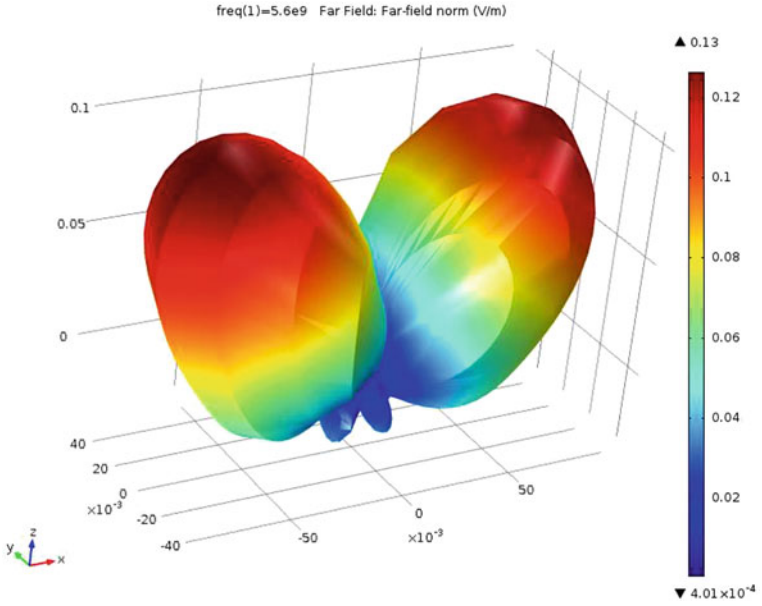


Fig. 3 3D radiation pattern

Optimization of input power feed

$$f = abs(comp1.emw.S11dB)$$

Setting up of the program allows us to seek maximum or minimum of objective function, so we use the fact that best result of optimization provides the absolute value of the S11 for one frequency f_x.

Optimizing by eFar

By eFar optimization—optimizing the size of the patch—W_patch, L_patch a L_stub

$$f = comp1.intop1(comp1.emw.gaindBEfar)$$

Type a search result—maximum. Searched solutions—the maximum of the resulting objects. The method used—Nelder-Mead limited to a maximum value of 50 for calculations. Patch antenna after completing optimizing by eFar is showed in Fig. 4. Further results of optimization by eFar are seen in Fig. 5.

Optimizing by S11

By S11 optimization—optimizing the size of the patch—W_patch, L_patch, W_line a L_stub.

$$f = abs(comp1.emw.S11dB)$$

Type a search result—maximum. Searched solutions—the maximum of the resulting objects. The method used—Monte Carlo with the maximum limited to 500 random calculations. Patch antenna after completing optimizing by S11 is shown in Fig. 6, (Figs. 7, 8).

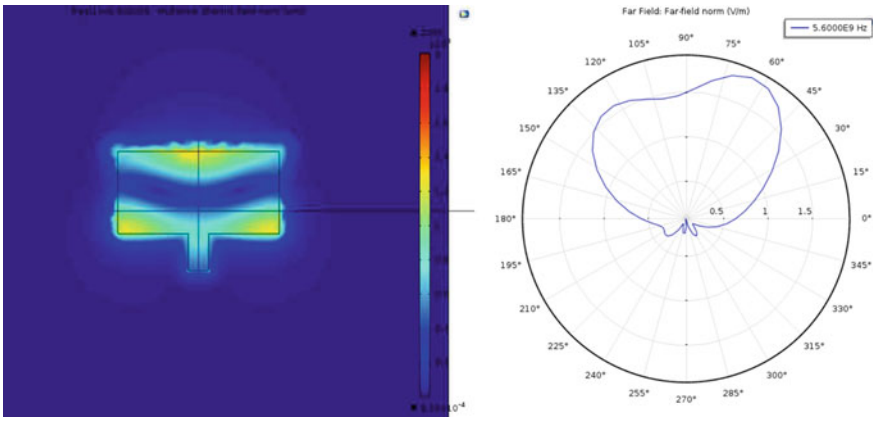


Fig. 4 Shape and radiation pattern after eFar optimization

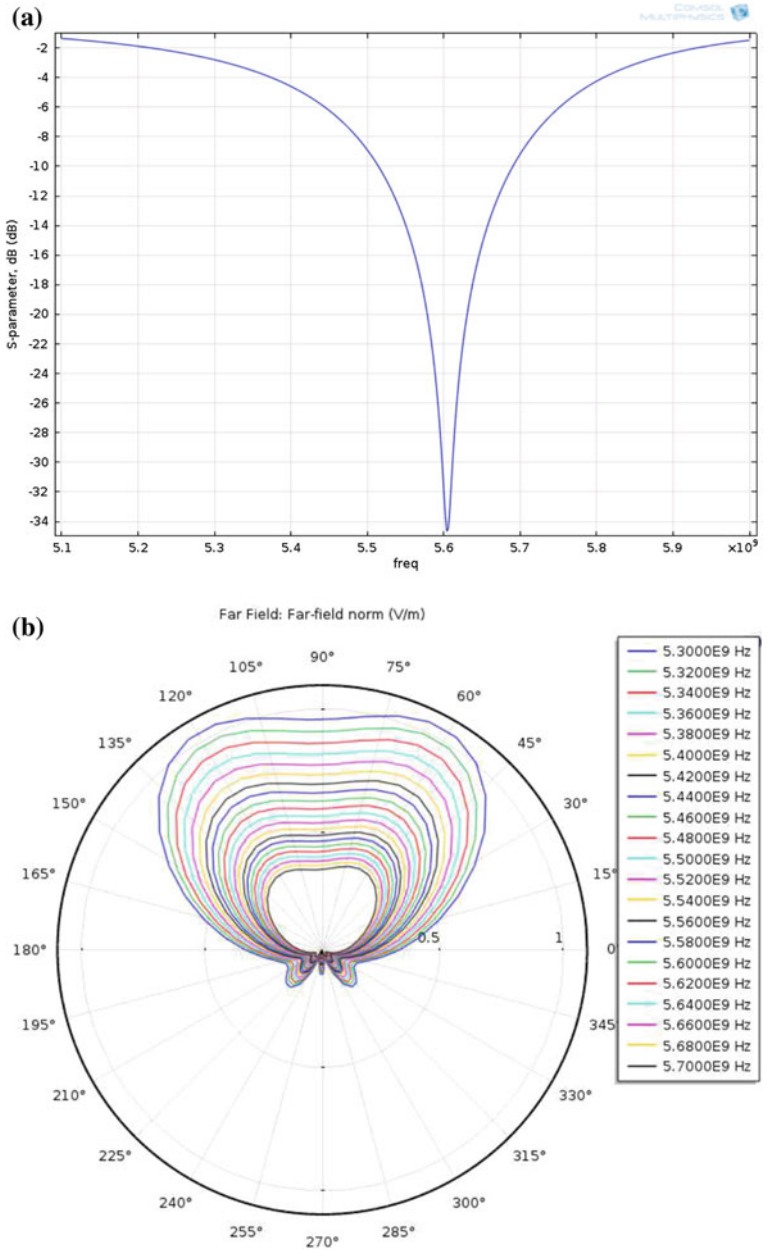


Fig. 5 **a** Further results of optimization by eFar. **b** Further results of optimization by eFar

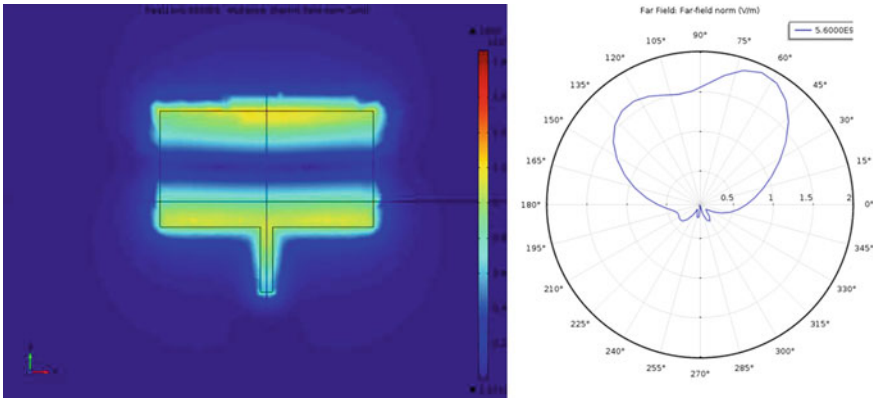


Fig. 6 Patch antenna after completing optimizing by S11

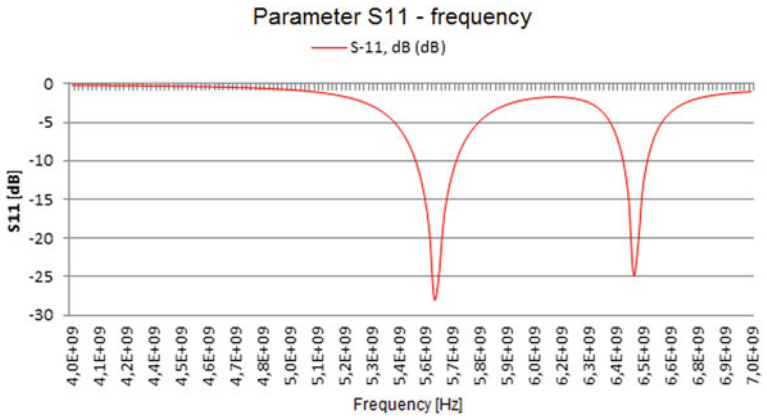


Fig. 7 Dependence of S11 on frequency

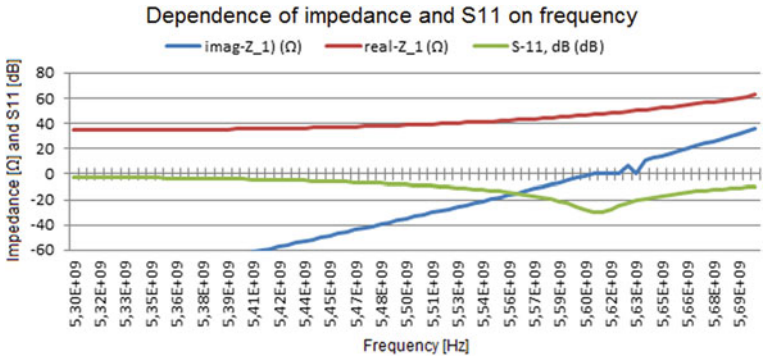


Fig. 8 Dependence of impedance and S11 on frequency

2.2 *Optimization of Antenna Arrays with Two Rectangular Patch*

Adjustment of dimensional substrate

In the Global definition section, Parameters—at first we are resize the width of the substrate at 120 mm in parameter table. Shift of the first patch antenna—we need to reach location, in other words, shift of the patch antenna symmetrically to the x-axis.

In Geometry—Patch we can find the section position and there we move patch antenna in the x-direction of minus L/2, it means we moved patch antenna about half of the distance L to the left. Click on the build selected for a redraw.

In Geometry—stub we can find the section position and the x-axis to specify the displacement as above. New parameter of shift will be this: −L/2 + W_stub/2 + W_line/2 and click to build all objects for a redraw.

The next patch antenna we draw as a copy of the original antenna. In the Geometry menu, click on Transforms and choose Copy then we select dif 1 to the section Input objects. Then at the Displacement and in the X-axis settings, we can place the element at a distance L and click on Build All button for redrawing.

The illustration below shows the basic structure for the distribution of elements in the configuration of 2 × 2. At first we place the components as close to each other as possible. It is also possible mutual overlapping elements using a multilayer board, but we greatly increases the complexity and cost of manufacturing, because of that it will remain in the configuration on the two-layer board (Fig. 9).

Then, we draw the decomposition of network for calculating and check if in MESH setting is highlighted edges with finer structure around power lines. Softening of computer networks around the power line will allow us to refine the calculation of the S11 parameter.

Setting the measuring probe

Setting the measuring probe—the aim of this calculation is to determine how much energy is emitted with non-powered patch antenna and then to optimize the

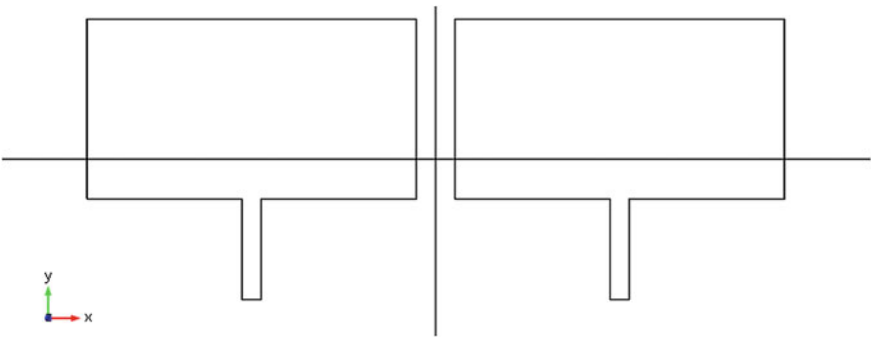


Fig. 9 Arrangement of elements combinations

distances between the antennas. First we will calculate the required variables on the power port of patch antenna 2, on which is placed the probe for the calculation.

In Definition section, select Probes and select Global Variable Probe.

Under Expression section in the upper right corner, select Model—Components—Electromagnetic Waves, Frequency Domain—Energy and Power—emw.intWe—Total electric energy and confirm. In section "table and plot units", select the unit μeV. Program offers the basic unit as Joule, but due to the expected size of the values, the observed number would be very small. Because of that we choose electron volt.

Checking before calculating

In the electromagnetic Waves section we find Perfect electric conductor 2 and check, whether we have selected all the copper surfaces, including newly added. At the selection must be 15,20,21,50 and 51. For Lumped Port 1 in the section Selection we retain only port 26 as a power source. For this simulation we erase port 56—this port is not powered.

Own calculation we perform for a given range of frequencies by adjusting in the Study section, Step 1: Frequency Domain. Optimization of the second patch antenna distance—mutual positions. Objective of this calculation is to determine how the position of patch antennas is affected to each other. This may happen in two ways—by optimizing the S11 parameters, power ports, or by the values obtained using probes placed on the patch antenna 2 or according to the total amount of radiated energy.

At first we dealt with how the relative position is affects. We used the same optimization function:

$$f = \text{abs}(\text{comp1.emw.S11dB})$$

3 Conclusions

The results of the data are in the table in Annex No. 5. Optimize of the parameter L was performed by Monte Carlo method, which was limited to 500 calculations and for distance L = 25 mm to 60 mm (Fig. 10).

In the figures above is shown the distribution of the field on two patch antennas, when is soldered the left patch antenna and also show directional characteristics of the antenna, which is evident from affecting to the radiation pattern of the second unpowered patch antenna. Table 1 shows the values for optimizing the relative position of the patch antennas 2 × 2 with a focus to achieving the best possible parameter values of S11.

In the graph is like an optimization function used absolute value of the S11 parameter, so the best value is the highest positive value (Figs. 11, 12).

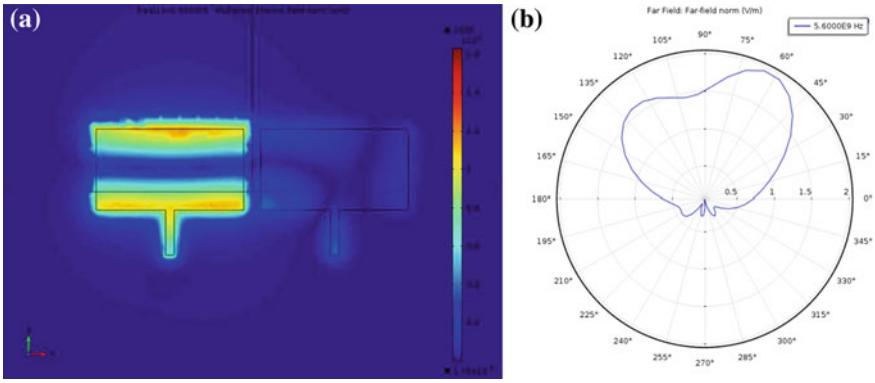


Fig. 10 **a** Display of layout field, **b** Directional characteristics 2 × 2

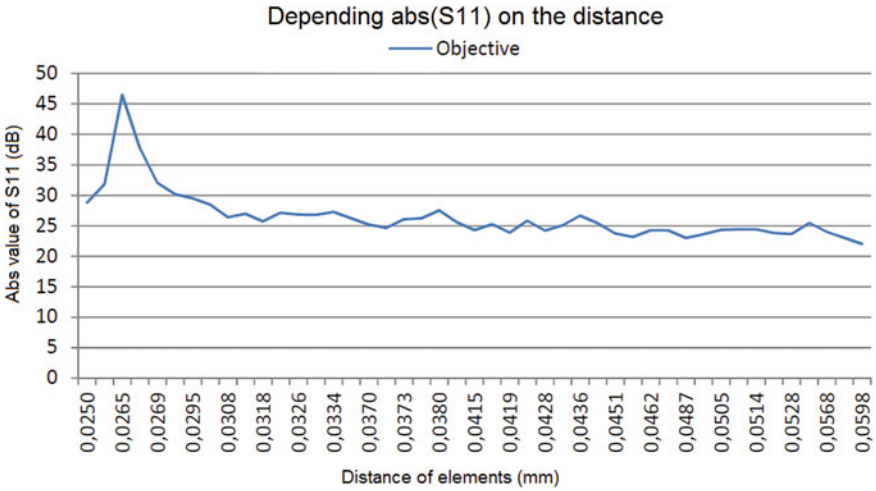


Fig. 11 Influence of location the second patch antenna to value of optimization function of the first patch antenna

Table 1 The best achieved adaptation with the second patch antenna

freq (Hz)	LP impedance (Ω)	S-11, dB (dB)
5.6000E9	50.157–0.44870i	−46.472

Optimizing on the best parameters of overall radiated energy

For the optimization it is useful to examine the total amount of radiated energy from antenna subsystem and relate it to the distance between patch antennas. We selected the solution of the calculation according of the total amount—"Total electric energy". In the menu Optimization, we set the variable optimization according to the above descriptions.

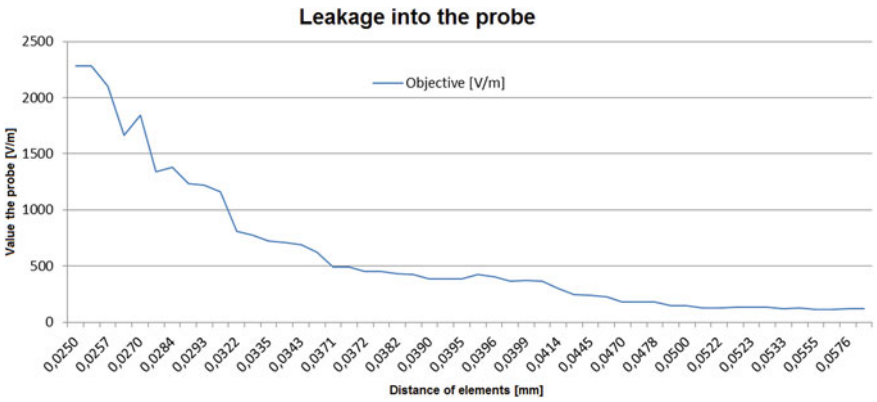


Fig. 12 Graph of leakage dependence into the second patch antenna and distance elements

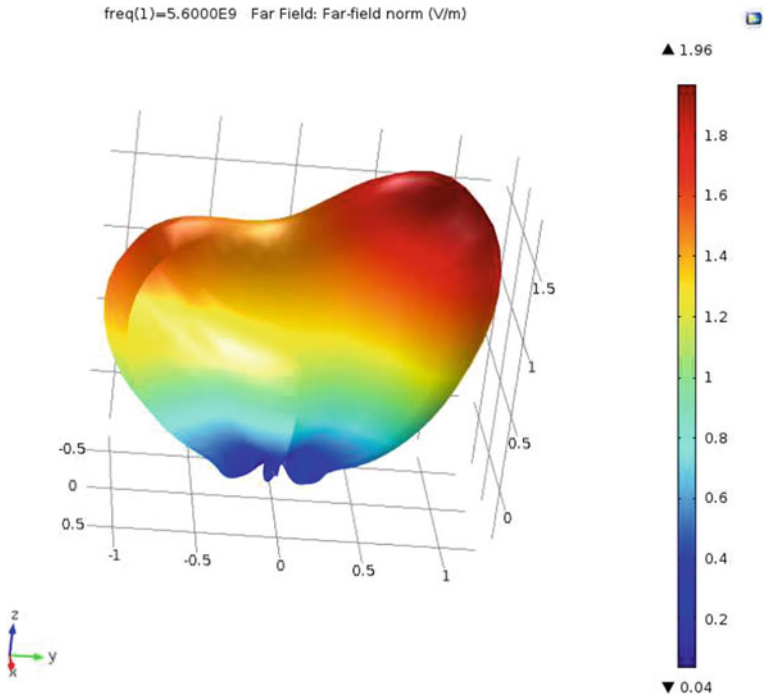


Fig. 13 3D simulation of radiation diagram with the one powered patch antenna and second unpowered patch antenna

Figure 13 shows a 3D view on the antenna radiation patterns. On the right side is the obvious influence radiating characteristics for secondary patch antenna.

Acknowledgments This work was supported by the project SP2016/143 "Research of antenna systems; effectiveness and diagnostics of electric drives with harmonic power; reliability of the supply of electric traction; issue data anomalies." The authors would like to express their thanks to the members of Department of Electrical Engineering at VSB—Technical University of Ostrava for their valuable instructions and support.

References

1. Yang, B., Adams, J.J.: Modal Q as a bounding metric for MIMO antenna optimization. In: Annual Review of Progress in Applied Computational Electromagnetics, May 2015. [7109724] Applied Computational Electromagnetics Society (ACES) (2015)
2. Du, J., Li, Y.G.: Optimization of antenna configuration for MIMO systems. IEEE Trans. Commun. **53**(9), Sept 2005
3. Berenguer, I., Wang, X., Krishnamurthy, V.: Adaptive MIMO antenna selection via discrete stochastic optimization. IEEE Trans. Sig. Process. **53**(11), Nov 2005
4. Powell, M.J.: The Bobyqa Algorithm For Bound Constrained Optimization Without Derivat Ves. Cambridge NA Report NA2009/06. University of Cambridge, Cambridge (2009)
5. Press W., Teukolsky, S., Vetterling, W., Flannery, B.: Numeri-cal Recipes: the Art of Scientific Computing, 3rd edn. Cambridge University Press (2007)
6. Frandi, E., Papini, A.: Coordinate search algorithms in multilevel optimization. In: Universit`a degli Studi dell'Insubria, Dipartimento di Scienza e Alta Tecnologia & di Energetica Sergio Stecco, Published online: 15 Nov 2013, Optimization Methods and Software, vol. 29, Issue 5 (2014)
7. Amar, J.G: The monte carlo method in science and engineering. In: University of Toledo. Published by the IEEE Computer Society, vol. 8, Mar/Apr 2006

Automation and Control of Energetic Systems Using Cogeneration Unit in Industry

Michal Špaček and Zdeněk Hradílek

Abstract Recently, we are trying to deal with energy as efficiently as possible. When it comes with nature-friendly technologies and energy saving. One of the possible solutions is to replace the cogeneration unit instead of gas boilers. The gas is burned in a combustion engine which is connected the shaft and an electrical generator that produces electricity. Heat energy is obtained by engine cooling and flue gases. Another possible application of cogeneration unit is overlaying of consumption peaks in power supply network when we are trying to reduce them.

Keywords Cogeneration unit · Electric output · Heating power · Accumulation tank · Energy balance

1 Introduction

One of these energy savings is applied in engineering company in the Czech Republic. Gas boiler and gas burners were replaced by cogeneration unit with storage tank. Thermal and electrical energy are produced by using cogeneration units burning natural gas. To cogeneration unit is connected a pipe for exhausting hot water into the technology and accumulation tank (AKU). Part of the heat (1.3) is thus supplied directly into the technology, and part of the heat (2/3) is supplied into the accumulation tank. The diagram of cogeneration unit technology and accumulation tank is displayed in Fig. 1. The technology uses produced heat is shown in Fig. 2, which shows heat exchanger and heated tanks with chemicals.

Warm water of the main heating circuit of cogeneration unit operates with a thermal gradient of 85/65 °C. For measuring of thermal energy is mounted col-

M. Špaček · Z. Hradílek (✉)
17. listopadu 15, Poruba, Ostrava, Czech Republic
e-mail: zdenek.hradilek@vsb.cz

M. Špaček
e-mail: michal.spacek.st2@vsb.cz

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_46

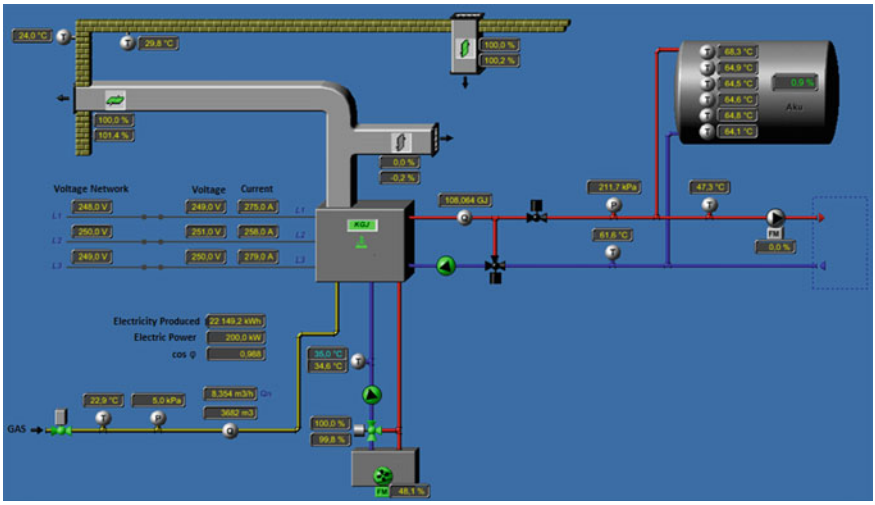


Fig. 1 Visualization of technology of cogeneration unit with accumulation tank and other auxiliary devices

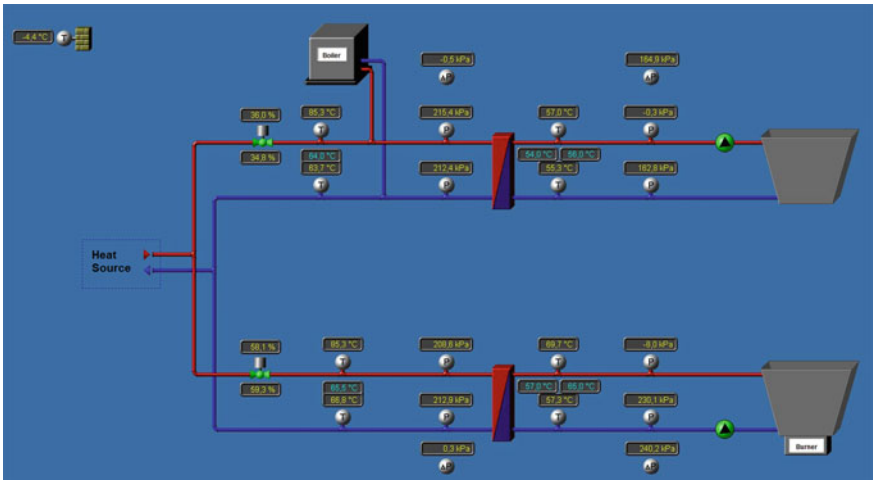


Fig. 2 Visualization of adjacent technologies using thermal energy

orimeter. For accumulation of thermal energy was used accumulation tank with a capacity of 60 m^3. This accumulation tank is thermally insulated.

To achieve greater engine power of cogeneration unit is installed a turbo compressor. Compressed natural gas is cooled using by closed warm water circuit with dry cooler. In the cooling circuit is necessary to maintain the temperature gradient of 38/35 °C.

While designing has been calculated that in the company will be consumed all electrical power. And there will be no supply into the network. For last six months it has started to changeable electricity consumption in the company. Overflows of electricity occurred from cogeneration unit into the network.

The company has requested the reserved power of 0 kW. Against of these overflows company wants to realize the regulation of cogeneration unit to prevent electricity overflows.

2 Technology Cogeneration Unit

2.1 Cogeneration Unit

There are in the company install cogeneration unit with electric power 199 kWe and thermal power 263 kW. Basic data of cogeneration unit are in Table 1.

Important parameter of cogeneration unit is energy balance, constituting a graphical representation of the energy flow. Energy balance shows the conversion of primary energy (natural gas, 100 %) into an electrical and thermal usable energy. Losses which occurs based on energy converting are also shown. Usable electrical energy rise by combustion in a combustion gas Otto engine and its rotary motion is converted in a synchronous generator to a current. Thermal energy rise also by combustion process in a gas Otto engine. Gas Otto engine is operated as a combustion engine with turbo blower and two-stage cooling of the mixture with the air ratio lambda 1.6. It is divided into heat in flue gases, collecting piping, engine block and engine lubricating oil and serves to heating e.g. heating water. The complete efficiency level of cogeneration unit module is sum result of the electric and thermally usable energy, and it is shown in Fig. 3 and Table 1 [1].

The module of block heat power station is unit ready for connection to a source of gas, electric network and heating circuit. It includes an air-cooled synchronous generator to produce three phase power with a voltage of 400 V, 50 Hz. The heating circuit is operated with temperature gradient 85/65 °C. Cogeneration unit module may be operated electrically or thermally, in the operating range of 50 −100 % of rated electric power (corresponding to 60−100 % of the thermal output). Installed cogeneration unit in the company is shown in Fig. 4.

Table 1 Parameters of cogeneration unit [2]

Parameters	Max. electric performance	199 kWe
	Max. heat performance	263 kW
	Engine performance	210 kW
	Fuel consumption	53 Nm^3/h
	Min. total efficiency	89.60 %
	Temperature gradient of heating circuit	65/85 °C
	Max. heat performance	10/50 °C

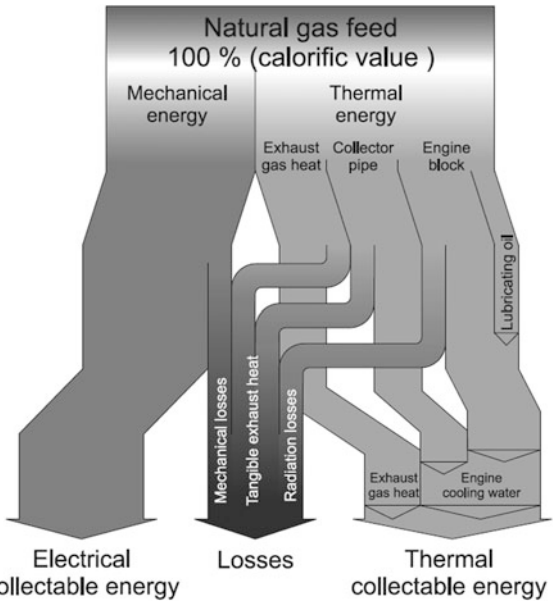


Fig. 3 Energy balance of module cogeneration unit

Fig. 4 Cogeneration unit 199 kWe/263 kW

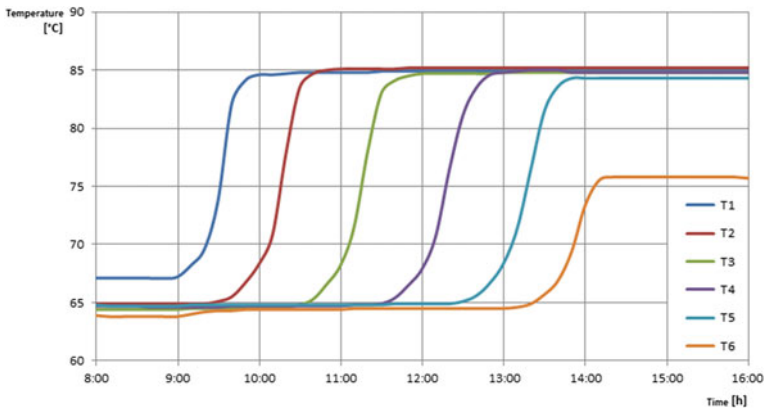


Fig. 5 Charding of storage tank

2.2 Storage Tank

Thermal energy is accumulated in a tank of circular cross section 3.40 m and a total length of 60 m^3, with the volume of 7.812 m. The casing of the tank is made from steel with a 100 mm thermal insulation thickness. The tank is in upper and lower part equipped with inputs and outputs for water. The tank is also equipped with six temperature sensors. A temperature gradient of water is 85/65 °C. Inside of the tank are input and output troughs, which ensure uniform inflow of water into the tank to avoid cooking and subsequent mixing of water in the storage tank.

Temperatures are measured with the help of PLC control, which is recorded at 10 min intervals into the PC memory.

Charging process of the storage tank is on the Fig. 5. Temperature sensors with which were recorded waveforms are spread over the entire height of the storage tank. Temperature sensor labeled T1 is located most highly, but temperature sensor T6 is located the lowermost. Temperature sensors T2, T3, T4 and T5 are located downwardly between the sensors T1 up to T6. From Fig. 5 you can see that at 9:00 am cogeneration unit has been activated when the temperature began to rise at the highest point of the storage tank. Depending on the time began to increase temperatures of the individual sensors towards the bottom of the tank. Lowest temperature sensor T6 was recharged only at 76 °C, to avoid overheating of cogeneration unit.

3 Connection of Cogeneration Unit

3.1 Connection of Cogeneration Unit into the Network

In Fig. 6 is shown connection diagram of cogeneration unit into the electrical. The company is supplied from a distribution network 22 kV, via two transformers T1

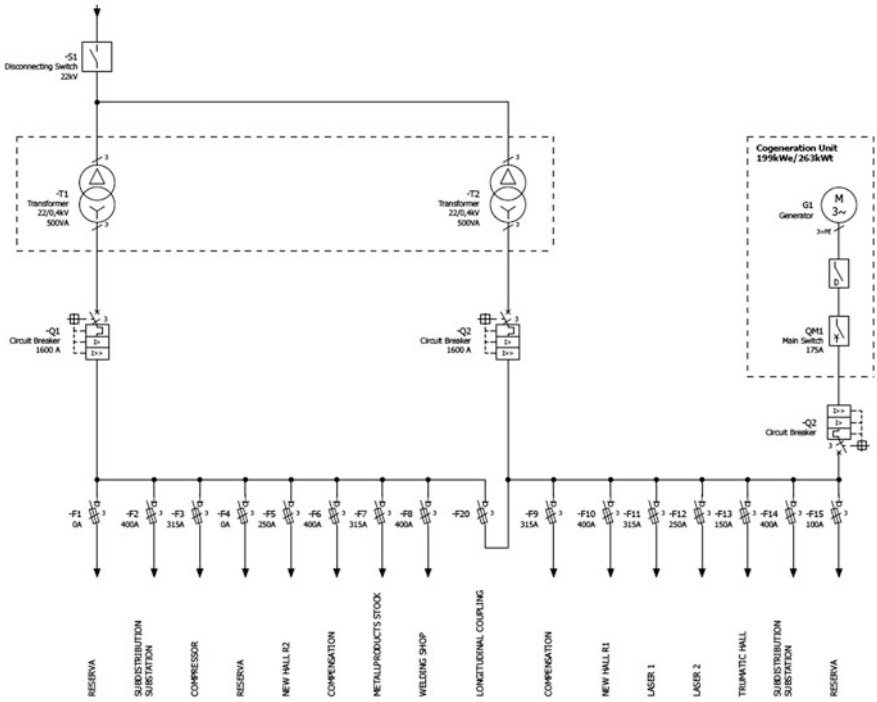


Fig. 6 Mono-polar wiring diagram of cogeneration unit into the network

and T2. Each of them has the performance of 500 kVA. These transformers can operate in parallel or separately. Currently they are working separately, and for their parallel collaboration is used longitudinal fuse switch F20. Transformers operate separately, mainly because of the lasers, which have a very variable consumption from the network, which causes fluctuations in line voltage. For this reason, the company has two types of network, "dirty" and "clean". Clean network is supplied from the transformer T1 on which is connected the lighting, offices and other appliances dependent on the quality of the network. To dirty network (transformer T2) are connected to lasers, welding centers and the cogeneration unit.

3.2 Control of Electric Power of Cogeneration Unit

Regulation methods of the cogeneration unit may be large amounts. One of them is continuous regulation that is used most often. Among other control methods belongs jumping method, in which we perfume regulation in individual steps (for example, 50, 60, 70, 80, 90 and 100 % of rated power). For cogeneration units is necessary to keep power control between 50 and 100 % of rated power. Below

50 % of rated power unit has very poor efficiency. The best method of unit operation is at a maximum (100 %) power, in which has the largest ratio produced electricity to heat energy. When designing units from the viewpoint of regulating electric power it is better to propose more units, which would be triggered sequentially according to the desired performance and would be regulated by the rated power. In our case, the unit will be operated at nominal power.

Because unit has an output of 199 kW it falls in the Czech Republic into the category of plants with an output of over 100 kW connected to distribution network. In case of danger and reliable operation of the electricity system is necessary for supervisory control to temporarily restrict or shut down active power supply from the manufactory of electricity [2].

The source must be able to adequately (quickly and accurately) respond to commands from the control room PDS to reduce active power gradually in the mode 0, 50, 75 and 100 % of the installed capacity.

At source is applied smoothly (not stepped) voltage regulation and reactive power (U/Q) according to the dispatcher or by system of automatic control.

To the cogeneration unit is also connected control unit, which provides to control center, operator DS transmission of measuring and signaling about cogeneration unit.

In the Fig. 7, dated 1.7.2015 is recorded cogeneration unit regulation using by RTU control of switchboard emitting signals from the control room. At 10:43 cogeneration unit was turned on when in time 10:48 to nominal value. This cogeneration unit worked in nominal mode until 11:30, when the command came from the dispatcher to reduce the performance of the unit to 75 % of rated power. After two minutes, the unit ran in the 150 kW and in this performance was operated until 12:10, when came the command from the control room to reduce power to

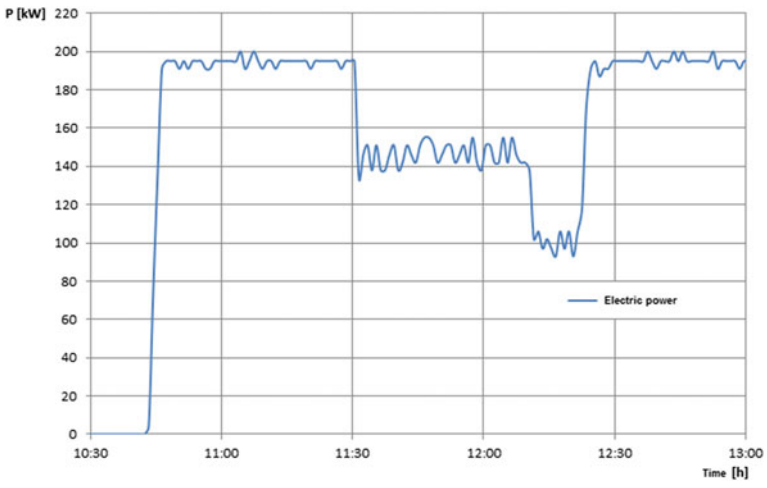


Fig. 7 Performance regulation of KGJ using by RTU switchboard

50 % of rated power (100 kW). Unit ran into this performance for one minute and was operated until 12:23, when came the command for start-up the unit at 100 % of rated power. This performance was achieved in a time of two minutes.

At cogeneration unit run at 75 and 50 % of nominal performance may be observed a significant power changes. These are caused by a short operation of cogeneration unit, when its performance is stabilized mostly after two hours of operation.

This cogeneration unit is controlled by a heat regime. When is operated only at 100 % of rated power. This type of unit can be operated from 50 to 100 % of the rated electric power. Regulation of electric performance will be performed following way. In the first phase on supply will be mounted current measuring transformer about parameters of 800/5 A, accuracy class 0.5. This output current will be connected to the transmitter simultaneously with the potentials from individual phases. Converter will be convert the calculated electric performance to 4-20 mA. When 4 mA will be correspond to −200 kW and 20 mA−525 kW. The output current loop will be connected to the control machine, which will give impetus of operation of cogeneration unit. A simplified diagram is shown in Fig. 8.

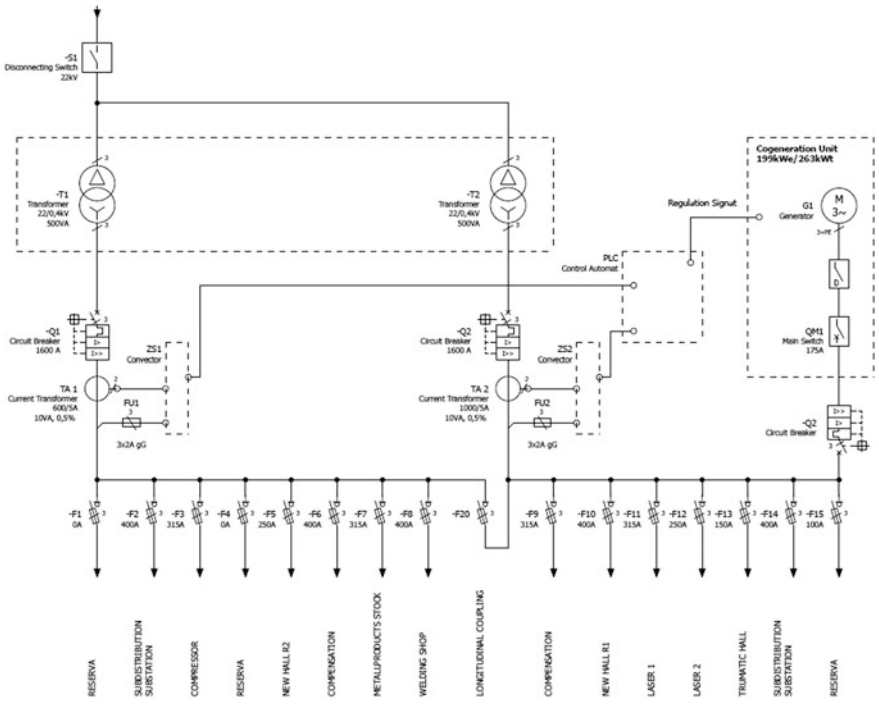


Fig. 8 Simplified wiring diagram of regulation

Further, to the control system will be connected individual information about the operation of cogeneration unit. Among which will include: a power failure, gas leaks, basic failure of cogeneration units. This information will be sent via GSM communication in SMS form to preprogrammed numbers [3, 4].

4 Conclusions

This article aims at verifying regulation of cogeneration unit not only from a theoretical point of view, but also from the practical. At the current regulating of heat and electrical performance by one source always arise problems. The biggest problem arises if we have total discharged an accumulation tank and we can't run the cogeneration unit because of the overflow of electrical power into the network. This problem is solved by closing the old (original) heating equipment (gas boiler, gas burners). For this reason there was installed accumulation tank which regulates heat consumption. One of the important factors is devise an algorithm that will turn on and then turn off running of the unit. Another factor is also the operation time at which is not recommended turn off the unit under running time of two hours. The aim of resolution is also the option of analysis of regulation regarding to variable loads in the industrial network. It is necessary to devise an algorithm to regulate the running of cogeneration source.

References

1. Kirtley, J.L.: Electric Power Principles. Massachusetts Institute of Technology, 404 pp. (2010). ISBN: 978-0-470-68636-2
2. Dincer, I., Rosen, M.: Thermal Energy Storage, 620 pp. (2010). ISBN: 978-0-470-74706-3
3. Momoh, J., Mili, L.: Operation and Electric Energy Processing Systems, 200 pp. (2010). ISBN: 978-0-470-47209-5
4. Špaček, M., Hradílek, Z., Moldřík, P.: Cooperation between photovoltaic and biogas power plants within 22 kV network. In: Electric Power Engineering. Czech Republic (2015)

Reliability Database of Industrial Local Distribution System

Jiri Drholec and Radomir Gono

Abstract Calculatin of the supply reliability is impossible without trustworthy input parameters. It is possible to find reliability indices of electricity transmission or distribution system but there is hardly any of local distribution system at heavy industry. The paper shows the way how to get reliable results—collection and processing of information and documentation from the dispatcher logs on the operation of the local distribution system within the premises of an industrial plant. The first part focuses on development of equipment passportization and processing of documentation for system operation. Additionally, the work deals with calculation of reliability indices of overhead lines and components. The final part brings the comparison of reliability indices of the input overhead lines of the industrial local distribution system with overhead lines of the distribution system of the Czech Republic.

Keywords Reliability · Local distribution system · Industrial plant · Control centre · Distribution plant · Failure rate · Mean time to repair · Faultless operation probability

1 Introduction

This paper deals with collection and processing of data from control centre logs in the company concerned with industrial power engineering [1], which is a complex and extensive enterprise focused specially on the needs of metallurgical and steel

J. Drholec · R. Gono (✉)
Department of Electrical Power Engineering, VSB - Technical University of Ostrava, Ostrava, Czech Republic
e-mail: radomir.gono@vsb.cz

J. Drholec
e-mail: jiri.drholec.st@vsb.cz

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_47

industry. Its scope of activities includes provision of services within the fields of water economy, combined heat and power production, gas engineering, power engineering and production of technical gases. This paper further deals with operation of the local distribution system (LDS) situated within the premises of the company focused on production and processing of pig iron as well as the metallurgical and machine engineering production [2]. The final part lays emphasis on calculation of reliability indicators and their comparison against the reliability indicators in the superior distribution system.

2 Collection of Data on Outages and Failures of Power Supply

The most data on reliability of power supply within LDS can be obtained by monitoring and evaluation of its operation [3]. Efficient records of downtime, outages and failures with subsequent analyses of their consequences is based on procurement of the set of data for evaluation of various reliability indicators. However, application of the reliability theory requires gathering and processing of large amount of monitoring data. That is why collection of primary data is vital to ensure their further sorting and processing.

As far as the heating plant and operator of LDS is concerned, the primary data is collected by the head supervisor of control centre maintaining permanent presence at this site. The records contain mainly this type of data:

- information about the location and cause of failure,
- circumstances preceding the failure occurrence,
- course of the failure,
- cause of failure and its further impact operation of LDS,
- type of equipment/component affected,
- time of failure occurrence/extinction.

All events associated with operation of LDS are recorded by the Head Operations Officer at the control centre manually, using the Operation Log that is available to the Head Supervisor of Power Supply Control (HSPSC) as the material for sign off during the shift in the Lotus Notes interface. The application is based on a text document only and it is not designed for filtering of any items or events. The report on a failure occurring during particular shift will be generated by the attending supervisor, this report is necessary. Once completed, the report will be forwarded to the person authorised and responsible for evaluation of reliable LDS operation for further processing. That will help with prevention of accidental omissions of less serious failures.

3 Implementation of Methodology for Operation Data

The entire LDS is required to provide highly reliable service with a very low failure rate to avoid any failures in supply of electric power to end users, supported by the method of operation and backup employed [4].

3.1 Scope of Power Distribution System

Precise enumeration of potential reliability indicators shall be based on accurate information about the number of elements and their voltage levels. That is why these elements need to undergo the passportization process [5]. The local distribution system in question contains approximately 1400 cells and MV and HV power cabinets in 52 distribution plants within the LDS operated by the heat production facility of this company.

1. Cable runs, cable ducts and bridges
   Figure 1 shows an overview of cable lines 6, 22 and 110 kV. The lengths of cable ducts are in kilometres and these include all the cable outlets, cable connections and cables linked with the transformer for power transmission. These do not include the cable lines to technological and distribution transformers, motors and technological cable connections operated by other plants of the industrial enterprise.
2. Equipment on MV and HV distribution plants
   The summary of equipment on MV and HV distribution plants makes use of the single-pole diagrams for these plants with subsequent passportization of all switches, disconnectors and other fittings in particular cells. For the passportization chart, refer to Table 1. The number of elements and line lengths are further complemented by other essential data for evaluation of failure rate of the system concerned, i.e. which is the element age.

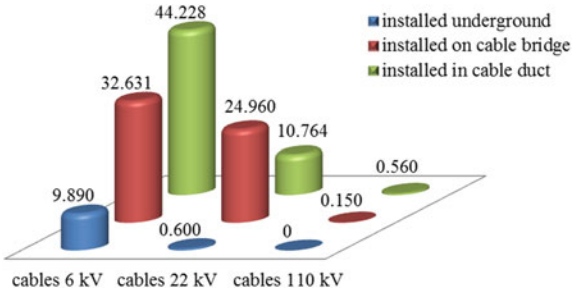


Fig. 1 Overview of the scope and installation of cable network rated for 6, 22 and 110 kV

Table 1 Passportization of equipment on MV and HV distribution plants

Device	Quantity	Device	Quantity
Bus disconnector 6 kV	1438	Outlet disconnector 22 kV	72
Bus disconnector 22 kV	132	Outlet disconnector 110 kV	1
Bus disconnector 110 kV	12	Transformer 110/22 kV	9
Disconnector ZN 110 kV	6	Transformer 22 kV/LV	3
Power switch 6 kV	826	Transformer 22/6 kV	20
Power switch 22 kV	72	Bus bar 6 kV	826
Power switch 110 kV	5	Bus bar 22 kV	78
Outlet disconnector 6 kV	430	Bus bar 110 kV	11

3.2 Processing of LDS Operation Documents

Gathering and subsequent analysis of a substantial amount of materials referring to operation of the LDS for the period of 2010–2014 allowed compiling the 5-year database to document the overall behaviour of the system with respect to its operation, maintenance and failures, including time stamps for each occurrence [6]. Calculation of maintenance indicators has been performed using an approach different to the regular standard [7]. Using the actual completion of maintenance operations on equipment, as initially defined by the record on shift changeover for HSPSC is more accurate than using values contained in the charts of internal regulations—the Revision Code to determine periodical performance of check-up and revision operations.

The volume of materials processed comprises 5 478 (5 years times 365 days (366 days in 2012) times 3 shifts per day) sheets in A4 format with "Shift Changeover Logs for HSPSC" duly completed by the Head Supervisor of Power Supply Control using the Lotus Notes interface.

The number of events linked to operation of the industrial LDS directly to be processed amounts to approximately 2 000 each year. That is up to 10 000 over the 5-year period. This is evidently not just a routine transcription of records on shift changeover by the HSPSC into electronic form, yet rather a specialised procedure conducted by the expert familiar with the system wiring. He must be also able to make a decision on specific categorisation of each event to filter anomalies that remain concealed to anyone not skilled in the field of electric power engineering.

3.3 Calculation of Element Reliability Indicators

Modest mathematical tool was employed for enumeration of the essential reliability indicators for elements over the 5-year period—Failure rate λ_p, Mean time to repair τ_p, Maintenance rate λ_u, Mean time to maintenance τ_u. These are listed in Table 2.

Table 2 Essential reliability indicators within the LDS observed

Device	λ_p	τ_p	λ_u	τ_u
	(year^{-1})[a]	(h)	(year^{-1})[a]	(h)
Bus disconnector 6 kV	0.004	9.077	0.231	15.011
Bus disconnector 22 kV	0.008	7.450	0.241	48.790
Bus disconnector 110 kV	0.000	0.000	0.000	0.000
Disconnector ZN 110 kV	0.000	0.000	0.600	48.466
Power switch 6 kV	0.025	5.538	0.238	13.845
Power switch 22 kV	0.039	5.702	0.289	62.522
Power switch 110 kV	0.000	0.000	0.000	0.000
Outlet disconnector 6 kV	0.002	6.358	0.193	15.810
Outlet disconnector 22 kV	0.000	0.000	0.244	63.282
Outlet disconnector 110 kV	0.000	0.000	0.000	0.000
Transformer 110/22 kV	0.333	4.403	1.111	74.679
Transformer 22 kV/LV	0.200	0.361	1.733	23.132
Transformer 22/6 kV	0.130	0.481	0.970	90.282
Cable 6 kV	0.032	8.104	0.284	108.640
Cable 22 kV	0.022	11.417	0.209	53.474
Cable 110 kV	0.000	0.000	3.380	51.922
Power line 110 kV	0.037	0.700	0.186	47.443
Bus bar 6 kV	0.000	0.000	0.014	15.421
Bus bar 22 kV	0.003	3.900	0.054	14.011
Bus bar 110 kV	0.000	0.000	0.018	5.250

[a]Cable lines are described by failure rate per km

3.4 Expression of Reliability Indicators on Power Lines

Probability of failure-free operation of power supply lines is not considered absolutely reliable, so the values will not be $R = 1$. The approach to this agenda is therefore very different and respects the actual operating conditions of power supply lines.

The actual values of R are not calculated using failure rates, maintenance rates, mean times to repair and mean times to maintenance for individual elements, because accurate information about reliability indicators of elements falling within the operations of distribution company is not available to the LDS operator. These values are calculated from outages recorded at the control centre without knowledge of lines (Table 3).

Table 3 Essential reliability indicators of power supply lines

Line 110 kV	λ_p	τ_p	λ_u	τ_u	R
	(year^{-1})	(h)	(year^{-1})	(h)	(–)
V1	1.20	0.197	3.60	15.439	0.99999460
V2	0.80	0.350	3.60	15.997	0.99999361
V3	2.60	0.378	3.20	23.870	0.99997756
V4	3.00	0.458	4.80	29.659	0.99996866
V5	1.80	0.500	5.60	85.955	0.99997946
V6	2.00	0.423	5.00	67.891	0.99998068
V7	2.00	0.463	2.20	31.129	0.99997885

4 Reliability Results Comparison

The importance of failure database with the option to determine reliability indicators is also based on the fact that there are actually certain databases showing failures of distribution networks within the Czech Republic, yet there is absolutely no failure database using data about industrial power distribution network to enable calculation of the essential reliability indicators. Unlike the reliability indicators for distribution networks within the Czech Republic, those are obviously different and cannot be subject to mutual comparisons.

Industrial grids differ by pattern, the method for connection into the grid or even geographical conditions, etc. Furthermore, power grids in heavy industry are also exposed to many adverse effects, e.g. presence of water, jolts, shaking, vibrations, heavier dust formation, high ambient temperatures or even occurrence of corrosive, chemical or other contaminating materials. That is why the element reliability results are very valuable for the existing voltage level of 6, 22 and 110 kV not subject to analysis yet.

Table 4 provides comparison of selected reliability indicators that have been enumerated by means of the operating design principles of 22/80 ČEZ for the period of 1975–1990 [8], updated results from the period of 2000–2014 [9, 10] and the 5-year period of 2010–2014 with respect to the local distribution system.

4.1 Comparison of Reliability Indicators for Power Lines

The level of quality in distribution systems is determined e.g. by the failure rate or system average interruption indexes. Figure 2 shows the comparison of failure rate at power supply lines V1–V7 into the LDS and the failure rate of the one Czech distribution system operator (DSO). It shall be pointed out that the significant differences between distribution grids operated by both entities prevent easy comparison of failure rates, since the results depend on the profile of particular

Table 4 Comparison of results for essential reliability indicators

Damaged equipment		ČEZ 22/80	Period	LDS
		1975– 1990	2000–2014	2010–2014
Cable 6 kV	λ (year^{-1})	Not included in the database	Not included in the database	0.032[a]
	τ (h)			8.104
Cable 22 kV	λ (year^{-1})	14.5	4.661	0.022[a]
	τ (h)	215	5.710	11.417
Transformer 110 kV/MV	λ (year^{-1})	0.04	0.058	0.333
	τ (h)	1300	0.231	4.403
Transformer MV/MV	λ (year^{-1})	Not included in the database	Not included in the database	0.130
	τ (h)			0.481
Transformer MV/LV	λ (year^{-1})	0.03	0.006	0.200
	τ (h)	2500	5.303	0.361
Power switch 6 kV	λ (year^{-1})	Not included in the database	Not included in the database	0.025
	τ (h)			5.538
Power switch 22 kV	λ (year^{-1})	0.015	0.012	0.039
	τ (h)	30	23.580	5.702

[a]With respect to the scope of LDS cable system, the resultant values of failure rate on cables rated 6 and 22 kV refer to each 1 km rather than 100 km, which is the case of the ČEZ cable system

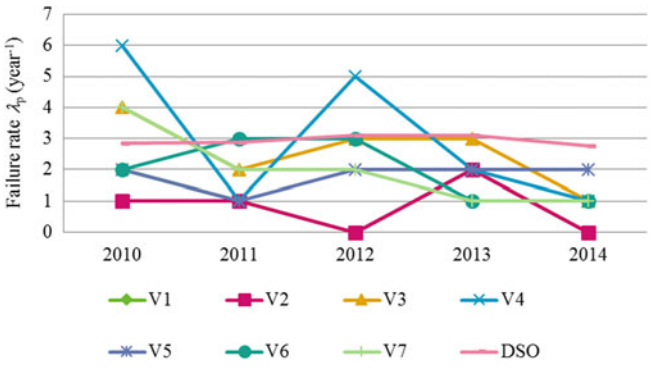


Fig. 2 Progress of failure rate on power supply lines

systems and other circumstances. Monitoring of progress of particular indicators over time is more important from this point of view.

The values of failure rate λp at the power supply line were approximately constant and showed rather a declining tendency over the past 2 years. One of the facts worth mentioning is the magnitude of failure rate on power line V4, which rose in years 2010 and 2012, compared to other power lines. These increases were due to the higher number of outages on line V4 caused by the superior grid.

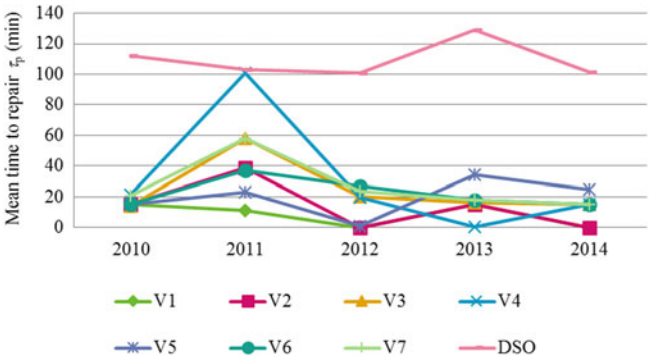


Fig. 3 Progress of mean time to repair on power supply lines

Value of mean time to repair τ_p indicators shown by Fig. 3 referring to power supply lines V1–V7 were also almost constant and their overall tendency was rather declining. Another thing worth mentioning is the mean time to repair on the line V4 with its value higher than other values in the year 2011. The increase was caused by short-circuit on the current transformer on the line V7 in the distribution plant R 110 kV in the superior system, which had a substantial impact on the voltage restoration period for the region supplied from V4 within the local distribution grid.

5 Conclusion

Gathering and subsequent analysis of a large amount of materials referring to operation of the industrial LDS in the period of 2010–2014 allowed compiling the database to document the overall behaviour of the system with respect to its operation, maintenance and failures, including time stamps for each occurrence. That will enable enumeration of the reliability indicators for significant elements within the system and their comparison against reliability indicators for other distribution systems.

Calculation of some reliability indicators has been performed using an approach different to the regular standard. Those are values of the maintenance rate and the mean time to maintenance. Both values have been calculated using the actual completion of maintenance operations on equipment, as initially defined by the record on shift changeover for HSPSC. Our approach is more accurate than using values contained in the charts of internal regulations—the Revision Code to determine periodical performance of check-up and revision operations.

Acknowledgments This research was partially supported by the SGS grant from VSB - Technical University of Ostrava (No. SP2016/95) and by the project TUCENET (No. LO1404).

References

1. Prokop, L., Kral, V., Hradilek, Z.: Software OCENS—a tool for outage costs and energy not supply calculation. Przeglad Elektrotechniczny **85**(3), 227–230 (2009)
2. Drholec, J., Gono, R.: Operation of industrial local distribution system. In: 2015 International Scientific Conference Electric Power Engineering (EPE), pp. 478–483
3. Kornatka, M.: Analysis of unreliability of low voltage power networks some branches of Distribution System Operators. Przeglad Elektrotechniczny **88**(10B), 303–306 (2012)
4. Barlow, R.E., Proschan, F.: Statistical Theory of Reliability and Life Testing: Probability Models. Holt, Rinehart and Winston, New York (1975)
5. Szkutnik J., Gawlak A.: Optimization-based method of dividing the network development costs. Electr. Eng. **93**(3) (2011)
6. Todinov, M.T.: Reliability and Risk Models: Setting Reliability Requirements. Wiley, Chichester (2005)
7. Brown, R.E.: Electric Power Distribution Reliability. Marcel Dekker, New York (2002)
8. Martínek, Z., Klor, T., Holý, J.: Reliability of the electrical power system. In: 15th International Scientific Conference on Electric Power Engineering (EPE), pp. 81–84 (2014)
9. Gono, R., Rusek, S., Kratky, M., Slivka, M.: Component reliability parameters of distribution network. In: 2015 International Scientific Symposium on Electrical Power Engineering (EPE), pp. 376–379
10. Petráš, J., Kurimský, J., Balogh, J., Cimbala, R., Džmura, J., Dolník, B., Kolcunová, I.: Thermally stimulated acoustic energy shift in transformer oil. Acta Acustica United Acustica **102**, 16–22 (2016)

Improved Search of Typical Projects of Private Houses with Using Data Mining

Alexander Greshnov

Abstract The goal of this study is to determine the methods and algorithms for solving search problems with uncertainty conditions for the selection of projects of private houses according to customer requirements. The problem of reducing the volume of the database sample is solved. The methods of data collection, storage and reporting are presented. Inputs are described. It is proposed to use intelligent analysis: segmentation, the search for associations, summarizing. The algorithms and selection of the most appropriate one are presented. The task of construction of the information system for the designer-analyst, providing a solution of this problem, is placed here. There are the requirements for it and a description of analogues in this paper.

Keywords Data mining · Search with uncertainty · Segmentation · The search for associations · Patterns search · Typical projects of private houses · A decision support system

1 Introduction

The construction of private houses is gaining great popularity in Russia and the world in general. Detail statistics information is presented by Federal State Statistics Service of the Russian Federation (web—http://gks.ru). The dynamics (2000–2014) of volume of new private buildings is shown in Fig. 1. In most cases, the customer has difficult problem to make the choice of specific design solutions for its future house. Often customer is ready to provide information about only the cost and size of the building, less often about the basic materials, and very rarely about the specific square of the rooms, heating technology, ventilation and others. For the designer there is the problem of selection of basic parameters of a private house and coordination them with the customer. Having a database of finished projects or wishes of

A. Greshnov (✉)
Ulyanovsk State Technical University, Ulyanovsk, Russia
e-mail: am.greshnov@gmail.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_48

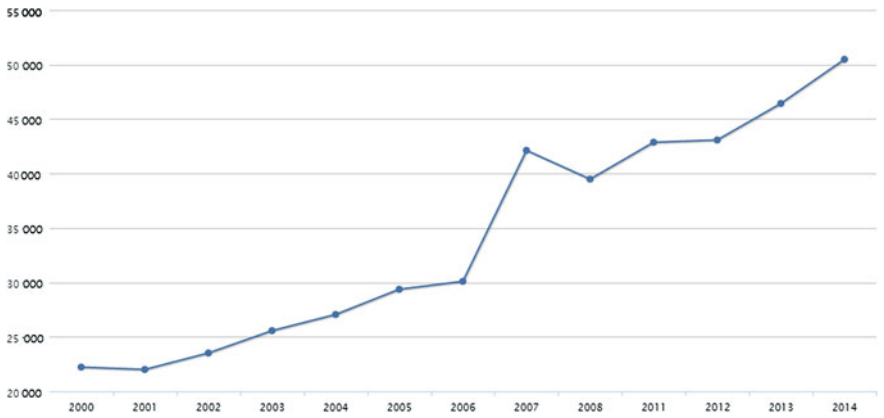


Fig. 1 Number of introduced individual housing buildings

former customers with the addition of expert data we can get sample from it, according to several of the requirements of the new customer. However, the result of getting sample will be received with quite a large number projects (especially, if the input data from the new customer will be minimum), which will require further analysis and the choice of several best options. Therefore, having a large amount of data about the projects of private houses, it is difficult to get useful information from this wealth of experience, which can help to create a project for a new customer.

Because of problem of heterogeneity and high-dimensional data exists, the solution is the use of data mining. To reduce the search results from the database with conditions of uncertainty it is necessary to detect a possible relationship between the characteristics (parameters) of house and choose the most common sets. The solution of this problem is the method for finding association rules [1].

As part of this work there are the following tasks:

1. Determination of the required parameters of house and allowable values;
2. Determination of ways to collect and store information about the projects of houses for the defined options;
3. Determination of ways to present the results of data analysis (recommended projects, individual design and technological solutions);
4. Determination of necessary methods and algorithms for data analysis.

Additionally, the task of the development of computer information system is presented. It will able to provide its users (employees of the design organization) support in the selection of the project of house for the customer from process of data collection to process of display the results of the analysis in an intuitive and convenient way.

Also the overviews of alternative methods, algorithms, and analogs of the proposed information system are presented.

This article is the first step in the description of the possible concept of the Data Mining application in the construction sector of economics, in this case, for the search of typical private houses projects.

2 Input and Output Data

2.1 Data for Analysis

During studying of construction of private houses 70 parameters has been identified, divided into 23 categories, which customers and designers pay more attention. Their potential values were determined. These options by categories are:

1. The cost of the house: minimum, maximum;
2. House area (without garage space): minimum, maximum;
3. The occupied house (with garage) land area: minimum width and length, maximum width and length;
4. Walls: type, cladding, color (facade) and coating (inside), the color coating (inside);
5. Roof: type, roofing material, color, self-cleaning, the presence of solar panels;
6. Basement (optional): the area and the height, comfort, decoration;
7. Attic (optional): the presence of windows, comfort, decoration;
8. Floors: number of floors, the number of stairs, the type of interfloor stairs, stairs material;
9. The climatic and other conditions: temperature, humidity, seismic resistance;
10. Rooms: amount, minimum and maximum area;
11. Kitchen: minimum and maximum area, compatible with the living room;
12. Hallway: minimum and maximum area;
13. Dressing room (optional): minimum and maximum area;
14. Bathrooms and wc: number of bathrooms, number of wc, type, arrangement;
15. Ceiling: level, coverage;
16. Flooring: floor covering, heated floor;
17. Windows: size, material profile, double pane windows, tinted, color, light shading;
18. Doors: type, material;
19. Ventilation: cooling, filtering;
20. Heating (+ hot water): the type of heater, the type of heating;
21. Fireplace (optional): type, amount;
22. Balconies/loggias: amount, size, glazing, comfort;
23. Garage (optional): capacity, comfort, decoration.

The collected data for these parameters are the starting point for data mining, as a result of which patterns and preferences of customers similarity of houses will be revealed when searching with uncertainty. As the request the values of only some house parameters may specified.

2.2 Collection and Storage

To store information about projects of houses the database of the following structure is proposed to use.

Tables for each house parameter with possible qualitative values include an identifier (id) and embodiment names (name) of values for this parameter. For example, the table "floor Heating" contains records (id–name): 1—water, 2—IR, 3—cable, 4—without heating, 5—other, and the table "wall Type" contains records (id—name): 1—brick, 2—concrete blocks, 3—gasblock, 4—board, 5—timber, 6—sandwich panels, 7—monolith, 8—other.

Table for storing information about the projects of houses includes a record identifier (id), the recording date (date) and field values for each parameter (numerical or foreign key from table of entries described above). Thus, we have 70 fields that describe the design of the house as a set of values of previously allocated parameters and two service fields (id and date). In this table entries are to be analyzed in this study as input data for the algorithms.

Collecting of customer feedback method is proposed to conduct the survey on the parameters, described above, followed by the recording of this information to the database. Previously created projects houses information can also be entered into the database. The more data is collected, then the result of the analysis is more valuable.

Database management system should support the optimal storage and processing of such data.

2.3 Presentation of Results

Results of the analysis data about house projects, during searching with uncertainty, must be provided, eventually, in the form of several, corresponding for search query, most typical projects of houses, described by set of values of the parameters, identified above in this article. Also found connections and dependences between the values of the parameters of the house will be useful for the analyst-designer. The results should be presented in a clear way as tables and graphs.

3 Data Mining

Knowledge Discovery in Databases (KDD)—the process of extraction knowledge from the data in the form of dependencies, rules, models that allow for modeling and forecasting of various processes. The sequence of actions to be undertaken to build the model in order to extract knowledge does not depend on the subject area. Stages of analysis:

1. Selection;
2. Pre-processing (cleaning);
3. Transformation;
4. Data Mining (Construction of models);
5. Interpretation/Evaluation.

The core of KDD is methods of Data Mining, enabling to detect patterns and knowledge. Data Mining—the process of decision-making support, based on a search for hidden patterns in the data, i.e., extracting information, which can be characterized as knowledge [2].

To solve the problem of searching in the database of projects of houses, which are relevant to customer requirements, it is proposed to use segmentation, the search for associations and summarizing. These processes of data mining within this study are described below.

3.1 Segmentation

Segmentation—selecting of segments of private houses. It is proposed to apply the method of grouping data, using a description of the data in the class—the segment. It is used for getting the original sample (filtering) data from the database, according to the basic specifications by customer. The data about projects are divided into segments by parameters, such as cost, space or processability, thus the amount of information for later analysis is reducing.

The grouping data algorithm. The essence of it for segmentation of the data is: for the selected parameter, e.g., home area of 150–250 m^2, we select data, in which the parameter coincides with a predetermined. Further, from these filtered data the most frequently repeated values for each parameter are finding. The output is a set of parameters with the corresponding values for the selected segment. Specifying a number of alternatives (i.e., different values) and of the support (i.e., a greater percentage (frequency of occurrence) to take the value of the parameter into account), the description of the segment to obtain a specific parameter values with their proportion (percentage) of all its values in the selected segment.

After receiving of a description of a certain segment of data, it can already be used in the selection of the project for the client, if the resulting value of the support options for the parameters are large enough, according to the analyst-designer. For example, for a segment of the selected area, the value of "material window profile"—"plastic" has support 89 %, it can be almost uniquely set for a typical project of the house in this segment.

3.2 Search for the Associations

To define common sets of objects in a large variety of sets it is proposed to use the search method of association. It is used to detect dependencies between the values of the house parameters, thus allowing to determine the most common variants of the house components (design and technological solutions). Input data are the results of filtering, which was received at step of segmentation.

Basic concepts. The ratio of the number of entries, which includes a set of F, to the total number of records is called the support of F and denoted Supp (F). During searching, analyst can specify the minimum support value of the sets he has interest. Set is called frequent (large item set), if it has support more than the minimum support value specified by the user.

The results, which are obtaining in solving this problem, were decided to present as association rules. In this regard, for their search there are two stages: to find all the frequent sets of objects and to generate association rules of the found parts of a set of objects.

There are three types of association rules:

- Useful rules. They contain the actual information that was previously unknown, but it has a logical explanation. Such rules can be used to make decisions that benefit;
- Trivial rules. They contain real and easily explained the information that is already known. Such rules can not benefit because either reflect the known laws of the study area, or past performance. Sometimes these rules can be used for verification implementation of decisions taken on the basis of previous analysis;
- Undetermined rules. They contain information that can not be explained. Such rules may be obtained based on outliers, or latent sensitive knowledge. Such rules can not be directly used for decision-making, as they is inexplicable and can lead to unpredictable results. To better understand the additional analysis is required.

Support shows the percentage of records, which support the rule. Confidence indicates the likelihood that the presence of recording in the set X implies the presence there in set Y. If confidence is greater then rule is better, in the rules, which was constructed on the basis of the same set. Unfortunately, the confidence can not determine the usefulness of the rule.

Improvement indicates whether the rule is more useful than random guessing. Improvement of the rule is the ratio of the number of records, which contain sets of X and Y, to the product of the number of records, which contain a set of X, and the number of records, which contain a set Y. If the improvement is greater than one, it means that with a set of rules to predict the presence of Y likely than random guessing, if less than one, then other [3].

Algorithms. A primitive way to find all sets of elements for association rules (simple search) requires a lot of run-time (about a power function (2^i), where i—the number of elements). There are several well-known algorithms for the solution of this problem, they are described below.

The FP-growth algorithm—one of the most effective and avoids not only costly procedure of the candidates generating, but also multiple scanning of the input set. The method is based on a pre-processing of the input and converting it to a special tree compact structure FP-tree (frequent pattern tree), and only then the computation of the popular-sets is doing [4]. Disadvantages of this algorithm: the construction of a tree—time consuming operation, in some cases, due to the large number of nodes and links, the size of FP-tree can far exceed the size of the input data set.

The first step of Eclat algorithm [5] is data converting to a vertical plan view (so-called set-TID), and next operations are working with it. In this representation support value is the ratio of power of the set to the total number of records. The steps of this algorithm are similar to the corresponding steps of the algorithm Apriori (see below), except calculation functions for support of the candidate, which now does not require scanning the database. Disadvantages of the algorithm: TID-sets may be too large, so the operation with them can take a long time, a large number of candidates is generating with low level of the minimum support value.

To solve this problem it is encouraged to use the most common algorithm for mining association rules—Apriori. Input data can be used as either all records in the database, or only the previously selected segment.

Apriori algorithm. The algorithm uses one of the properties of the support, which states: support of any set of elements can not exceed the minimum support of any of its subsets. The algorithm defines common sets by several stages. In the i-th stage, the all common i-element sets are defining. Each stage consists two steps: the first—the formation of candidates (candidate generation), the second—counting of support values of candidates (candidate counting). The algorithm is completed either at the completion of the specified number of steps, or in the absence of a certain step of partial sets because of the further generation of candidates is not possible [6]. Algorithm Advantages: simplicity; a rapid decrease of the number of generated candidates by setting high minimum support, or when relatively sparse basic set is presented. Disadvantages: multiple scanning of base set; a large number of generated candidates, when the data set is too large, or too low support value is set.

Input data for the algorithm of search for association rules (in this case, Apriori) must be converted to a binary form. The number of columns in the transformed data table equals the number of elements, which present in the set of records. In the corresponding column value is 1, if the item is presented in the record, and 0—otherwise. The number of columns, which are based on the possible values of the house parameters, can exceed 200 items.

The choice of this algorithm is due to its relatively simple implementation and the possibility of applying optimizations to reduce the number of scans of input sets, reduce the number of generated candidates and parallelization [7]. One of the most effective methods of implementation of the program is described in Borgelt "Efficient Implementations of Apriori and Eclat" [8].

3.3 Summarization

Summarization—a description of the specific groups of objects from the analyzed data set. It is proposed to convert the received results of segmentation and search for associations to linguistic form, using spreadsheets and other imaging agents.

Because of the summarization algorithm is more dependent on the results of the previous steps of data mining, then there is no universal method. For the segmentation results sentences are based in the form: "In the segment of private houses projects with up to 250 m^2 typical ones are following: Project 1 – the cost is 5 million rubles, the type of walls: gasblock, floors: 2, … Project 2 …". For results of search for associations—"Among the projects sample configuration sets of house are following: 1) living area is 200–300 m^2, 2 parking spaces in the garage, full sized basement, 3 balconies, 2)…."

4 The Proposed System

System solutions for the problem of typical house projects search with uncertainty conditions, which are based on the wishes of the customer, must provide the following functions. The system should be easy to use and it is obliged to reduce the time costs of analyst-designer. The basic requirements are following.

Database:

- Providing access to add (edit, delete) houses projects in the database, in the form of parameter values described above;
- Providing access to clean database;
- Providing access to adding wishes of customers in the database with a questionnaire forms.

Data mining:

- Enabling the analysis of segmentation methods, research associations and summarization based on the data from the database;
- Allowing to set the customer wishes to specify the parameters of the house (search query);
- Enabling to set analysis algorithms;
- Analysis progress is showing.

Results of the analysis:

- Providing access to viewing the results of solving segmentation tasks, search for associations and summarization;
- Access to view the consolidated result—the typical configurations of the houses for the search query;
- Opportunity to save the results.

Wishes of the customer should be compared with the parameters of the structural elements of the house and their price (optional). The result is selected several projects of house, which are described as the set of values of parameters of structural elements, that are relevant to the customer's wishes.

The famous buildings design systems (ArchiCAD, AutoCAD, ArCon and others) do not have built-in data analysis functions to search for versions of the house project or individual optimal design solutions, according to several customer requirements, from previously collected data about the houses projects and customers' wishes.

To solve this problem, the program product for support of construction of private houses, using the built-in analytics on the basis of Data mining, is proposed. The system can be designed as a separate module to Computer-Aided Design (CAD) systems for giving advices during creating a project of house.

5 Analogues of the Proposed System

Systems, which are designed to solve problems, which are above stated in the proposed system, were not found in this domain during searching on the Internet. Results of search for systems, which are used in the construction of houses, are following.

There are construction management systems (e.g., ActiveMap Building from web—http://gradoservice.ru), the systems to urban planning and tasks management (e.g., Information systems of Scientific and Practical Center "The development of the city" from web—http://gisa.ru). However, in their description there is not information on the results, which are obtaining in the form of standard solutions (projects of houses). Below analytics and statistics systems are presented, which can solve some problems, using Data Mining.

5.1 1C: Enterprise

Among the universal data analysis systems it can be identified quite popular and productive software 1C: Enterprise, which includes the data analysis component, presented on web—http://1c-dn.com. On the information page, the software provides information on the basis of which we can judge that the system has a fairly high functionality and has the ability to organize on its platform the proposed system, because it implements the required algorithms, except segmentation (not stated in the summary).

5.2 STATISTICA

Also among the universal systems it can pay attention to the STATISTICA system is powered by StatSoft (an example of segmentation) from web—http://statsoft.ru. On the software information page it provides detailed information about the data segmentation and usage of this Data Mining problem. The interface is quite simple in appearance, but for the end user, i.e., for the designer-analyst, it may not be the most convenient and informative, because a lot of specific data analysis parameters are here.

It was not found similar analogues of the proposed system. For greater system specifications it must be more detailed study and optimization of algorithms for the building sphere and the construction of detached system for the most comfortable perception of it by designer-analyst. In the system the emphasis should be done at the convenience of the user perception, the data connection with the sphere of building, particularly with projects of private houses.

6 Conclusions

For the problem of typical house projects search with uncertainty conditions, which are relevant to customers' wishes, from the available data about the projects for goal to reduce the resultant sample such methods of intellectual analysis as segmentation, associations search and summarization were determined. Their algorithms, input data, the method of collecting and storing output data were described. The requirements for information system, realizing this task and designed for the designer-analyst of private houses, were presented. Descriptions of analogues of the proposed system were presented.

The test version of the software system of supporting the design of private houses, which uses methods of Data Mining and allows the designer to reduce the time costs in the selection of options for the customer's house, was developed. It is the.NET 4.0 solution (C#) with WinForms user interface and SQLite database. Experimental input data was presented by 300 records of house projects. More than 100 search queries were done. For example, searching by 10 parameters of house, the results volume was reduced average 80 %, compared with simply data filtering. In the worst case, the analysis process takes about 2 minutes. Also more than 70 frequently occurring sets of house parameters (with different support value and size) were found in all experimental input data.

The goal of the future work is to optimize the algorithms for a particular tasks and the creation of a full-fledged software product as a decision support system for designer-analyst of the construction company. It is also considering the possibility of using some of the customers' personal data such as age, family composition, ethnicity, status in society, and others, during analyzing of the data for more personalized recommendations of house parameters values.

References

1. Crawford, I.M., Lomas, R.A.: Factory analysis—a tool for data reduction. Eur. J. Mark. **14**(7), 414–421 (1980)
2. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco (2011)
3. Zhang, C., Zhang, S.: Association Rule Mining. Models and Algorithms. LNAI, vol. 2307, pp. 25–46. Springer, Heidelberg (2002)
4. Han, J., Pei, J., Yin, Y., Mao, R.: Mining of frequent patterns without candidate generation: a frequent-pattern tree approach. In: Data Mining and Analysis Discovery, vol. 8, no. 1, pp. 53–87. Kluwer Academic Publishers, Massachusetts (2004)
5. Zaki, M.: Scalable algorithm for association mining. IEEE Trans. Knowl. Data Eng. **12**(1), 372–390 (2000). IEEE Press, New York
6. Wang, P., Shi, L., Bai, J., Zhao, Y.: Mining association rules based on Apriori algorithm and application. In: International Forum on Computer Science-Technology and Applications, vol. 1, pp. 141–143. IEEE Computer Society, Los Alamitos (2009)
7. Pol, U.: Design and development of Apriori algorithm for sequential to concurrent mining using MPI. Int. J. Comput. Technol. **10**(7), 1785–1790 (2013). CIR
8. Borgelt, C.: Efficient implementations of Apriori and Eclat. Workshop on Frequent Itemset Mining Implementations. ACM Press, New York (2003)

Author Index

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3