Advances in Intelligent Systems and Computing 450

Ajith Abraham
Sergey Kovalev
Valery Tarassov
Václav Snášel *Editors*

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

Volume 1

Springer

Advances in Intelligent Systems and Computing

Volume 450

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.

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 1

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 978-3-319-33608-4 ISBN 978-3-319-33609-1 (eBook)
DOI 10.1007/978-3-319-33609-1

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 during 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 above-mentioned conferences 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 M. Kovalev, Rostov State Transport University, Russia
Václav Snášel, VŠB-TU Ostrava, Czech Republic
Ajith Abraham, MIR Labs, USA

Conference Vice-Chair

Valery B. 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

Organizing Committee Chair

Local Organizing Committee

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 I
Invited Papers

Cognitive Reasoning Framework: Possibilities, Problems, Prospects

Oleg Anshakov and Tamás Gergely

Abstract The paper discusses a general scheme of constructing different systems of artificial intelligence and data mining. This scheme interprets various intelligent technologies as kinds of reasoning. All of these kinds of reasoning aim to cognition and formation of domain models. We assume that reasoning has a referential character, i.e. reasoning can use semantic arguments as well as syntactic rules of deduction. We use some non-classical logics to formalize cognitive reasoning.

Keywords Reasoning · Cognition · Data mining · Domain model · Non-classical logics · Cognitive reasoning framework

1 Introduction

In book [1], we introduce cognitive reasoning framework (CRF) as three-leveled structure, which includes conceptual, formal, and implementational levels.

(i) At the conceptual level we present an informal model of a cognizing agent (CA). This model consists of different units that implement various functions of cognizing agent. The model also depicts how information circulates between the agent's units and how information transforms due to some unit's activity. At this level we also describe the behavior of a cognizing agent (reasoning, perceiving, memorizing and so on).

(ii) At the formal level we define some classes of non-classical logics and a general scheme of creating formal theories. We consider two kinds of deduction technique: traditional and original (non-monotonic). The second kind leads to

O. Anshakov (✉)
Russian State University for the Humanities, Moscow, Russia
e-mail: oansh@yandex.ru

T. Gergely
Applied Logic Laboratory, Budapest, Hungary
e-mail: gergely@all.hu

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 450, DOI 10.1007/978-3-319-33609-1_1

constructing the so-called modification calculi and modification theories. Here 'modification' means that one can revise truth valuation of sentences within an inference of a specific type.

(iii) At the implementation level we propose an object model of a cognitive reasoning system (CRS) that includes objects and collections for representing the cognizing agent's units and its activity. At this level, one can express:

- Data representation for the results of perceiving
- Knowledge representation and manipulation
- Formalization of cognitive reasoning

We expect that the above object model is quite adequate representation of the so-called "cognitive engine" which can be consider as the "brain" of an artificial cognizing agent.

In this paper, we revise substantially the logical language for CRF. In [1], we use logical languages with special connectives, which represent characteristic functions for different sets of truth values. These connectives are called J operators. They was introduce by Rosser and Turquett [2].

In [1], we introduce original inference technique, which we call *modification calculi*. Namely modification calculi allow us to represent dynamic and non-monotonic inference where we can use semantic consideration that was interpreted syntactically. However, the logical languages from [1] permit us to define modification calculi only for first order logics. These languages have no expressive tools for defining this technique for propositional logics.

In this paper, we introduce special connectives for cause-effect relation, which allow us to define a logical language, that permit us to consider modification calculi for propositional logics. This lead us to a noticeable simplification of the example of modification inference with temporal contradiction (compare inference from Sect. 3 in this paper and inference from example 14.2.28 in [1]).

Notice that in [1], we have to use different logical languages for different logics with J operators. The logical language from this paper *can be the same* for different logics with different sets of truth values. This allows us to consider interesting problems for suitable class of logics. However, in this paper, we regard only the simplest logic from this class.

2 Cognizing Agent: General Scheme

Now we consider a general scheme of a cognizing agent. We expect that this scheme represents the structure and the behavior of the various kinds of these agents. We suppose that cognizing agent can be (for example) a living entity (particularly a human being), or a group of them or a technical system, which can adapt to changing environmental conditions.

Note that the structure we discuss below includes units, which are contained in different models of cognitive architecture, that are defined by different researchers. Our contribution consists in detailed description of a cognizing agent's behavior in various situations. The set of constituents of a cognizing agent is quite usual.

The most common architecture of cognitive systems include the following subsystems:

- A sensing subsystem, which can receive signals from the environment
- An information-processing subsystem, which creates a "world model" for the agent; this model can be used by the agent for making decisions
- An effecting subsystem, which performs actions, that are conclusions of the agent's decisions: we suppose that the actions can modify environment

We suppose that one can use different methods and techniques for constructing the world model of a cognizing agent. For instance, the world model can be formed by methods of machine learning or soft computation.

The basic constraints of the proposed model are as follows:

- A cognizing agent extracts knowledge by the use of *reasoning in a broad sense*. The agent can infer deductively, argue plausibly, and use computational procedures.
- A cognizing agent acts *discretely*; its activity includes alternating phases of perceiving and reasoning;
- A cognizing agent forms a *history* of its activity

We require the agent's activity to be discrete because this makes formation of algorithms for the agent's behavior more easy. However, we think that even natural cognizing agents cannot perceive and analyze information simultaneously.

Note that we mean a cognizing agent as a *researcher*, but not as a hunter or a soldier. Therefore, we are interested in its abilities such as capacity of extracting knowledge, accuracy of reasoning, etc.

We consider the history of activity of a cognizing agent as a theory of a special type. This theory will be called *open cognitive* or *quasiaxiomatic theory*.

A cognizing agent consist of the following subsystems:

- A sensing subsystem
- An information-processing subsystem
- An effecting subsystem
- The long-term memory for storing the history of the cognizing agent's activity

The *sensing subsystem* receives signals from the environment where we can distinguish the physical and the information part. The signals from the physical environment are used for forming data. The information environment is understood as a set of other cognizing agents or their components which have the possibility to transfer information. One can interpret signals from the information environment as additional information, advice or instructions (Fig. 1).

Long-term memory stores history of the discrete activity of the cognizing agent, which we consider here as a theory of a special kind. This history can be interpreted

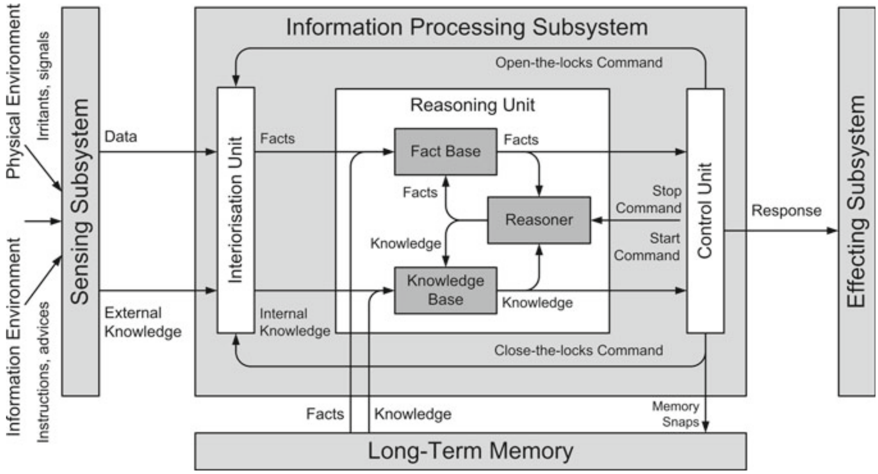


Fig. 1 General scheme of cognizing agent

as a sequence of *knowledge states*. The functioning of the long-term memory will be described in detail in the next chapter.

Consider the information-processing subsystem more detailed than the other subsystems. *The information-processing subsystem* contains the following units:

- The interiorization unit
- The reasoning unit
- The control unit

The agent uses *interiorization unit* for the accumulation of the information arriving from the environment and for its transformation into an internal representation (an internal format) of the cognizing agent. The data and knowledge are collected in the corresponding buffers (the *data buffer* and the *knowledge buffer*). This accumulation can occur in parallel with other actions of the cognizing agent, for example cognitive reasoning.

When it becomes necessary, the data and knowledge from the buffers will be transformed by corresponding converters (the *data converter* and the *knowledge converter*) into an internal format of the cognizing agent. The data in the internal format of the cognizing agent are called *facts*. The transformation of the data and external knowledge into facts and internal knowledge is called *interiorization* of data and knowledge.

The *reasoning unit* is the kernel of the considered subsystem. It contains the *fact base*, the *knowledge base* and the *reasoner* (the mechanism of reasoning). The fact base and the knowledge base together form the short-term (working) memory. The reasoner introduces changes in the working memory during its operation (both in the fact base and in the knowledge base). Changes in the knowledge base can be interpreted as acquisition of new knowledge, and changes in the fact base can be interpreted as generation of predictive hypotheses.

3 Modification Calculi and Modification Theories

Modification theories locate at the formal level of CRF. These are non-monotonic and based on specific non-classical logics. In [1], we use the so-called finite pure J (FPJ) logic to formalize modification theories. Here J means J operator, which was introduced by Rosser and Turquett in [2]. The more general class of logics with J operators than the class of FPJ logics was considered in [3]. The algebraic models of such logics was studied in [4]. The class of FPJ logics is interesting. However, now we can apply more convenient logics for formalization of modification theory. Particularly, we can represent modification rules by means of some propositional logics.

Consider an example of such logics and rules.

3.1 Syntax

Definition 1 The *alphabet* of the language of logic **M** contains the following groups of symbols:

(i) A countable set of internal propositional variables
(ii) A countable set of external propositional variables
(iii) The classical propositional connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow, \bot, \top$
(iv) The special connectives $?, !, \overset{?}{\rightarrow}, \overset{!}{\rightarrow}, :\ \rhd$.

As for the arities of the above connectives, we say:

- $\bot$ and $\top$ are nullary connectives;
- $\neg, ?,$ and $!$ are unary;
- $\wedge, \vee, \rightarrow, \leftrightarrow, \overset{?}{\rightarrow}$, and $\overset{!}{\rightarrow}$ are binary.
- $:\ \rhd$, these two symbols are used for representing a ternary connective.

Definition 2 Let us define some kinds of *atomic formulas*:

(i) The internal propositional variables are called *internal atomic formulas*.
(ii) The expressions $?p, !p, p \overset{?}{\rightarrow} q, p \overset{!}{\rightarrow} q$, and $p : q \rhd s$, where p and q are internal propositional variables, are called *J-atomic formulas*.
(iii) Each external propositional variable and each J-atomic formula are *external atomic formulas*.

Remark 1 The informal interpretation of $?p$, $!p$, $p \overset{?}{\rightarrow} q$, and $p \overset{!}{\rightarrow} q$ are the following:

- $?p$ is "p is unknown"
- $!p$ is "p is known"
- $p \overset{?}{\rightarrow} q$ is "it is unknown that p cause q"

- $p \overset{!}{\rightarrow} q$ is "it is known that p cause q"
- $p : q \rhd s$ is "it is known that in instance p of phenomenon q there exist a circumstance s"

Definition 3 The internal atomic formulas are called *internal formulas*. Let us inductively define external formulas as follows:

Basis: Each external atomic formula and each J-atomic formula are *external formulas*. $\bot$ and $\top$ are *external formulas*.

Induction step: Let φ and ψ be *external formulas*. Then $(\neg\varphi)$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \rightarrow \psi)$, and $(\varphi \leftrightarrow \psi)$ are *external formulas*.

Agreement 4 We have to agree on omission of parentheses for simplifying the representation of external formulas. We set the priority of the connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$ as follows:

(i) $\neg$, !, and ? have the same priorities, their priority is the highest;
(ii) $\wedge$ and $\vee$ have the same priority and their priority is lower than priority of $\neg$;
(iii) $\rightarrow$, $\leftrightarrow$, $\overset{!}{\rightarrow}$, $\overset{?}{\rightarrow}$, and $: \ \rhd$ have the same priority and their priority is lowest.

We agree to omit the external parentheses of an external formula and to omit parentheses if it is possible due to priority of its connectives.

3.2 Semantics

Definition 5 The *algebra of formulas of* **M** is a two-sorted algebra

$$\mathcal{F} = \left\langle \mathcal{E}, \mathcal{I}; \bot, \top, \neg, \wedge, \vee, \rightarrow, \leftrightarrow, ?, !, \overset{?}{\rightarrow}, \overset{!}{\rightarrow}, : \ \rhd \right\rangle$$

where $\mathcal{E}$ is the set of external formulas, $\mathcal{I}$ is the set of internal formulas, the operations of this algebra are define as usual. This algebra is an absolutely free algebra of its own type.

Definition 6 The *algebra of logic* **M** is a two-sorted algebra

$$\mathcal{M} = \left\langle \{\mathbf{0}, \mathbf{1}\}, \{?, !\}; \bot, \top, \neg, \wedge, \vee, \rightarrow, \leftrightarrow, ?, !, \overset{?}{\rightarrow}, \overset{!}{\rightarrow}, : \ \rhd \right\rangle$$

where **0** and **1** are classical **False** and **True**, respectively, the informal interpretation of ?, ! are **Unknown** and **Known**, respectively, operations $\bot$, $\top$, $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$ are the usual classic logical operation, operations ?, !, $\overset{?}{\rightarrow}$, $\overset{!}{\rightarrow}$, and $: \rhd$ are defined as follows:

$$?\,a = \begin{cases} \mathbf{1} & a = ? \\ \mathbf{0} & \text{otherwise} \end{cases} \qquad !\,a = \begin{cases} \mathbf{1} & a = ! \\ \mathbf{0} & \text{otherwise} \end{cases}$$

$$a \overset{?}{\to} b = \begin{cases} \mathbf{1} & a = \,! \text{ and } b = \,? \\ \mathbf{0} & \text{otherwise} \end{cases} \qquad a \overset{!}{\to} b = \begin{cases} \mathbf{1} & a = \,? \text{ or } b = \,! \\ \mathbf{0} & \text{otherwise} \end{cases}$$

$$a : b \rhd c = \begin{cases} \mathbf{1} & a = \,? \text{ or } c = \,? \text{ or } b = \,! \\ \mathbf{0} & \text{otherwise} \end{cases}$$

Definition 7 We can define a *tautology* by the usual way. We also can define *semantic entailment* as usual. We write $\vDash \varphi$ if φ is a tautology (of course φ is an external formula). We write $\Gamma \vDash \varphi$ if φ is *semantic conclusion* of Γ where Γ is a set of external formulas.

Remark 2 Let φ be an external formula. Then $\vDash \varphi$ if and only if $\emptyset \vDash \varphi$.

3.3 Calculus

Definition 8 (*Axioms*) Let φ, ψ, and θ be external formulas, c, p, and q be internal propositional variables; then the following external formulas are *axioms*:

(1) $\varphi \to (\psi \to \varphi)$
(2) $(\varphi \to (\psi \to \theta)) \to ((\varphi \to \psi) \to (\varphi \to \theta))$
(3) $\varphi \wedge \psi \to \varphi$
(4) $\varphi \wedge \psi \to \psi$
(5) $\varphi \to (\psi \to \varphi \wedge \psi)$
(6) $\varphi \to \varphi \vee \psi$
(7) $\psi \to \varphi \vee \psi$
(8) $(\varphi \to \theta) \to ((\psi \to \theta) \to (\varphi \vee \psi \to \theta))$
(9) $(\varphi \to \psi) \to ((\varphi \to \neg\psi) \to \neg\varphi)$
(10) $\neg\neg\varphi \to \varphi$
(11) $(\phi \leftrightarrow \psi) \to (\varphi \to \psi)$
(12) $(\phi \leftrightarrow \psi) \to (\psi \to \varphi)$
(13) $(\psi \to \varphi) \to ((\psi \to \varphi) \to (\phi \leftrightarrow \psi))$
(14) $\top$
(15) $\neg\bot$
(16) $?p \leftrightarrow \neg\, !p$
(17) $\left(p \overset{?}{\to} q\right) \leftrightarrow \neg \left(p \overset{!}{\to} q\right)$
(18) $(c : p \rhd q) \leftrightarrow \left(!c \to \left(q \overset{!}{\to} p\right)\right)$

Definition 9 (*Modus ponens*) The rule

$$\frac{\varphi, \quad \varphi \to \psi}{\psi}$$

is called *modus ponens*.

Definition 10 We can define an *inference*, a *proof*, and the relations of *derivability* and *provability* as usual. We write $\Gamma \vdash \varphi$ if φ is derivable from Γ where Γ is set of external formulas (the set of hypotheses). We write $\vdash \varphi$ if φ is provable.

Remark 3 Let φ be an external formula. Then $\vdash \varphi$ if and only if $\emptyset \vdash \varphi$.

Theorem 11 (Soundness Theorem) *Let Γ be a set of external formulas, φ be an external formula. Then*

$$\Gamma \vdash \varphi \quad \textit{implies} \quad \Gamma \vDash \varphi$$

Theorem 12 (Completeness Theorem) *Let Γ be a set of external formulas, φ be an external formula. Then*

$$\Gamma \vDash \varphi \quad \textit{implies} \quad \Gamma \vdash \varphi$$

3.4 Modification Calculus

Definition 13 Let us define some non-reliable rules, which are the additional rules for **M**:

(i) The rule

$$\frac{p \overset{?}{\rightarrow} q, \quad c_1 : q \rhd p, \quad c_2 : q \rhd p}{p \overset{!}{\rightarrow} q}$$

where $c_1 \neq c_2$ (i.e. c_1 and c_2 are different internal variables) is the *induction rule*.

(ii) The rule

$$\frac{?q, \quad !p, \quad p \overset{!}{\rightarrow} q}{!q}$$

is the *prediction rule*.

(iii) The rule

$$\frac{?p, \quad !q, \quad p \overset{!}{\rightarrow} q}{!p}$$

is the *abduction rule*.

Remark 4 The informal explanation of modification rules is the following:

(i) Suppose we do not know that p causes q. Assume that two different instances (c_1 and c_2) of one phenomenon q have the same circumstance p. Then we can suppose that we know p to cause q. This is informal comprehension of the *induction* rule. Note that this rule is a rather superficial formalization of John Stuart Mill's first canon (method of agreement): *If two or more instances of the* *phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon*. (See [5].)

(ii) Suppose we do not know q and we know p. Assume that we know p causes q. Therefore we can suppose that we know q. This is informal comprehension of the *prediction* rule. This rule is similar to *modus ponens*. Such a rule is often called *deduction*. But the term "deduction" is usually applied to reliable reasoning. However, this rule is actually non-reliable because the assertion "we know p causes q" is not necessary true but only known.

(iii) Suppose we do not know p and we know q. Assume that we know p causes q. Therefore we can suppose that we know p. This is informal comprehension of the *abduction* rule. This rule is used to form a possible explanation for the occurrence of a phenomenon: q is, why? because p is, so p causes q.

Remark 5 Note that we cannot use the rules from Definition 13 in the same manner as we use reliable rules such as modus ponens. We obtain a contradiction immediately. Therefore we have to define a special dynamic inference for to represent updating of estimation of internal atomic formulas. In this inference, we can hide rows that lost their relevance. We also hide all rows that depend on hidden rows.

Notice that formal definition of dynamic inference in our sense is too long and it demands to define some auxiliary constructs. Therefore we have to introduce some constructs informally.

The table below represents an example of modification inference. The last column contains the number of the inference step (i.e. row number) where the current row has been hidden. The number 0 (in the last column) means that the current row is not hidden.

#	Formula	Reason	Hidden at
1	$!p$	Hypothesis	0
2	$?q$	Hypothesis	13
3	$p \overset{?}{\to} q$	Hypothesis	7
4	$d_1 : q \rhd p$	Hypothesis	0
5	$d_2 : q \rhd p$	Hypothesis	0
6	$\left(p \overset{!}{\to} q\right) \to (!p \to !q)$	Additional hypothesis	0
7	$p \overset{!}{\to} q$	Induction (3, 4, 5)	0
8	$!p \to !q$	Modus ponens (7, 6)	0
9	$!q$	Modus ponens (1, 8)	0
10	$!q \to (?q \to \bot)$	Tautology	0
11	$?q \to \bot$	Modus ponens (9, 10)	0
12	$\bot$ (Contradiction)	Modus ponens (2, 11)	13
13	$!q$	Prediction (2, 1, 7)	0

This example displays the so-called temporal contradiction. The contradiction appears at the step (row) number 12 and disappears at the step (row) number 13.

4 Open Problems and Promising Areas of Research

Here we discuss some open problems that connected with cognitive reasoning framework. We group these problems by themes.

Theme 1. Modification calculi: further development
Modification calculi have many restriction, which reduces their ability for expressing the natural reasoning process. Relaxing these restrictions allows us to solve different important problems.

Some problems:

- Problem 1.1: to define various logical languages (propositional and first-order) for modification calculi
- Problem 1.2: to define logics without J operators that can be used in modification calculi
- Problem 1.3: to define generalized modification inference that can use more complicated algorithms of formalized reasoning

Theme 2. Kinds of similarity: formalization and investigation
Recognizing the similarity is the main part of induction. We can regard similarity as the common part (meet) of objects under consideration. This kind of similarity is used in the JSM method proposed by V.K. Finn (see [6, 7]), more informal description of the JSM method is in [8]. Another approach to similarity is to measure the degree of similarity of different objects. We can find technologies that based on this approach in many work on machine learning and data mining. It is interesting that calculating the degree of similarity can be combined with general scheme of JSM method (see [9]).

Some problems:

- Problem 2.1: to find regularities that are common for similarity as meet (intersection) and similarity as degree of likeness
- Problem 2.2: to develop algorithms for calculating or constructing similarity and minimize the complexity of these algorithms;

Theme 3. Modification alculi with generalised quantifiers
We increase the abilities of modification calcili to express informal reasoning if we use generalized quantifiers. For example we can define various statistical quantifiers that are similar to quantifiers from [10].

One problem:

- Problem 3.1: to define and to study modification calculi with generalized quantifiers

References

1. Anshakov, O., Gergely, T.: Cognitive Reasonimg: A Formal Approach. Springer (2010)
2. Rosser, J.B., Turquette, A.R.: Many-valued logics. Br. J. Philos. Sci. **5**(17), 80–83 (1954)
3. Anshakov, O.M., Finn, V.K., Skvortsov, D.P.: On axiomatization of many-valued logics associated with formalization of plausible reasonings. Stud. Logica. **48**(4), 423–447 (1989)
4. Anshakov, O., Rychkov, S.: On finite-valued propositional logical calculi. Notre Dame J. Formal Logic **36**(4), 606–629 (1995)
5. Mill, J.S.: A System of Logic. University Press of the Pacific, Honolulu (2002)
6. Finn, V.K.: Plausible inferences and plausible reasoning.Itogi Nauki i Tekhniki. Ser. Teor. Veroyatn. Mat. Stat. Teor. Kibern. **28**, 3–84 (1988). [English version in: *Journal of Soviet Mathematics*, **56**:1 (1991), 2201–2248.]
7. Finn, V.K.: Timely notes about the JSM method for automatic hypothesis generation. Autom. Documentation Math. Linguist. **43**, 5, 257–269 (2009)
8. Anshakov, O.M.: The JSM method: a set-theoretical explanation. Autom. Documentation Math. Linguist. **46**(5), 202–220 (2012)
9. Lyfenko, N.D.: An approach to text data categorization based on the ideas of J.S. Mill. Autom. Documentation Math. Linguist. **49**(6), 202–212 (2015)
10. Hajek, P., Havranek, T.: Mechanizing Hypothesis Formation. Springer (1978)

A Survey of Current Challenges in Manufacturing Industry and Preparation for Industry 4.0

Ateeq Khan and Klaus Turowski

Abstract Evolving business needs and shorter product life cycle demand for new methods and services from production facilities. In the past few decades, research and technology has been advanced at the rapid pace on enterprise level. Technology used at production level is still quite old and has not gained much attention in comparison to enterprise level. Enterprises are interested to invest more resources on production level to overcome problems and satisfy goals of stakeholders. The paper discuss the term Industry 4.0 and why we need changes in our traditional manufacturing systems (due to new business models, companies' competition, and innovation gap). We find-out what are the current challenges faced by organizations with the help of a survey. We also list new opportunities and applications possible by introducing new tools and technologies and provide initial feedback of a scenario from case company. At the end, we discuss how to solve these challenges, scenarios and summarize our paper.

1 Introduction

In the past three decades, there are significant development took place in information technology in general. The developments have revolutionized the way we live our lives and perform business operations/processes at work. The challenge of competition posed by certain countries (China, India) forced developed countries to focus more on innovation, more value, and services. Therefore, there is a trend of growth in budget allocated for next wave of manufacturing as reported by various surveys [1, 2]. Companies are eager to introduce new technologies to improve quality, efficiency of resources, reduce risks, and to remain competitive in the market. These new developments made possible for organizations to serve the customers in new

A. Khan (✉) · K. Turowski
Magdeburg Research and Competence Cluster, University of Magdeburg, Magdeburg, Germany
e-mail: ateeq.khan@ovgu.de

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 450, DOI 10.1007/978-3-319-33609-1_2

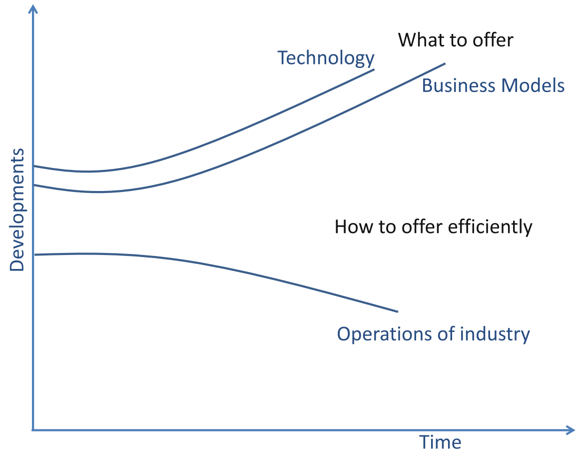


Fig. 1 Relation to develop strategy

ways and generated new business models to create value for customers and revenue for themselves. The Fig. 1 (adapted from [3]) shows the relation between developments in technology, business model, and their application in industrial operations with respect to time. As shown in the figure that there is significant difference in developments of business models, technology vs. development in industrial operations. This also justifies the need to better understand technology, business models and their applicability in organizations to achieve maximum benefits.

New business models have been introduced by various companies in order to provide more value and services to customers [4, 5]. These new business models not only attracted majority of the customers but also created new customers (and their segments) to which values/services can be provided.

A company which fails to cope the technology challenges also face the challenge of introducing new products/services, innovation, and business models. This puts the company in enormous competition where costs have to be reduced each year. A company which employs latest technology coupled with an innovative business model is guarantee for its success. There are many successful examples of big companies and start-ups. Technology has been advanced at rapid pace on enterprise level. Technology used at production level has not gained much attention in comparison to enterprise level. However, these technological developments were not much applied on industrial production level and stands as old as start of the third revolution with basic IT functionalities with monolithic structure. This was due to critical nature of the systems and their potential economic impact on organization operation. If a manufacturing system stops, the whole production line will be stopped and results loss to company. The gap between application level and shop-floor or machine level is also quite obvious. Enterprises are interested to invest more resources on production level to overcome problems.

In this paper, we discuss about industry 4.0. As there is quite potential of improvement at production level, we want to know what are the current challenges faced by companies with the help of a survey. We also provide promising future scenarios and discuss initial results of one of the future scenario in a multinational manufacturing company. At the end we provide a summary and outlook.

2 Industry 4.0 and Future Manufacturing

In order to compete with other uprising countries (like India, China) and offer more value to the customers, developed countries started to apply advance technologies on production level. In USA and some other countries, such initiatives are termed as fourth industrial revolution, Internet of Things (IoT), or next generation systems. Whereas in Germany, it is driven by German government (Bundesministerium für Bildung und Forschung) [6], and referred as Industry 4.0 (industrie 4.0) [7]. There are various definition exists for industry 4.0 by various groups and companies according to their needs and understanding. They also relate the term with other terms like IoT, Cyber Physical Systems (CPS), Smart Systems, and Digital Factory.

We define Industry 4.0 as a revolution enabled by application of advanced technologies (like IT) at production level to bring new values and services for customers and organization itself. The will also bring flexibility and quality in production systems to fulfill demands of new innovative business models and services quickly (service oriented architecture and network communication at production level). The digitalization and virtualization are tools to bring end-to-end services throughout a product life-cycle (design till recycle) and in a cost effective way for customers.

A formal definition of industry 4.0 is defined in [7] as follows:

"Industry 4.0 will involve the technical integration of CPS into manufacturing and logistics and the use of the Internet of Things and Services in industrial processes". Powerful autonomous physical systems connected with one another and environment will perform operations intelligently (smart systems). These interconnected systems referred as CPS [8, 9] communicate each other to fulfill the tasks. Cyber-Physical Production Systems comprise smart machines, warehousing systems and production facilities that have been developed digitally and feature end-to-end ICT-based integration, from inbound logistics to production, marketing, outbound logistics and service. The IoT is a network of devices. These devices can be small, e.g., sensor in a fridge or it can be a robot working inside a car manufacturing factory. The no. of IoT devices till 2020 will be around 50 billion, an estimate by Cisco [10]. Such huge amount of connected devices opens up the door for new opportunities and new use cases in every field. Industry and academics will find new use cases and services which can be offered to various industries. Although currently IoT use cases at production level are less and organizations does not know how we can take advantage from it. As stated in [7], in future manufacturing, collaboration is the focal point.

Whether such collaboration exists between CPS, companies, universities, institutions, and other partners, the granularity of such collaboration can vary from cases to cases. Cloud platform and services will play a significant role in this context, allowing systems and partners to work from anywhere, communicate and collaborate in real time cloud environment. Such collaboration exists between our institution and our partners. We perceive the requirements of next generation systems and services (like seamless integration, secure services, and smart systems) and conduct the research on industry 4.0 with the help of industrial partners. We investigate solutions for current problems and how new and innovative services/systems can be created and evaluated for next industrial revolution.

3 Research Methodology

For our research in this paper, we use one of the qualitative techniques called case research strategy. As industry 4.0 and smart manufacturing is relatively new research area, practice based problems, and poses new challenges, case research strategy is a best candidate for it as discussed in [11–13]. In this paper, we want to know and understand the stakeholder's expectations, requirements and the potential challenges industry 4.0 poses in the natural settings. Since current challenges, future expectations from industry 4.0 have been limited investigated and lack of case data in production environment from companies. We also want to discuss future scenarios from industry 4.0 perspective and challenges associated with them. We prepared short questionnaire and distributed in ICT exhibition. We also get insights by informal interviews, various company's documents, and talks with industrial experts and consultants regarding current problems and challenges in production environment.

4 Current Challenges

In this section, we provide an overview of current challenges in manufacturing industry. To find-out, what the top challenges manufacturing industry is facing, we prepare a short questionnaire and distributed it in an information technology exhibition. We also include the feedback from informal interviews, company documents, talks with industrial experts and consultants involved in production environment. These companies belong to diverse industry segments. Although due to different industry segments and complex nature of their businesses, challenges are also diverse but there are also some common challenges. From our results, we discuss top three challenges out of fifteen, in this paper due to space limitations.

4.1 Data Integration

In our data-driven world, we generate data in various ways. In production environment, data is generated and collected from different machines sensors, process data, product data, quality data, plant data, logistics data, data from partners, and infrastructure data; all contribute into explosion in data size. Usage of IoT devices increased rapidly in last decade, which also contribute largely in the amount, heterogeneity, and speed of the data generated at the production level.

Such data poses various challenges and demands new methodologies for storing, processing, and management. New algorithms, models, products, and visualizations techniques are required to use and gain the actual benefits from the data. Data engineers are required to analyze such data and to find correlation between data streams and to gain new insights from the data which were not thought earlier.

Specifically, there is a problem in which plethora of intermediate solution exists for data management within a company; it ranges from storing and exchanging data in form of printouts, emails, excel sheets, proprietary applications, and using heterogeneous database solutions between various departments or production halls. Lack of standardized approach for data management is still one of the concerns in big companies. Redundant data is stored in various departments of the company, in different data formats with minor extensions or enrichments. Such data silos raise the amount of data redundancy, inconsistency, and different interpretation of data. Software licenses, updates, hardware, and skilled personnel costs to manage such landscape heterogeneity are burden in a competitive production environment. Decision made on inconsistent data leads to incorrect decision.

Other challenge is availability of the needed data for analysis purpose. Currently reports are periodically made, pre-defined, and results are exported into other programs, e.g., in Microsoft Excel, for decision making. The significance of real time analysis of production data enable us to make decisions on current data in real time, which leads to cost reduction and improved performance.

If systems are well integrated then above-mentioned problems can be eliminated and hardware, software, and operational costs can be reduced. Such integration is although quite difficult to execute because changes in data structure and understanding of various source data models are needed. If such data have to transfer into other systems then knowledge of system landscape transformation, ETL (extract, transform, and load) methods, and expertise in management are required.

4.2 Process Flexibility

As product life cycle in this decade is shorter than before [14, 15]. Individualized and customized products also become reality. Such individualization and customization requires flexibility at production level in a cost effective manner. In order to provide such flexibility, production environment should be adaptable at the process level.

Technology, currently used at shop floor level is inadequate and does not support the process flexibility [16]. Although such adaptation are performed frequently at various levels, e.g., at process level, changes in database, programs, data types, and even in the production sequences or flow but results in higher costs. Variability management mechanisms are needed for flexibility at various levels in production as it is required in other fields [17]. Traditionally, processes and systems at production level are developed and managed isolated over the time in various departments. Change management at production level is quite challenging. As processes span over various departments, a clear process ownership is also missing in case of adaptation or changes. Change structure is also needed because sometimes it is not possible to keep the required change in the specific area and will impact the whole landscape due to dependencies. In case of changes, required changes are transferred in form of printouts or using email communication. Often these changes are handled individually in each department without any specific standards which raises the complexity and costs of managing such changes. There is a need to bring process standardization and synchronizations between various company departments to provide flexibility in an effective manner.

4.3 Security

Security is also a top concern now and it will be the major concern in future for industries. Industries want to keep their people, products, and production facilities environment secure from security risks. The trend of using smart devices in production is increasing. IoT devices will be around 50 billion till the end of this decade [10]. These devices will be used in homes, factories and everywhere. On one hand connectivity of these devices provides great advantages to ease our lives. On the other hand it poses greater risk from security perspective. Monitoring of such devices, used in production, is also a challenge from software and hardware perspective, which is often ignored. All devices whether industrial machines, computer, tablets, or smart phones needs to be updated on regular basis whether to avoid threats or due to configuration changes installed in these devices spread across the geographical location or inside factory. Keeping track of updates and management of such devices is a tedious task too. As some of IoT devices used at production level have very limited processing capabilities which requires new tools or methods, and measurements, to keep the devices secure instead of tradition methods. Serious measures are needed to restrict the threats posed by the malfunctioning or hacked devices. There are already various examples already happened where production facilities are targeted, e.g., security holes exploited in programmable logical controllers deployed in factories [18]. It is also possible that manufactured electronic products may contain viruses from production facility, when delivered in the market, which results in heavy fines for company or product returns.

5 Future Scenarios and Initial Results

Manufacturing industry has to cope with various challenges as mentioned in previous section. Despite of those challenges, in the following, we present some of the future scenarios from industry 4.0 perspectives. The scenarios also reflect the challenges mentioned in previous section.

5.1 Process Integration Within and Across Enterprise Boundaries

Product life-cycle involves series of processes, from design to production, service and feedback from customers. These processes can belong within the same enterprise or distributed across enterprise boundaries. Process integration is quite challenging in this case due to various technologies, interfaces, standards, methods and unique characteristics in each enterprise involved. Integrated processes across the enterprise will enable to optimize and make decisions in real time. Logistics can be well optimized and out of stock or over production cases, both results in revenue losses can be eliminated. Suppliers can access to live data at shop-floor level and know when to provide the required material for better resource planning and will reduce unplanned outage or overstock situations. Existing processes can be optimized and will be executed faster.

In case of companies having more than one manufacturing facilities, whether in same geographical location or scattered around the globe, cross plant manufacturing and planning makes more sense if data from facilities is available and integrated. Production load can be distributed from one plant to the other plants for optimal resource usage. Business processes can be analyzed across plants [19] to find out which plant is better and what we can learn from one plant or how we can develop best practices for specific industry or products for the whole organization. Other key performance indicators (KPIs) can also be measured, e.g., comparing employee productivity, production capacity, quality metrics, effectiveness, failure rate, and producing products better, faster and in efficient way. There is also a trend in which instead of selling end products, companies sell their know-how or other services. A company can allow other companies or partners to use state of the art manufacturing facility, competency and knowledge know-how as a service to develop their own product. In this case integrated process across enterprise boundaries is a real challenge where companies have to exchange information and applying processes at hired facility in a secure and confidential way.

5.2 Real-Time Information Access on Hand-Held Devices

Real-time data access in a production is very vital whether it is related to products, processes, or machines operating in the factory. Traditionally, real time information access for processes was not available at shop floor level. In case of change in processes or actions, workers or machines have to wait until instructions are manually transferred or data is loaded in the production system. Future factories demand a close integration between ERP and shop-floor and real time access of data at production level for real time execution. Data collected from machines and business processes is filtered, analyzed, and then delivered in required format to provide insights which in return will help to give better process control, optimize, and reduce overhead costs. Customized worker plans can be generated in real time for day to day operations to reduce the possibility of errors and faults. Real-time access of needed information on mobile devices fosters the development of applications using new front-end technologies, e.g., HTML5, SAP Fiori from in-memory database to give personalized experience and enable efficient business interactions. In our case company, production process and product data is delivered on handheld devices to employees working at the shop-floor level at right time. Due to instant availability of data, resources whether machines or human perform their tasks faster based on current data; hence Feedback from the employees was very positive and overall process was also optimized. Some employees complaint about the size and weight of the device. For some jobs handling an extra device was not comfortable. One reason was that they were not used to handle such devices in their day to day operations. Displaying such information on handheld devices in real-time will enable to enhance the performance and reduce the operational costs. Companies can also compare the production data within plants, e.g., by peer to peer comparison of machines or production facilities. Machine states and processes follow to accomplish tasks; all generates data and logs which can be used to find the track the steps performed in a plant. Such data can be compared with data from other plants to learn how to improve or optimal way.

Similarly, in case of faults in machines at production level, e.g., machine faults can be directly reported at ERP level and necessary measures can be initiated immediately saving time needed for production. The production environment display always current KPIs based on real time data.

5.3 Predictive Maintenance

Maintenance of machines is an important area which every organization has to address. Organizations try to carry out planned maintenance based on different strategies like operating hours; number of products processed, or after a certain time. A machine condition monitoring system can be introduced to avoid unplanned maintenance. Machines equipped with sensors generates huge amount of data. Such data, e.g., containing machine temperature, vibrations, speed, pressure, state, and other values, records the operating condition in which machine operates. Historical data collected regarding machines operating conditions can play a vital role. Current state of the machine is compared with historical data and with other data in different dimensions (product quality, and wastage data). Models can be developed to predict which part of machine or machine is going to fail or vulnerable in a production environment. Prompt actions can be taken in case of vulnerability to avoid breakdowns. Such actions will increase the reliability of a machine. Predictive maintenance can be performed by seeing when machine is going to fail or which parts should be replaced before machine actually fail.

Machines or hardware manufacturers can collect data from machines to provide remote diagnostics and offer maintenance services from their locations. Such data can also be useful for them to know in which conditions their machines are operating and what they can learn from such data. Remote setting of parameter or operating conditions or providing early warning in case of machine is over used or wrongly used as compared to what it is made for. They can also send their maintenance staff to repair or diagnose the problem. Such data can be collected by the machines and transmitted to the machine manufacturer. Other option is such data is collected by the production facility and then those enterprises can collaborate to produce or offer better services.

6 Discussion and Suggestions

In previous sections, we list challenges faced by many organizations in their production environment found from our survey. We also provide some motivating future scenarios and initial findings of one of the scenario. Given this state of affairs, there is no silver bullet to solve those challenges at once; solutions are needed to cope with them. We already discussed that poor data management is also the reason for revenue losses as it is also reported in an other survey [20]. In future, database sources heterogeneity will be increased and spread across the globe. Tools and techniques are needed to have a scalable solution which incorporates the requirements within and across enterprise boundaries and to suppliers and other manufacturers. Industry standards should be needed for seamless integration of such heterogeneity.

New tools and platforms are needed to integrate data at production and enterprise level. It will remove the information silos and bring integrated solutions. A holistic security approach is needed to develop a trusted environment for data exchange between partners. Future platforms for collaboration are needed where stakeholder can share and exchange data easily ensuring data security and confidentially of share data. Industry 4.0 demands to introduce new technology to remain competitive in future and requires a complete roadmap and long term investments. A smooth-less transition plays as important role for migration. A comprehensive approach is needed for seamless transition in such projects. During this transition, it is important that routine processes of factory should continue as before. Planned downtime is less risky, and costs less as compared to unplanned downtime. Although it is quite difficult to estimate the loss due to downtime as it also depends on industry or factors involved, e.g., production loss, material loss, low quality of product manufactured, unproductive resources, and costs involved to overcome all these issues. The overall cost reduces significantly if it is know which part is going to fail and should be replaced. If production is effected or stopped due to unplanned failure, costs are very high, e.g., averaging around 22,000 dollars/minute as mentioned in a survey [21].

We believe that a step-wise approach is needed, first to solve current challenges and to improve current situation. By doing so, it enables production environment to gain quick benefits and make production ready for future scenarios or to gain advantage from industry 4.0. Such step-wise or modular approach can come from service-oriented domain depending on the individual nature of the problem, e.g., for data exchange problem, introducing service bus for data exchange or using wrapper patterns [22] for departments to provide necessary services.

To enable industry 4.0 scenarios firstly concrete problems should be eliminated. As we discussed only few challenges and future scenarios, obviously there are also other future scenarios, e.g., additive manufacturing, cloud manufacturing, self-organizing manufacturing, which will bring also new challenges. Future scenarios mentioned above and others will also bring new aspects of different dimension which should be studied carefully, when implementing or running such project, to keep the organization competitive.

There are various kinds of risks involved in running industry 4.0 projects. One of the hindrances in such projects is related to people. Taking people in confidence and motivating them to participate in Industry 4.0 projects is very important, otherwise they will not allow projects to run successfully. Changing mind-set of people working in the organization is also challenge for industry 4.0 projects. People resist changing the ways they used to follow. So, a clear directive is needed from higher management for activities like data integration or related to industry 4.0 projects. It is also important to increase safety, security, and awareness about industry 4.0 in employees, by training the workforce, as they play a key role to enable new scenarios. Hiring of skilled staff and then keeping the know-how or skilled person with the organization is quite challenging.

7 Summary and Outlook

In our work, we show the importance of industry 4.0 and how it will enable manufacturing industry to improve and optimize processes. Research is focusing to find out what we are new opportunities and other scenarios which are not thought before by collecting data from machine level and integrating systems across the boundaries. In our study, we investigate what are the current challenges industry is facing which should be solved. We also discuss the new scenarios which are possible in manufacturing industry to gain the benefit from industry 4.0. We also give hints briefly how to solve those challenges and how we can enable new scenarios. Although there is no silver bullet which will solve all problems in industry but still frameworks or step by step approaches are needed to realise scenarios in near future. These new scenarios may pose new challenges which should be considered when realizing such scenarios.

References

1. Falk, M., Klien, M., Schwarz, G.: Large manufacturing firms plan to increase their investments in 2015. results of the wifo spring 2015 investment survey. WIFO Monatsberichte (monthly reports) **88**(7), 581–591 (2015)
2. Tassey, G.: Competing in advanced manufacturing: the need for improved growth models and policies. J. Econ. Perspect. 27–48 (2014)
3. London, T., Hart, S.L.: Next generation business strategies for the base of the pyramid: New approaches for building mutual value. Pearson Education India (2010)
4. Figalist, D.H.: On the way to industrie 4.0. driving the digital enterprise. VDMA Informationstag, Industrie 4.0 Konkret (Sep 2015), www.industrie40.vdma.org
5. Kallenbach, D.R.: The internet of things: Connected products, connected manufacturing. internet (May 2015), www.bosch.com, bosch Software Innovations GmbH
6. MacDougall, W.: Industrie 4.0 smart manufacturing for the future. Germany Trade and Invest (Jul 2014). http://www.gtai.de/GTAI/Content/EN/Invest/_SharedDocs/Downloads/GTAI/Brochures/Industries/industrie4.0-smart-manufacturing-for-the-future-en.pdf
7. Kagermann, H., Helbig, J., Hellinger, A., Wahlster, W.: Recommendations for implementing the strategic initiative industrie 4.0: Securing the future of german manufacturing industry; final report of the industrie 4.0 working group (2013)
8. Lee, E., et al.: Cyber physical systems: design challenges. In: 2008 11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing (ISORC), pp. 363–369. IEEE (2008)
9. Tseng, M.M., Hu, S.J.: Mass customization. In: CIRP Encyclopedia of Production Engineering, pp. 836–843. Springer (2014)
10. Cisco: Internet of things (iot). http://www.cisco.com/web/solutions/trends/iot/portfolio.html (Sep 2015)
11. Benbasat, I., Goldstein, D.K., Mead, M.: The case research strategy in studies of information systems. MIS Q. 369–386 (1987)
12. Eisenhardt, K.M.: Building theories from case study research. Acad Manage. Rev. **14**(4), 532–550 (1989)
13. Eisenhardt, K.M., Graebner, M.E.: Theory building from cases: opportunities and challenges. Acad. Manage. J. **50**(1), 25–32 (2007)
14. Hofmann, J., Bick, P.D.W.: Are variant producers the pioneers? ROI Dialog **43**, 6–7 (2015)
15. Hofreiter, B., Huemer, C.: Flexible workflow management in service oriented environments. In: Agent-Based Service-Oriented Computing, pp. 81–111. Springer (2010)
16. Zhang, Y., Huang, G., Qu, T., Sun, S.: Real-time work-in-progress management for ubiquitous manufacturing environment. In: Li, W., Mehnen, J. (eds.) Cloud Manufacturing, pp. 193–216. Springer Series in Advanced Manufacturing, Springer, London (2013). http://dx.doi.org/10.1007/978-1-4471-4935-4_9
17. Khan, A., Kästner, C., Köppen, V., Saake, G.: The pervasive nature of variability in SOC. In: 9th International Conference on Frontiers of Information Technology (FIT'11). Islamabad, Pakistan, pp. 69–74. IEEE Computer Society, Los Alamitos, CA, USA (Dec 2011)
18. Zetter, K.: Serious security holes found in siemens control systems targeted by stuxnet. http://arstechnica.com/security/2011/08/serious-security-holes-found-in-siemens-control-systems-targeted-by-stuxnet/ (Apr 2011)

19. Lodhi, A., Köppen, V., Wind, S., Saake, G., Turowski, K.: Business process modeling language for performance evaluation. In: 47th Hawaii International Conference on System Sciences, pp. 3768–3777 (Jan 2014)
20. Vespi, C.Y.: From overload to impact: An industry scorecard on big data business challenges. http://www.oracle.com/us/industries/oracle-industries-scorecard-1692968.pdf (Jul 2012)
21. Research, N.: Stopped production costs 22,000 dollars per minute. http://advancedtech.info/news-releases/04/21/2006/new-survey-of-auto-industry-manufacturing-executives-shows-stopped-production-costs-an-average-22-000-per-minute-en-us (Mar 2006), advanced Technology Services, Inc., Accessed 02 Feb 2016
22. Khan, A., Kästner, C., Köppen, V., Saake, G.: Service variability patterns. In: De Troyer, O., Bauzer Medeiros, C., Billen, R., Hallot, P., Simitsis, A., Van Mingroot, H. (eds.) Advances in Conceptual Modeling. Recent Developments and New Directions, ER 2011 Workshops. Lecture Notes in Computer Science, vol. 6999, pp. 130–140. Springer (Nov 2011)

A Comparison of Stochastic and Fuzzy Orderings

Alexander Lepskiy

Abstract There are many applications where the comparison of histograms (discrete random variables, fuzzy set on discrete universal set) is required with the help of relationship of type "more-less". There are many approaches to solving this problem. The relations between some popular stochastic and fuzzy orderings are investigated in the article. The simple formulas for calculating the number of comparisons obtained, as well as established relationships between the various comparisons. The new approach for comparison of histogram is proposed in this paper too. This approach is based on the calculation of minimum directional transform of one histogram in another histogram.

Keywords Stochastic ordering · Fuzzy ordering · Minimal transform

1 Introduction

There are many applications where the comparison of random variables (probability distributions) is required with respect of relationship of type "more-less". A similar problem is also relevant for the comparison of fuzzy numbers and fuzzy random variables. In particular, a comparison of random (fuzzy, fuzzy random) variables used in the theory of reliability and risk [1, 2], in biomedical research [3], in economics [4, 5], in the models of social welfare [6, 7], in the simulation of fuzzy preferences [8], in decision making under uncertainty [9, 10] etc.

There are several approaches to the construction of the order relation on the set of random (fuzzy, fuzzy random) variables. The studies of different methods of ordering

The study was implemented in the framework of The Basic Research Program of the Higher School of Economics. This work was supported by the grant 14-07-00189 of RFBR (Russian Foundation for Basic Research). The author thanks Dr. Timur Lepsky for his helpful comments.

A. Lepskiy (✉)
Higher School of Economics, 20 Myasnitskaya Ulitsa, Moscow 101000, Russia
e-mail: alex.lepskiy@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 450, DOI 10.1007/978-3-319-33609-1_3

random variables (at the time of the 2007) are summarized in the monograph [11]. The some methods of ordering of fuzzy numbers analyzed in [10]. The problem of ordering of fuzzy random variables considered in [12, 13].

Theoretical studies of stochastic (fuzzy) orderings include an analysis of the relationships between the different orders, the study of orders on the certain classes of random (fuzzy, fuzzy random) variables, properties of these relations [11], the study on the stability to distortions [14] etc. The relationships between the some orderings of random (fuzzy) variables will be analyzed in Sect. 2.

The interpretability of the used comparison relation is one of the major issues in the choice of a particular method of ordering in applied problems. The method of ordering is determined naturally when the problem formulated in some applications. For example, the natural order is given by a hazard rate order in the theory of risk and reliability [2], by a mean residual life order in the insurance theory [11]. The choice of a particular method of ordering is not obvious in many other applications.

The new method for comparing of discrete random (fuzzy) variables will be introduced in Sect. 3. This method is based on the calculation of the minimum transformation which must be performed above one histogram (variable) to convert it into another histogram. Such a method has a good interpretability in solving a number of problems, as will be shown. In general, this method can be described as a directional version of the procedure of calculating the EMD-metric [15].

2 Stochastic and Fuzzy Orderings

We shall consider only discrete random or fuzzy variables in this article. Discrete random variables and fuzzy set on a discrete universal set can be described using histograms. The histogram has a view $U = (x_i, u_i)_{i\in I}$, where $(x_i)_{i\in I}$ is an ordered ascending vector different arguments of histogram (i.e. $x_i < x_{i+1}$, $i \in I$), $(u_i)_{i\in I}$ is a vector of non-negative values of histogram, I is a some index set. We have $u_i = P\{U = x_i\}$, $i \in I$ in case of discrete random variable (which is also denoted by U). If U be a fuzzy variable then $u_i = \mu_U(x_i)$, $i \in I$, where μ_U is a membership function on the universal set $(x_i)_{i\in I}$.

We want to set up some relation of order R on the set of histograms $\mathcal{U} = \{U\}$. The specific type of relationship is determined by features of the problem. For example, in the case of ranking a set of objects on histograms, in which a higher value corresponds to a higher rank, it should be the relation of total preorder (reflexive, complete and transitive relation). If histograms U and V are in the relation R (i.e. $(U, V) \in R$)), then we will denote this through $U \succcurlyeq V$ and we will speak that U is greater than V. If $U \succcurlyeq V$ and $V \succcurlyeq U$ then we will treat these histograms as equal and we will denote by $U \sim V$.

We will also assume that the relation R should be in accordance with the condition, that ordering arguments of the histogram ascend their importance: if $U' = (x_i, u'_i)$, $U'' = (x_i, u''_i)$ are two histograms for which $u'_i = u''_i$ for all $i \neq k, l$ and $u'_l - u''_l = u''_k - u'_k \geq 0$, then $U'' \succcurlyeq U'$ for $k > l$ and $U' \succcurlyeq U''$ for $k < l$.

2.1 Stochastic Orderings

We will present few examples of pairwise comparison of histograms-probability distributions of discrete random variables $U = (x_i, u_i)_{i\in I}$ and $V = (y_j, v_j)_{j\in J}$.

A comparison on the mathematical expectation: $U \succcurlyeq_{me} V$, if $E[U] \geq E[V]$.

In general $U \succcurlyeq_{mef} V$, if $E[f(U)] \geq E[f(V)]$, where f is some function (utility function).

Stochastic dominance: $U \succcurlyeq_{sd} V$, if $F_U(t) \leq F_V(t)$ for all $t \in \mathbb{R}$,
where $F_U(t) = \sum_{i:x_i<t} u_i$ be a distribution function of random variable U. This is one of the most popular methods of stochastic ordering. But this ordering may be determined not on all set of $\mathcal{U}^2$, where $\mathcal{U} = \{U\}$ be a some set of histograms. The generalization of this comparison on the set of so-called probability boxes (p-boxes) [16] was considered in [17]. The probability box is the set of distribution functions that bounded by the two distribution functions $\underline{F} \leq \overline{F}$. Such constructs are investigated in the framework of imprecise probability theory and are widely used, for example, in the theory of risk.

Because the $E[U] - E[V] = \int_{-\infty}^{\infty} \left(F_V(t) - F_U(t)\right) dt$ is true, then the comparison of the average value follows from the stochastic dominance: $U \succcurlyeq_{sd} V \Rightarrow U \succcurlyeq_{me} V$.

Stochastic precedence (see, e.g. [18]):

$$U \succcurlyeq_{sp} V, \quad \text{if} \quad P\{U \geq V\} \geq P\{U \leq V\} \Leftrightarrow P\{U \geq V\} \geq \frac{1}{2}\left(1 + P\{U = V\}\right).$$

If random variables $U = (x_i, u_i)_{i\in I}$ and $V = (y_j, v_j)_{j\in J}$ are independent then $P\{U \geq V\} = \sum_{(i,j):\, x_i \geq y_j} u_i v_j = \int_{-\infty}^{\infty} F_V(t) dF_U(t)$. In particular, the last formula implies that $U \succcurlyeq_{sd} V \Rightarrow U \succcurlyeq_{sp} V$ in the case of independent continues random variables because $P\{U \geq V\} = \int_{-\infty}^{\infty} F_V(t) dF_U(t) \geq \int_{-\infty}^{\infty} F_U(t) dF_U(t) = \frac{1}{2}$ if $U \succcurlyeq_{sd} V$.

In general the stochastic precedence is not transitive relation. The some generalization of notion of t-transitivity (with triangular norm t) considered in [19]. The relation of stochastic precedence will be transitive relation with respect to this new type of transitivity (so called cycle-transitivity). Some basic directions of applications of stochastic precedence in problems of statistical analysis are indicated in [20]. The relationship between these and other stochastic orders investigated in several papers (see [11, 18]).

Theorem 1 *The following relations hold:*

(a) $U \succcurlyeq_{sd} V \Rightarrow U \succcurlyeq_{me} V$;
(b) $U \succcurlyeq_{sd} V \Rightarrow U \succcurlyeq_{sp} V$, *if* U, V *are independent random variables.*

Note that the relation of the stochastic precedence $\succcurlyeq_{sp}$ depends on the joint distribution of two random variables as opposed to the relationship established by comparing the mathematical expectations $\succcurlyeq_{me}$. Therefore, the stochastic precedence does not follow from stochastic dominance in the general case of dependent variables. For example [20], if we have two normal random variables $U \sim N(m_U, \sigma_U^2)$ and $V \sim N(m_V, \sigma_V^2)$ then $U \succcurlyeq_{sd} V \Leftrightarrow m_U \geq m_V$ and $\sigma_U^2 = \sigma_V^2$, but $U \succcurlyeq_{sp} V \Leftrightarrow m_U \geq m_V$.

In addition, stochastic dominance between the variables of $U = (x_i, u_i)_{i \in I}$ and $U = (x_i, u_i)_{i \in I}$ depends on the ordinal relation between arguments $(x_i)_{i \in I}$ and $(y_j)_{j \in J}$, as well as from the ordinal relation between the cumulative probabilities $\left(\sum_{i=1}^k u_i\right)_k$ and $\left(\sum_{j=1}^k v_j\right)_k$. Stochastic precedence depends on the ordinal relations between arguments and the absolute values of probabilities. But the comparison on mathematical expectation depends on the absolute values of the arguments and probabilities.

2.2 *Fuzzy Orderings*

The fuzzy set U with membership function $u_i = \mu_U(x_i)$ defined on the universal set $X = (x_i)_{I \in I} \subseteq \mathbb{R}$ can be associated with each histogram $U = (x_i, u_i)_{i \in I}$ in this case. Sometimes we will assume that the fuzzy set U which defined on the $\mathbb{R}$ with a step membership function μ_U, corresponds to the histogram $U = (x_i, u_i)_{i \in I}$, where $\mu_U(t) = u_i$ for $x_i < t \leq x_{i+1}$, $i \in I$ and $\mu_U(t) = 0$ for $t \notin (\min x_i, \max x_i]$. The mathematical formalism of comparison of so-called fuzzy numbers [10] developed to a greater extent in the theory of fuzzy sets. Fuzzy numbers are a generalization of conventional (crisp) real numbers. In our case the histogram-fuzzy set $U = (u_i)_{i \in I}$ will be a fuzzy number, if all so called α-cuts $U_\alpha = \{t : \ \mu_U(t) \geq \alpha\}$ are non empty convex closed sets for all $\alpha \in (0, 1]$ and $\max_{i \in I} u_i = 1$. Many ways of comparing fuzzy numbers have appeared in recent decades (see references and review in [10]). We note that some methods of ranking fuzzy sets (for example, the centroid method, see below) are not sensitive to the requirement of convexity α-cuts, but the other (mainly those which use α-cuts) are sensitive. This requirement will certainly be satisfied if the histogram is unimodal. Below we mention some of the main approaches to the comparison of fuzzy numbers.

The comparison methods with the help of the ranking function (index). A certain function (index, defuzzification operation) $F : \mathcal{U} \to \mathbb{R}$ is determined in this case and we believe that $U \succcurlyeq V$ $(U \sim V)$, if $F(U) \geq F(V)$ $(F(U) = F(V))$. Some functions of segment boundaries of α-cuts $U_\alpha = [u_1(\alpha), u_2(\alpha)]$ of fuzzy number are considered often as indices ranking. The following functions are examples of indices ranking:

- Adamo index [21]: $A_\alpha(U) = u_2(\alpha)$, where fixed number $\alpha \in (0, 1]$ characterized the measure of risk of wrong decisions;
- generalized Yager index [22]: $Y_\lambda(U) = \frac{1}{2} \int_0^1 (\lambda u_1(\alpha) + (1 - \lambda) u_2(\alpha)) d\alpha$, where coefficient $\lambda \in [0, 1]$ characterized the level of optimism of decision maker;
- centroid index $C(U) = \sum_i x_i u_i \Big/ \sum_i u_i$. If U is a random value then $C(U) = E[U]$.

Let U be a fuzzy number and $F_U(z) = \int_{-\infty}^z \mu_U(t) dt$ (analog of distribution function), $|U| = \int_{\mathbb{R}} \mu_U(t) dt$ be a cardinality of fuzzy number, $a_U = \inf\{t : \ \mu_U(t) = 1\}$

and $b_U = \sup\{t : \ \mu_U(t) = 1\}$ are borders of kernel of fuzzy number. The following theorem is true for generalized Yager index.

Theorem 2 *If U is a fuzzy number and $|U| < \infty$, then*

$$Y_\lambda(U) = a_U + \lambda\,|U| - F_U(a_U).$$

The last formula implies that ranging with respect to generalized Yager index of fuzzy numbers of equal cardinality does not depend on parameter but it depends only from left or right sides of distribution functions: if $|U| = |V|$, then $Y_\lambda(U) \geq Y_\lambda(V) \Leftrightarrow a_U - F_U(a_U) \geq a_V - F_V(a_V)$. The following example shows us that the dominance with respect to Jager index (i.e. $U \succcurlyeq_{Y_\lambda} V$, if $Y_\lambda(U) \geq Y_\lambda(V)$) in general does not follow from stochastic dominance.

Example 1 Let $V = (0, 2, 2)$ be a triangular fuzzy number, $|V| = 1$; U is a fuzzy number with membership function $\mu_U(t) = \begin{cases} \frac{18t-5}{36}, \ \frac{5}{18} \leq t < 1, \\ 1 + 2b - 2bt, \ 1 \leq t \leq c, \\ 0, \ t \notin \left[\frac{5}{18}, c\right], \end{cases}$ where $b = \frac{169}{620}$, the value $c > 1$ determined from the conditions $|U| = 1$ and $\mu_U(t) \geq 0 \ \forall t$: $c \approx 2.417$. Then $a_U = 1$, $a_V = 2$ and $F_U(t) \leq F_V(t) \ \forall t$, i.e. $U \succcurlyeq_{sd} V$. On the other hand we have $Y_\lambda(U) = a_U + \lambda\,|U| - F_U(a_U) \approx \lambda + 0.87$, $Y_\lambda(V) = a_V + \lambda\,|V| - F_V(a_V) = \lambda + 1$. Thus $Y_\lambda(U) \leq Y_\lambda(V) \ \forall\lambda$, i.e. $V \succcurlyeq_{Y_\lambda} U$.

However, the following statement is true.

Corollary 1 *If U and V are fuzzy numbers with equal finite cardinality $|U| = |V| < \infty$ and $a_U \geq a_V$, then $U \succcurlyeq_{sd} V \Rightarrow U \succcurlyeq_{Y_\lambda} V$ for every $\lambda \in [0, 1]$.*

The comparison methods based on the calculation of distance to the reference fuzzy number U_0. Kerre index is an example of this approach [23]: $K(U) = d_1(U, U_0)$, where fuzzy number $U_0 = \max\{U : U \in \mathcal{U}\}$ (it is calculated using the Zadeh's extension principle [10]), d_1 is a Hamming distance. Then $U \succcurlyeq V$, if $K(U) \leq K(V)$.

We can consider the pair Kerre index of two fuzzy numbers U and V: $K(U, V) = d_1(U, \max\{U, V\})$. Then $U \succcurlyeq_K V$, if $K(U, V) \leq K(V, U)$. The following theorem about comparison with the help of a pair Kerre index is true.

Theorem 3 *If U and V are fuzzy numbers with equal finite cardinality $|U| = |V| < \infty$, then*

$$K(V, U) - K(U, V) = 2\left(F_V(c) - F_U(c)\right),$$

where

$$c = \begin{cases} \underset{[\min\{b_U, b_V\}, \max\{a_U, a_V\}]}{\arg\min} \left(\max\{\mu_U(t), \mu_V(t)\}\right), \text{if } \min\{b_U, b_V\} \leq \max\{a_U, a_V\}, \\ \max\{a_U, a_V\}, \text{otherwise}. \end{cases}$$

Corollary 2 *We have under the same conditions:*

(a) $U \succcurlyeq_{sd} V \Rightarrow U \succcurlyeq_K V$;
(b) $U \succcurlyeq_{Y_\lambda} V \Rightarrow U \succcurlyeq_K V$ *for every* $\lambda \in [0, 1]$.

Notice that the comparison of fuzzy numbers with equal cardinality with the help of Yager index and Kerre index has a point character for distribution functions (see Theorems 2 and 3). But the comparison of fuzzy numbers with the help of mathematical expectation has an integral character: $E[U] - E[V] = \int_{-\infty}^{\infty} \left(F_V(t) - F_U(t)\right) dt$. Therefore the comparison with the help of Yager index and Kerre index does not follow from comparison with the help of mathematical expectation and vice versa.

The methods based on the calculation of the pairwise comparison index. Baas–Kwakernaak index is an example of this approach [9]: $BK(U, V) = \sup_{x_i \ge x_j} \min \left\{u_i, v_j\right\}$. This index is a "fuzzy" analogue of calculating the probability $P\{U \ge V\}$ for independent random variables U and V, if operations of addition and multiplication replaced on the min and sup respectively.

This index calculated simply without solution of optimization problem with the help of following property. If there are two numbers $u' \in \mathrm{Ker}\, U$ and $v' \in \mathrm{Ker}\, V$, $u' > v'$, then $BK(U, V) = 1$ and $BK(V, U) = hgt(U \cap V)$, where $\mathrm{Ker}\, U = \{x_i \,:\, u_i = 1\}$ is a kernel of fuzzy number, $hgt(U) = \sup_i u_i$ is a height of fuzzy set. Let $U \succcurlyeq_{BK} V$, if $BK(U, V) \ge BK(V, U)$. The relation $\succcurlyeq_{BK}$ is transitive in contrast to the "similar" stochastic precedence. The relation $\succcurlyeq_{BK}$ does not follow from $\succcurlyeq_{sp}$ and vice versa.

The pair Kerre index $K(U, V) = d_1(U, \max\{U, V\})$ is another example of a pair comparison index which determines the relation $U \succcurlyeq_K V$ if $K(U, V) \le K(V, U)$. The following example shows us that the relation $\succcurlyeq_K$ does not transitive.

Example 2 Let $U = (-\delta, 1, 1)$ and $W = (0, 0, 1 + \delta)$ are two triangular fuzzy numbers which depend from the parameter $\delta > 0$, and $V = (0, 0, 1, 1)$ is a rectangle fuzzy number. Then

$$\mu_{\max\{U,V\}}(t) = \begin{cases} \frac{t+\delta}{1+\delta}, & t \in [0, 1], \\ 0, & t \notin [0, 1], \end{cases} \qquad \mu_{\max\{V,W\}}(t) = \begin{cases} 1, & t \in [0, 1], \\ \frac{1+\delta-t}{1+\delta}, & t \in (1, 1 + \delta], \\ 0, & t \notin [0, 1 + \delta], \end{cases}$$

$$\mu_{\max\{W,U\}}(t) = \begin{cases} \frac{t+\delta}{1+\delta}, & t \in [0, 1], \\ \frac{1+\delta-t}{1+\delta}, & t \in (1, 1 + \delta], \\ 0, & t \notin [0, 1 + \delta]. \end{cases}$$

We have $K(U, V) = d_1(U, \max\{U, V\}) = \frac{\delta^2}{2(1+\delta)}$, $K(V, U) = d_1(V, \max\{U, V\}) = \frac{1}{2(1+\delta)}$. Therefore $U \succcurlyeq_K V$ and $U \not\sim_K V$, if $K(U, V) < K(V, U)$, i.e. $0 < \delta < 1$. By analogy we have for a pair of fuzzy numbers $V$and W that $K(V, W) = d_1(V, \max\{W, V\})$ $= \frac{\delta^2}{2(1+\delta)}$, $K(W, V) = d_1(W, \max\{W, V\}) = \frac{1}{2(1+\delta)}$. Therefore $V \succcurlyeq_K W$ and $V \not\sim_K W$, if

$K(V, W) < K(W, V)$, i.e. $0 < \delta < 1$. Finally we have for a pair of fuzzy numbers U and W that $K(U, W) = d_1(U, \max\{U, W\}) = \frac{\delta^2}{1+\delta}$, $K(W, U) = d(W, \max\{U, W\}) = \frac{1}{2(1+\delta)}$. Therefore $W \succcurlyeq_K U$ and $W \nsim_K U$, if $K(W, U) < K(U, W)$, i.e. $\delta > \frac{\sqrt{2}}{2}$. Thus we have that the triplet U, V and W does not transitive with respect to relation $\succcurlyeq_K$, if $\frac{\sqrt{2}}{2} < \delta < 1$.

Each of these and many other comparing methods are not free from disadvantages. For example, Kerre index is weakly sensitive when comparing the large number of fuzzy numbers, but Baas–Kwakernaak index is insensitive when comparing two fuzzy numbers with intersecting kernels. The unimodality of histograms required for Adamo and Yager indices. In addition, the choice of the method of comparison of fuzzy numbers is often difficult to interpret in a specific application in terms of this application. Therefore, the problem of development of a comparison method which would be interpreted well in terms of the solutions of a certain class of tasks is an actual problem.

3 A Comparison of Histograms by the Method of the Minimum Transformations

The calculation of the minimum "directional" transformation, that maps the one histogram on the other histogram, is the basic idea of the proposed approach for comparison of histograms. The problem of determining the distance between histograms with the help of a minimum transformation is a well known "similar" problem. This problem related to the solution of the transportation problem, which was formulated in 1781 by G. Monge and (in general) in 1942 Kantorovich [24] (Monge–Kantorovich transportation problem). Solution of the transport problem has led to the notion of Kantorovich–Rubinstein metric (transportation metric). This metric been rediscovered many times with certain variations: Wasserstein metric, Mallows metric, EMD metric (Earth Movers Distance metric) [15] etc.

If $U = (x_i, u_i)_{i\in I}$, $V = (y_j, v_j)_{j\in J}$ are discrete random variables, then Mallows metric is determined as follows. Let h_{ij} be a flow between the "points" x_i and y_j, that satisfy the conditions (the plan of transportation):

$$h_{ij} \geq 0, \quad \sum_i h_{ij} = v_j \ \forall j, \quad \sum_j h_{ij} = u_i \ \ \forall i \tag{1}$$

and $d_{ij} = \left|x_i - y_j\right|$ be a distance between the points x_i and y_j. Then Mallows metric is equal

$$d(U, V) = \inf_{(h_{ij})\in\mathcal{H}} \sum_{i,j} h_{ij} d_{ij}. \tag{2}$$

It is known (see, e.g. [25]), that Mallows metric of two random variables U and V with distribution functions F_U and F_V respectively is calculated using the simple formula

$$d(U,V)=\int_0^1 \left|F_U^{-1}(x)-F_V^{-1}(x)\right|dx=\int_{-\infty}^{\infty}\left|F_U(t)-F_V(t)\right|dt. \tag{3}$$

The distance $d(U,V)$ satisfies all axioms of a metric as follows from the above formula. If $U=(x_i,u_i)_{i\in I}$, $V=(y_j,v_j)_{j\in J}$ are discrete fuzzy variables, then the following transportation plan will be instead of plan (1)

$$\tilde{\mathcal{H}}:\qquad h_{ij}\ge 0,\quad \sum_i h_{ij}\le v_j\ \forall j,\ \ \sum_j h_{ij}\le u_i\ \ \forall i,$$
$$\sum_i\sum_j h_{ij}=\min\left\{\sum_i u_i,\sum_j v_j\right\} \tag{4}$$

and EMD metric is equal

$$d(U,V)=\sum_{i,j}h_{ij}^{opt}d_{ij}\Big/\sum_{i,j}h_{ij}^{opt},$$

where $H^{opt}=(h_{ij}^{opt})$ is an optimal plan of transportation (the solution of problem (2)–(4)).

We can consider different variants of comparison of histograms on base of calculation of their minimum transformation. We give some examples.

(a) Let

$$r_1(U,V)=\sum_{i,j}h_{ij}^{opt}(x_i-y_j). \tag{5}$$

Then $U\succcurlyeq_{tr1}V$, if $r_1(U,V)\ge 0$. Note that we have for random variables

$$r_1(U,V)=\int_0^1\left(F_U^{-1}(x)-F_V^{-1}(x)\right)dx=E[U]-E[V]=\int_0^1\left(F_V(x)-F_U(x)\right)dx.$$

By other words this comparison for random variables is equivalent to comparison of mathematical expectation: $U\succcurlyeq_{tr1}V\Leftrightarrow\succcurlyeq_{me}V$.

(b) Let

$$r_2(U,V)=\sum_{i,j}h_{ij}^{opt}(x_i-y_j)_+, \tag{6}$$

where $(t)_+=\frac{1}{2}(t+|t|)$. Then $U\succcurlyeq_{tr2}V$, if $r_2(U,V)\ge r_2(V,U)$. We have $r_2(U,V)+r_2(V,U)=d(U,V)$ and $U\succcurlyeq_{tr2}V\Leftrightarrow r_2(V,U)\ge\frac{1}{2}d(U,V)$. Moreover we have

$r_2(U,V) = \int_0^1 \left(F_U^{-1}(x) - F_V^{-1}(x)\right)_+ dx = \int_0^1 \left(F_V(x) - F_U(x)\right)_+ dx$ and $r_1(U,V) = r_2(U,V) - r_2(V,U)$. Therefore $U \succcurlyeq_{tr2} V \Leftrightarrow U \succcurlyeq_{tr1} V \Leftrightarrow U \succcurlyeq_{me} V$.

(c) Let

$$r_3(U,V) = \inf_{(h_{ij})\in\mathcal{H}} \sum_{i,j} h_{ij}\varphi(x_i - y_j), \tag{7}$$

where nondecreasing function φ satisfies the condition $\varphi(t) = 0$ for $t \le 0$. Then $U \succcurlyeq_{tr3} V$, if $r_3(U,V) \ge r_3(V,U)$. The function φ defines the restrictions on the long-distance transportation.

The comparisons (5)–(7) have very clear interpretation. The comparison $U \succcurlyeq_{tr} V$ means that the "cost of transportation" from histogram U to V is less than from V to U. For example, if we considered the problem of development of university ranking on histograms of USE (Unified State Exam) of applicants that were enrolled in the universities, then inequality comparison $U \succcurlyeq_{tr} V$ means that we have to transfer more applicants with higher scores from U university in V university than on the contrary for equalization of their histograms.

Note that the relation $\rho(U,V) = r_2(U,V)\big/d(U,V)$, if $U \ne V$ and $\rho(U,U) = \frac{1}{2}$ otherwise is a so called probabilistic relation (reciprocal relation, ipsodual relation), i.e. [0,1]-valued relation ρ satisfying $\rho(U,V) + \rho(V,U) = 1$. The stochastic precedence $P\{U > V\} + \frac{1}{2}P\{U = V\}$ is another example of probabilistic relation. These relationships are widely used for expressing the result of the pairwise comparison of a set of alternatives in decision making theory, fuzzy set theory etc.

Example 3 Let we have two random variables $x_i = \{1,4,7\}$, $u_i = \left\{\frac{1}{3}, \frac{5}{12}, \frac{1}{4}\right\}$ $y_j = \{2,4,8\}$, $v_j = \left\{\frac{5}{12}, \frac{1}{6}, \frac{5}{12}\right\}$ and $\varphi(t) = t^2$ if $t > 0$. Then we have

(a) $E[U] = \frac{15}{4}$, $E[V] = \frac{29}{6} \Rightarrow r_1(U,V) = E[U] - E[V] = -\frac{13}{12} < 0$ and $V \succcurlyeq_{tr1} U$.
(b) $r_2(U,V) = \frac{1}{6}$, $r_2(V,U) = \frac{5}{4} \Rightarrow r_2(U,V) \le r_2(V,U)$ and $V \succcurlyeq_{tr2} U$;
(c) $r_3(U,V) = \frac{1}{3}$, $r_3(V,U) = \frac{13}{4} \Rightarrow r_3(U,V) \le r_3(V,U)$ and $V \succcurlyeq_{tr3} U$.

4 Summary and Conclusion

The present research has a two aims. The study of relations between the known stochastic and fuzzy orderings was the first aim of this research. The simple formulas for calculating the number of comparisons obtained, as well as established relationships between the various comparisons. These results explain why the rankings which were received by different fuzzy methods are closer to each other than to ranking with the help of the mean value. The comparisons with the help of Yager index or Kerre index have a point character for distribution functions (see Theorems 2 and 3). But the comparison with the help of mathematical expectation has an integral character with respect to distribution functions.

The study of new approach to comparison of discrete random or fuzzy variables with the help of calculation the minimum transformation between two histograms was the other task of this paper. This comparison is the directional analog of transportation metric. The proposed comparison is calculated as a solution of linear programming problem. It has many advantages. In particular, it has a good interpretability for many applied problems. This method does not assume the unimodality of the compared histograms as a methods of comparison of fuzzy numbers. At the same time this method is equivalent to the comparison with the help of the mean value in the simplest case.

References

1. Navarro, J., Rubio, R.: Comparisons of coherent systems using stochastic precedence. Test **19**(3), 469–486 (2010)
2. Stochastic Orders in Reliability and Risk. Lecture Notes in Statistics, vol. 208. Springer, Heidelberg (2013)
3. Kottas, A.: Bayesian semiparametric modeling for stochastic precedence with applications in epidemiology and survival analysis. Lifetime Data Anal. **17**, 135–155 (2011)
4. Bawa, V.S.: Optimal rules for ordering uncertain prospects. J. Financ. Econ. **2**(1), 95–121 (1975)
5. Shorrocks, A.F.: Ranking income distributions. Economica **50**, 3–17 (1983)
6. Rothschild, M., Stiglitz, J.E.: Some further results on the measurement of inequality. J. Econ. Theory **6**, 188–204 (1973)
7. Sen, A.K.: On Economic Inequality. Oxford University Press, Oxford (1973)
8. Fodor, J., Roubens, M.: Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer Academic Publishers, Dordrecht (1994)
9. Baas, S.M., Kwakernaak, H.: Rating and ranking of multiple-aspect alternatives using fuzzy sets. Automatic **13**, 47–58 (1977)
10. Wang, X., Ruan, D., Kerre, E.E.: Mathematics of Fuzziness—Basic Issues. Springer, Heidelberg (2009)
11. Shaked, M., Shanthikumar, J.G.: Stochastic Orders. Springer, New York (2007)
12. Aiche, F., Dubois, D.: An Extension of Stochastic Dominance to Fuzzy Random Variables. Lecture Notes in Computer Science, vol. 6178, pp. 159–168. Springer, Heidelberg (2010)
13. Piriyakumar, J.E.L., Renganathan, N.: Stochastic orderings of fuzzy random variables. Inform. Manag. Scien. **12**(4), 29–40 (2001)
14. Lepskiy, A.: On the Stability of Comparing Histograms with Help of Probabilistic Methods. Procedia Computer Science, vol. 31, pp. 597–605. Elsevier, Amsterdam (2014)
15. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth movers distance as a metric for image retrieval. Int. J. Comput. Vision **40**(2), 99–121 (2000)
16. Ferson, S., Tucker, W.: Sensitivity analysis using probability bounding. Reliab. Eng. Syst. Saf. **91**(10–11), 1435–1442 (2006)
17. Montes, I., Miranda, E., Montes, S.: Stochastic Dominance with Imprecise Information. Comp. Stat. Data Anal. **71**, 868–886 (2014)
18. Boland, P.J., Singh, H., Cukic, B.: The stochastic precedence ordering with applications in sampling and testing. J. Appl. Prob. **41**, 73–82 (2004)
19. De Schuymer, B., De Meyer, H., De Baets, B., Jenei, S.: On the cycle-transitivity of the dice model. Theor. Decis. **54**(3), 261–285 (2003)
20. Arcones, M.A., Kvam, P.H., Samaniego, F.J.: Nonparametric estimation of a distribution subject to a stochastic precedence constraint. J. Am. Stat. Assoc. **97**(457), 170–182 (2002)

21. Adamo, J.M.: Fuzzy decision trees. Fuzzy Sets Syst. **4**, 207–219 (1980)
22. Yager, R.R.: A procedure for ordering fuzzy sets of the unit interval. Inf. Sci. **24**, 143161 (1981)
23. Kerre, E.: The Use of Fuzzy Set Theory in Electrocardiological Diagnostics. In: Gupta, M., Sanchez, E. (eds.) Approximate reasoning in decision-analysis, pp. 277–282. North-Holland Publishing Company, Amsterdam (1982)
24. Kantorovich, L.: On the Translocation of Masses. C.R. (Doklady). Acad. Sci. URSS (N.S.) **37**, 199–201 (1942)
25. Bickel, P.J., Freedman, D.A.: Some asymptotic theory for the bootstrap. Ann. Stat. **9**, 1196–1217 (1981)

Part II
Data Mining and Knowledge Discovery in Intelligent Information and Control Systems

Time Series Forecasting Using Combination of Exponential Models and Fuzzy Techniques

T. Afanasieva, N. Yarushkina, D. Zavarzin, G. Guskov and A. Romanov

Abstract In this paper, we develop the collective of several methods of time series modeling including exponential smoothing models. We make an assumption, that the creation of complicated methods for modeling and forecasting of time series is optimal if we know all features of subject area and we able to include them in model. To improve the quality of forecasting unknown kind time series, we propose to use the aggregate results of several models. Using a fuzzy approach allows to create models with more options.

Keywords Time series · Exponential smoothing · Fuzzy models

1 Introduction

Time series modeling is a permanently urgent problem, due to the fact that time series allow modeling a huge number of processes of all sorts in the course of time. Considering wide distribution of time series application there are a lot of methods and models for its expression and forecasting.

T. Afanasieva (✉) · N. Yarushkina · D. Zavarzin · G. Guskov · A. Romanov
Information Systems Department Ulyanovsk State Technical University,
Severny Venec 32, 432027 Ulyanovsk, Russia
e-mail: tv.afanaseva@mail.ru

N. Yarushkina
e-mail: jng@ulstu.ru

D. Zavarzin
e-mail: dzavarzin91@gmail.com

G. Guskov
e-mail: guskovgleb@gmail.com

A. Romanov
e-mail: romanov73@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 450, DOI 10.1007/978-3-319-33609-1_4

A large number of studies confirms that simple statistical methods of forecasting time series of business data provide more accurate predictions in most cases, than complex techniques. Researchers often try to develop new, more accurate methods of forecasting, but they ignore their complexity. Fildes and Makridakis [1], Makridaki, Hibon [2], Small and Wong [3] suggest that needs new methods of forecasting, which are resistant to oscillations that occur in business data.

Main goal of this paper is to combine exponential smoothing methods, fuzzy techniques and efficient aggregation model.

2 Overview of Time Series Modeling Methods

Many statistical tools, for example, regression analysis, moving average, exponential moving average and etc. are used in traditional forecasting. In competitive business environment precise forecasting is very important and affects the business strategy development. Exponential smoothing model is popular among short-term forecasting tools. However, smoothing parameters are hard to detect due to the lack of information, unknown system structure or high uncertainty. Many researchers proved fuzzy tools for time series forecasting are powerful models for such problems solution.

Song and Chissom [4] were the first who gave a definition to fuzzy time series in 1993. Fuzzy methods for time series modeling generally consist of three main stages—fuzzification, determination of fuzzy logical relations and fuzzy inference.

Our experiments shown that the using of Vovk aggregating algorithm [5] for forecasting provides more quality compared to the basic algorithms and also to the simple composition algorithms.

In the literature has been described that the quality of combination forecast is often superior to the constituent forecasts [6]. Past experience has shown that the only one best model in sample period could not be better for predicting future values.

Time series are generally encounter with a time-varying conditions. A good strategy to resolve problems is to develop some good prediction models and to combine forecasts. Several studies (e.g., those, Makridakis and Hibon [2]; Stock and Watson [7]) combine predictions for a better prediction.

We propose to use the array of exponential smoothing methods and fuzzy techniques as basic algorithms for aggregation method Vovk [5]. Set of basic algorithms is generated in accordance with the classification presented below in Sect. 2 and involves the decomposition on trend and seasonal component of time series.

We used Box—Cox transformation for stabilization of dispersion and receiving stable estimates of seasonal factors [8]. Box and Cox proposed a transformation of a time series variable y_t, $t = 1,\ldots, n$, that depends on the power parameter λ in the following way:

$$y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, \ \lambda \neq 0 \\ \log\ y_t, \ \lambda = 0 \end{cases} \quad (1)$$

It is obvious that all values *y* have to be positive. We did shift of a time series for time series at which there are negative values *y*.

$$y(\lambda) = \begin{cases} \frac{(y+c)^{\lambda}-1}{\lambda}, \ \lambda \neq 0 \\ \log\ (y_t + c), \ \lambda = 0 \end{cases} \quad (2)$$

where *c*—parameter, such that $y + c > 0$ for all *y*.

We transform time series with daily periodicity which consist of a large number of days and have the big horizon of forecasting in time series with week frequency on the following formula:

$$z_i = \sum_{t=i*7-6}^{i*7} y_t \quad (3)$$

where y_t—time series, z_i—aggregated time series with a weekly periodicity.

The horizon of forecasting is reduced by 7 times, which makes the problem of forecasting more manageable. We build a forecast for aggregated time series, followed by a simple linear interpolation of weekly forecast. In addition to classical methods, the array includes the realization of the author's methods and our proposed enhanced modifications:

1. Fuzzy method based on the integration of fuzzy model and the traditional exponential smoothing [9];
2. Approach by Reinhard Viertl, which builds a model of the time series in the form of fuzzy sets over which operates similar to the classical methods of exponential smoothing [10].
3. Fuzzy method of Direct Set Assignment (DSA) for time series prediction, combined with exponential smoothing [11].

Vovk aggregating algorithm creates a forecast using at any given time forecasts of basic algorithms [5]. To improve the quality of forecasting we offer to restrict availability for aggregation basic methods with low quality.

2.1 Exponential Smoothing

Exponential smoothing is related to adaptive methods of time series forecasting. Exponential smoothing is very important in business forecasting due to availability of an adaptive mechanism, simplicity of implementation and easy interpretation of the result. The method of automatic forecasting with using of exponential smoothing models has become very popular in business. Exponential smoothing relates to short-term forecasting methods.

2.2 Integration of Fuzzy Modeling and Exponential Smoothing

In the work [9] approach of fuzzy time series of Song and Chissom [4] is developed. As well as in the original variant there are several steps for implementation:

Step 1. The determination of a universe for time series values and the specification of segmentation fuzzy intervals. If the length of segment interval has been defined, the universe can be divided into equal parts.

Step 2. The definition of fuzzy set based on the universe and time series temporal data.

Step 3. The fuzzification of temporal data.

Step 4. The finding of logical relations and their grouping according to the current state of fuzzificated data.

Step 5. Forecasting.

Step 6. Defuzzification. Centroid method is used for obtaining of numerical results.

Huarng and others [12] proposed to use method of forecasting which aggregates global information of fuzzy relations and local information about the latest oscillations for computation of forecasting value.

$$Forecasted\_value = w_g \times Glob\_info + w_l \times Local\_info \tag{4}$$

where w_g and w_l—adaptive weight, and $w_g + w_l = 1 \;\; 0 \le w_g, w_l \le 1$.

2.3 Reinhard Viertl Fuzzy Modeling

By analogy with classical representation of time series the author represents [13] exponential smoothing for fuzzy time series as following:

$$\tilde{y}_t = \beta \cdot \tilde{y}_{t-1} \oplus (1 - \beta) \cdot \tilde{x}_t \tag{5}$$

Where—β smoothing parameter, $\tilde{y}$—smoothed series, $\tilde{x}$—fuzzy set.

The author considers fuzzy components:

$$\tilde{x}_t = \tilde{m}_t \oplus \tilde{s}_t \tag{6}$$

where $\tilde{x}$—fuzzy set, $\tilde{m}$—fuzzy trend component, $\tilde{s}$—fuzzy seasonal component. In this case it makes possible to develop a model similar to the models which are described in Sect. 3 but all operations are implemented with fuzzy sets.

For the implementing of fuzzy time series smoothing it is necessary to define operations on fuzzy sets. In order to implement the methods it is necessary to use the following operations: set addition, multiplication, division, multiplication by scalar. Operations on fuzzy sets, according to the author's description, are the operations on borders and nodes of membership function. In this paper triangular symmetrical membership function was selected.

2.4 Fuzzy Models Implementation (DSA)

The method DSA is based on fuzzy logic and the extrapolated methods, which were introduced by Song & Chissom and Chen. New functions in fuzzification of DSA method which describe membership functions as well as the overlapping between fuzzy sets are very important. This was not implemented in the methods of Song & Chissom and Chen. DSA method also provides a latent capability to implement inner and external forecasting. Though it is possible to measure unlimited number of fuzzy sets, in this investigation the maximum number of sets is 20. Earlier it was considered that it was optimal to take seven fuzzy sets, [14].

In this investigation membership intervals were determined for every fuzzy set and each interval is connected with definite grade of membership in the range [0,1]. The number of intervals should be enough for making out the difference of membership rate of observation and differs from the model. This simplified fuzzification of the results transfers input fuzzy set to the output module. Though it is possible to transfer more than one fuzzy set of observation to the output module, one set was selected because it is the simplest approach and it is the best starting point for the development of DSA method. The set of rules "If … then" is used in DSA module in order to make conclusions about the relation between input fuzzy sets. For every condition and consequence of couple of sets define a future state.

In defuzzification module of the rule "If … then" is used the center of defuzzificaton set for transformation of fuzzy forecasting to numerical forecasting. As a rule if there exits one and only one set for fuzzy forecasting then numerical forecasting is a center point of a fuzzy set, and if there two or more sets for fuzzy forecasting then numerical forecasting is a center of all fuzzy sets of fuzzy forecasting.

2.5 Combinations and Collectives of Models

Experience has proven that integration of models into a combination or a collective provides better results than the results of each model separately.

The main idea of the approach to a combination is a calculation of weights for the models. The project implements a weight calculation method reviewed by Stephan Kolassa in [15]. When calculating forecast we use model value taking into account its weight later on. The same way it is possible to calculate weights with another criteria.

Another approach to building of models' combination was suggested by LIANG Ai-hua, MEI Chi, E Jia-qiang, SHI Zhang-ming in [4] and its application. This approach is based on fuzzy models weights.

Vovk's aggregating algorithm considers the forecasting process as a triplet:

$$< \Omega, \Gamma, \lambda > \tag{7}$$

where Ω—an outcome set, Γ—a set of feasible predictions, $\lambda: \Omega \times \Gamma \rightarrow R$—loss function [5].

Forecasting algorithm takes the form of the function A:Ω* → Γ, which gives a forecast. Aggregating algorithm operates several models by weighting expert predictions errors in exponential space and then providing forecast with the decision rule:

$$\gamma_{T+1} = \sum \log\left(\sum_{j=1}^{N} \exp\left(-\mu Loss_{A_j}\left(\overleftrightarrow{\omega}\right)\right) p_0^{(j)}\right); \tag{8}$$

where p—expert weight at the initial time, $\mu > 0$.

3 Classification of Models

In this section we show models, that we use for forecasting. By analogy with the methods of exponential smoothing it is proposed to use parameter w.

$$Forecasted\_value = w \times Glob\_info + (1 - w) \times Local\_info. \tag{9}$$

In order to use this approach it is necessary to determine how to define seasonal and trend components for their using in additive or multiplicative model. It is proposed to take as a seasonal component the value Glob_info. It was proposed to determine the seasonal component according to the system of fuzzy logical deduction.

Therefore, it has become possible to implement for methods combining methods without trend component, without seasonal component and with these components.

The list of models is:

- without trend and seasonal components: $Forecasted\_value = Local\_info$. In this case smoothing is replaced by fuzzification.
- without trend component but with additive seasonal component of period p:

$$Forecasted\_value = \gamma * Local\_info + (1-\gamma) * Season\_info.$$

- additive trend component without seasonal component:

$$Forecasted\_value = \delta * Local\_info + (1-\delta) * Glob\_info.$$

- additive trend component and additive seasonal component of period p:

$$Forecasted\_value = \delta * \gamma * Local\_info + (1-\gamma) * Season\_info + (1-\delta) * Glob\_info$$

According to Reinhard Viertl notation, the formal representation of the methods will be the following:

- Model without trend and seasonal components $\tilde{x}_t = \tilde{y}_t$.
- Model with a seasonal component: $\tilde{x}_t = \tilde{y}_t \oplus \tilde{s}_t$.
- Model with a multiplicative seasonal component: $\tilde{x}_t = \tilde{y}_t * \tilde{s}_t$
- Model with additive trend component: $\tilde{x}_t = \tilde{y}_t \oplus \tilde{m}_t$..
- Model with a multiplicative trend component: $\tilde{x}_t = \tilde{y}_t \odot \tilde{m}_t$.
- Model with an additive trend and additive seasonal components: $\tilde{x}_t = \tilde{y}_t \oplus \tilde{m}_t \oplus \tilde{s}_t$.
- Model with an additive trend and multiplicative seasonal components:

$$\tilde{x}_t = (\tilde{y}_t \oplus \tilde{m}_t) \odot \tilde{s}_t$$

- Model with a multiplicative trend and additive seasonal components:

$$\tilde{x}_t = \tilde{y}_t \odot \tilde{m}_t \oplus \tilde{s}_t.$$

- Model with a multiplicative trend and multiplicative seasonal components:

$$\tilde{x}_t = \tilde{y}_t \odot \tilde{m}_t \odot \tilde{s}_t$$

- The last model is DSA.

We implement a combination of models using weights obtained by information criteria [15], and a combination using fuzzy weights [4] and a collective obtained by Vovk aggregating algorithm [5].

For the correct operation of the algorithm the time series data must be normalized.

4 Experiment Results

Let's examine the proposed approach on CIF-dataset [13]. The first step is prediction by each model the time series to the test interval. The length of the test interval corresponds to the desired length of the forecast interval for each time series (the value is in the data set). We used the symmetric mean absolute percentage error (SMAPE) as our error measure:

$$SMAPE = \frac{1}{M}\sum_{m=1}^{M}\frac{|\hat{y}_m - y_m|}{(|\hat{y}_m| + |y_m|)/2} \tag{10}$$

where y_m is the actual time series value, $\hat{y}_m$ is the forecast, and M is the size of the test period.

The results of the each model used as basic algorithms for the Vovk aggregating algorithm. We choose models with best results. The Table 1 shows the quality of forecasts by criterion SMAPE.

Classical methods were the best for daily time series only in 10 % cases, fuzzy methods were the best for monthly time series in 78 % cases and for quarterly time series in 92 %, for yearly time series in 90 % cases. Results are shown depending on the forecasting horizon in Table 2.

Table 1 Quality of forecasts for different time series

Type of time series	Classic average SMAPE, %	Fuzzy average SMAPE, %	Vovk SMAPE, %
Daily	16,67	17,96	41,22
Monthly	8,05	6,13	4,59
Quarterly	14,61	6,68	31,19
Yearly	9,67	2,53	26,42
Average	12,25	8,32	25,85

Table 2 Quality by forecasting horizon

Forecasting horizon, points	Classic average SMAPE, %	Fuzzy average SMAPE, %
4	10,83	5,19
5	13,45	3,52
12	5,96	4,84
18	0,30	0,59
30	21,77	21,58
90	8,86	11,67
180	45,35	41,27
Average	14,95	12,26

By analogy with analysis first group of time series: fuzzy methods were the best for 4 forecasting points in 87 % cases, for 12 forecasting points in 89 % cases, for 90 forecasting points in 87,5 % cases, for 180 forecasting points in 50 % cases. The classical methods were the best for 18 forecasting points in 100 % cases. Fuzzy methods were the best in 100 % cases for 5 and 30 forecasting points. Fuzzy models were the worse at 40 % on SMAPE, but classical models were the worse at 60 % on SMAPE. Fuzzy models inferior the classical models by 40 % SMAPE in the worst cases, classical methods inferior the fuzzy models by 60 % SMAPE.

5 Conclusion

In this paper we short describe our approach to the combination of several fuzzy and classical models. Group of fuzzy models shows better quality, but not always stable [16, 17]. Our approach with data transformation and combination of several models intended to increase methods stability with a time-varying conditions.

References

1. Fildes, R., Hibon, M., Makridakis, S., Meade, N, et al.: Generalising about univariate forecasting methods: further empirical evidence. Int. J. Forecast. **14**, 339–358, ISSN: 0169-2070 (1998)
2. Makridakis, S., Hibon, M. The M3-competition: results, conclusions and implications. Int. J. Forecast.**16**, 451–476 (2001). (Remus, W., O'connor, M.)
3. Small, G., Wong, R.: The Validity of Forecasting. 2002 Pacific Rim Real Estate Society (PRRES) Conference. Christchurch, New Zealand
4. Jiang, A., Mei, C., Jia-qiang, E., Shi, Z.: Nonlinear combined forecasting model based on fuzzy adaptive variable weight and its application. J. Central South Univ. Technol. **17**(4), 863–867
5. Vovk, V.G.: Universal forecasting algorithms. J Inf. Comput. **96**, 245–277 (1992)
6. Forecast combinations. In: Elliott, G., Granger, C.W.J., Timmermann, A. (eds.), Handbook of Economic Forecasting, Elsevier Publication, pp. 135–196.
7. Combination forecasts of output growth in a seven-country data set. J. Forecast. **23**, 405–430
8. Box, G.E.P., Cox, D.R.: An analysis of transformations. J. Roy. Stat. Soc. B **26**, 211–252 (1964)
9. Peng, G., Wang, J., Ren, P.,Gao, H., Luo, Y.: A new improved forecasting method integrated fuzzy time series with exponential smoothing method. Int. J. Environ. Pollut. **51**, Nos. ¾, 206–221 (2013)
10. Viertl, R.: Statistical Methods for Fuzzy Data. Wiley (2011)
11. Plouffe, J.S.: The Development and Validation of Fuzzy Logic Method for Time-series Extrapolation. University of Rhode island (2005)
12. Huarng, K. Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets Syst. № 12, 387–394 (2001)
13. http://irafm.osu.cz/cif/cif-dataset.txt
14. Song, Q., Chissom, B.: Forecasting enrollments with fuzzy time series. Part I. Fuzzy Sets Syst. **54**, 1–9 (1993)

15. Kolassa, S.: Combining exponential smoothing forecasts using Akaike weights. Int. J. Forecast. **27**(2), 238–251 (2011)
16. Perfilieva, I., Yarushkina, N., Afanasieva, T., Romanov, A.: Time series analysis using soft computing methods. Int. J.General Syst. 42(6) (2013)
17. Perfilieva, I., Novák, V., Romanov, A., Yarushkina, N.: Time Series Grouping and Trend Forecast Using F1-Transform and Fuzzy Natural Conference: The 11th International FLINS Conference (FLINS 2014) LOGIC. doi:10.1142/9789814619998_0026

Time Series Knowledge Mining Based on Temporal Network Model

Sergey Kovalev, Andrey Sukhanov, Alexey Averkin and Sergey Yarushev

Abstract This paper presents a novel model of knowledge representation for time sequences. This approach is based on a new class of the temporal network models. The basis of proposed technique is temporal associative rules describing the causal relations between process states and temporal shapes, which take place in past. These shapes are represented by the sequences of parametric temporal relations. The technique of automatic training based on empirical data for temporal network models is also proposed.

Keywords Time series knowledge discovery · Temporal relations · Associative temporal rules · Temporal data mining · Temporal shapes

The reported study was funded by RFBR, according to the research project (Grants No. 10-07-00158 and No. 10-01-00058) and partially supported by Grant of SGS No. SP2015/151, VSB—Technical University of Ostrava, Czech Republic.

S. Kovalev · A. Sukhanov (✉)
Rostov State Transport University, Rostov-on-Don, Russia
e-mail: andrey.sukhanov.st@vsb.cz; iiti16@rgups.ru

S. Kovalev
e-mail: drewnia@rambler.ru

A. Sukhanov
VŠB—Technical University of Ostrava, Ostrava, Czech Republic

A. Averkin
Dorodnicyn Computing Centre of RAS, Moscow, Russia
e-mail: averkin2003@inbox.ru

S. Yarushev
Dubna International University for Nature, Society and Man, Dubna, Russia
e-mail: sergey.yarushev@icloud.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 450, DOI 10.1007/978-3-319-33609-1_5

1 Introduction

Technologies for knowledge discovery in empirical data are based on the tools of Data Mining [1] and, hence, they can be efficiently used as a framework for problem decision in various analytical and practical domains. Among them, the particular interest is produced by the task of temporal data classification [2]. This task comes when intelligent systems for technological processes control as well as systems for dynamic data control and monitoring diagnosis systems of dynamical type should be constructed for industrial, transport and manufacturing domains.

The model selection for knowledge representation is very important in the area of Data Mining, when the key objective is knowledge discovery in temporal data. These models should satisfy to the set of requirements, which is divided into general and specific (or application associated) ones. Representation model for temporal data depends on way, by which preliminary information about observed process is presented. Preliminary information is usually presented in form of time series, which are referred to the data obtained from measurements and/or sensor systems describing dynamics of initial time process. Time series elements can be represented by continuous as well as discrete (symbolic) values. Continuous time series are often transformed into symbolic form because discrete representation is more convenient for processing [3, 4].

This paper presents a novel approach for temporal data modeling and considers a new technique for time series knowledge mining from symbolic sequences. The paper is organized as follows. Section 2 presents a problem statement for discrete time series representation and time series knowledge mining. Section 3 describes the notion of temporal network model. Section 4 shows the illustrative example of temporal network. Section 5 proposes the method for optimal temporal network construction. Section 6 gives the fuzzy interpretation of proposed framework. Conclusions and future work are discussed in Sect. 7.

2 Problem Statement

Symbolic time series can be defined as a sequence of consecutive symbols from finite state space Q. Consequence of symbols is represented by the connection using the fixed time intervals t on discrete time scale T:

$$S = (s(t_1), s(t_2), \ldots, s(t_n))(s(t_i) \in Q).$$

Let q be a symbol from Q, which is labeled as objective. Temporal model of elementary knowledge (TMK) for symbol q reflects the causal relationships between q and temporal shape in the past. It can be computed in the form of temporal production rule, which is expressed as follows:

$$\text{IF } G, \text{ THEN } q, \tag{1}$$

where q is the objective symbol, G is the description of the temporal shape, which take place before symbol q is obtained.

As knowledge element, TMK should meet the requirements [5, 6], such as completeness, consistency and interpretational fitness. Completeness of TMK is applied for temporal production rule system. It means that each symbol $q = s(t_i)$ has at least one rule, which determines this symbol obtainment. Temporal network consistency of TMK means that temporal shape observation definitely indicates following q. Interpretational fitness of TMK characterizes the possibility of rule representation in form of compact verbalized descriptions. Such representation considers utilizing an optimal numbers for variables in temporal rule and for temporal index as well.

Let the considered temporal rule be satisfied to above mentioned requirements. Semantics of the temporal rule is defined by a temporal logical formula providing the rule antecedent. This formula describes the generalized shape of one or many temporal scenarios, which take place before q is obtained in time series S. Temporal formula variables are presented by temporal relations τ^k, which have a meaning "In k steps before", and calculated for each time step t_j and objective symbol q as:

$$g\tau^k \quad \Leftrightarrow \quad (q = s(t_i), \quad g = s(t_j), \quad i - j = k). \tag{2}$$

Notion of temporal relation considers that the parameter k called temporal index characterizes the time interval, when symbols of time series are recorded.

Temporal formula Φ describing an antecedent of temporal rule has the form of the following expression:

$$\Phi = (\mathop{\&}_{j} \varphi_{i_j}) \mathbin{\&} \overline{(\mathop{\&}_{n} \varphi_{i_n})} \mathbin{\&} \cdots \mathbin{\&} \overline{(\mathop{\&}_{m} \varphi_{i_m})}, \tag{3}$$

where $\varphi_{i_j}, \varphi_{i_n} \ldots \varphi_{i_m}$ are the temporal relations described by (2).

Temporal formula (3) includes two types of relations, which characterize both generalizing and detailing features for its temporal shape. Conjunctive group $\&\varphi_{i_j}$ from (3) characterizes generalizing features of constructed temporal shape. These features are describe the "raw" description of presented area in temporal feature space. Inverse conjuncts $\overline{(\&\varphi_{i_n})}$ describe the distinctive properties for "improper" temporal shapes, which belong to area under impact of generalizing feature $\&\varphi_{i_j}$ because of overgeneralization.

Each equation of (3) describes a certain temporal scenario, which is obtained before observed time step t_i and leads to the objective symbol appearance. Therefore, Eq. (3) presents the prediction model for an objective symbol.

The task of knowledge discovery from time series S is to form a complete system of consistent temporal rules (3), which are satisfied to interpretational fitness requirement, for given objective symbol $q \in Q$. Interpretational fitness is established by minimization of variables as well as by maximization of temporal indexes for the antecedents.

Proposed statement is referred to a class of the problems for inductive generalization of data based on feature representation. To solve similar problems, various approaches are proposed [7–9]. The proposed framework reflects temporal specifications of initial data taking into account the new criterion, which characterizes interpretational fitness of time series knowledge model.

3 Temporal Network Model

Let TMK be presented by temporal network model [10], which is a modified form of growing pyramidal network [8]. Temporal network is defined for given objective symbol $q \in Q$ of time series S and may be represented by network graph with three layers, where each layer presents certain level of temporal data generalization. The first layer of network corresponds preliminary features, which are represented by relations $\varphi_i = g\tau^k$, the second and third ones present conjunctive groups of generalizing $\&\varphi_{i_j}$ and detailing $\overline{(\&\varphi_{i_n})}$ features, respectively.

Formal representation of temporal network may be expressed as described below.

Let S be a time series and $q = s(t_i)$ be an objective symbol of this time series. Let parameter L denote the width of window, inside of which causal relations are calculated between the symbols of S, and l be the given value of L $(l \in N)$. Let $g = s(t_{i-1}), p = s(t_{i-2}), \ldots, h = s(t_{i-l})$ be a sequence in past of the objective symbol q. In this case, the temporal scenario, which takes place before symbol q is obtained, is expressed as follows:

$$e_i = \{s(t_{j=i-k})\tau^k | k = 1, 2, \ldots, l\}. \quad (4)$$

Equation (4) denotes scenario from time series S in respect of symbol $s(t_i)$.

Let $E = \{e_i | i = l+1, l+2, \ldots, n\}$ denote the set of all scenarios from S. E can be divided into two groups: positive E^+ and negative E^-:

$$E^+ = \{e_i \in E \mid s(t_i) = q\},$$

$$E^- = \{e_i \in E \mid s(t_i) \neq q\}.$$

Model of temporal network is specified as an acyclic graph $H = \langle X, \Gamma \rangle$, where the vertices are divided into three layers $X = T \cup G \cup D$ in such a manner that the layer T corresponds the input set of preliminary features $\varphi_i = g\tau^k$, the layer G corresponds the set of generalizing features and the layer D corresponds the set of detailing features. Mapping Γ specifies the structure of interlayer relations, where the vertices of T are connected by arcs with the vertices of second and third ones. By this connections, the conjunctive and detailing feature groups are constructed, respectively. Vertices from second layer with connection of vertices from third one show the correspondence between generalizing and detailing features.

It should be noted that first (input) layer of network graph H is defined uniquely for given time series S with window width l, while second and third layers as well as mapping Γ may be defined by various ways. However, second and third layers should satisfy to the set of requirements arising from the problem statement.

Let vertices $g \in G,\ d \in D$ be related to sets $T(g) = \Gamma^{-1}(g)$ and $T(d) = \Gamma^{-1}(d) \cap \mathrm{T}$, which are constructed by input vertices connected with the vertices of G and D, respectively.

The generalizing assumption, according to which vertices of generalizing layer should be constructed, is calculated as follows:

$$(\forall e \in E^{+} \exists g \in G) \quad (\mathrm{T}(g) \subseteq e). \tag{5}$$

In other words, each positive sample e^{+} from generalizing layer should have the vertex of g, which has feature set included in this sample.

Vertices of detailing layer should conform to the thresholding condition:

$$\begin{gathered} (\forall g \in G, \forall e^{-} \in E^{-}(g)\ \exists d \in \Gamma(g)) \\ (T(d) \subseteq e^{-})\ \&\ (\forall e^{+} \in E^{+}(g)\ T(d)\ \not\subseteq e^{+}), \end{gathered} \tag{6}$$

where $E^{-}(g) = \{e \in E^{-} \mid T(g) \subseteq e\}$ is the set of negative samples and $E^{+}(g) = \{e \in E^{+} \mid T(g) \subseteq e\}$ is the set of positive ones.

Actually, this assumption reformulates the rule for control vertices allocation from growing pyramidal networks proposed in [8]. According to the thresholding condition, each negative sample $e^{-} \in E^{-}(g)$ should have detailing feature $d \in \Gamma(g)$, feature set $T(d)$ of which belongs to e^{-} and does not belong to any positive sample $e^{+} \in E^{+}(g)$. This requirement is satisfied if and only if each positive sample $e^{+} \in E^{+}(g)$ has relation $\phi_{i_j} \in TT(d)$ has, which belongs to e^{-} and does not belongs to e^{+}, i.e. it belongs to subtraction $e_i - e_j$.

It should be noted that network graph H is organized in the manner, where each subgraph, which is formed by vertex g with connection of subset of the detailing vertices $d_i \in \Gamma(g)$ together with the correspondent feature subsets $T(g)$ and $T(d_i)$, is uniquely described by Eq. (3). The generalizing vertex corresponds to the generalizing conjunct of a temporal formula. Detailing vertices connected with the generalizing vertex characterizes the inverse conjuncts.

4 Example of Temporal Network

The simple example of temporal network for time series $S = (xxxyxyyxxxyxyyxxxyxyy)$ is presented on Fig. 1. Here, symbol x is an objective and window width $l = 4$.

Layer T corresponds to the input set of preliminary features. It is represented as a set of temporal relations for the given example:

$$T = \{x\tau^1, x\tau^2, x\tau^3, x\tau^4, y\tau^1, y\tau^2, y\tau^3, y\tau^4\}.$$

The layer of generalizing features is represented by the set of relations:

$$G = \{y\tau^4, x\tau^4 \,\&\, x\tau^1, y\tau^1 \,\&\, x\tau^3\}.$$

Two groups of relations represent the detailing layer:

$$D = \{x\tau^2 \,\&\, x\tau^3, x\tau^3\},$$

where conjunct $x\tau^2 \,\&\, x\tau^3$ has a role of the detailing feature for generalizing feature $y\tau^4$ and relation $x\tau^3$ has a role of the detailing feature for generalizing feature $x\tau^4 \,\&\, x\tau^1$. Accordingly, the connections of graph are established.

Above described model of temporal network has the structure combining the shapes of the scenarios, which are obtained in the past of x. In particular, the structure of Fig. 1 implements the temporal associative rule of inference:

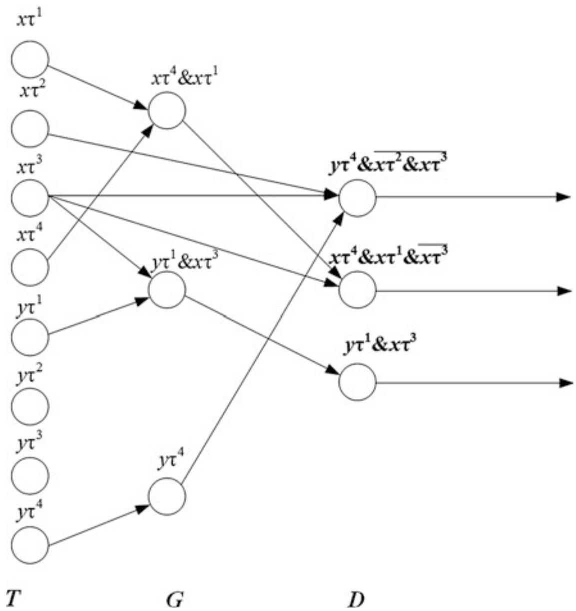


Fig. 1 Structure of temporal network for the given example

$$y\tau^4 \,\&\, \overline{(x\tau^2 \,\&\, x\tau^3)} \vee x\tau^4 \,\&\, x\tau^1 \,\&\, \overline{x\tau^3} \vee y\tau^1 \,\&\, x\tau^3 \quad \Rightarrow \quad s(t_i) = x.$$

A complete system of temporal rules is formed by constructing the temporal network for all symbols from a time series and matching the correspondent temporal formulas as antecedents for network.

5 Temporal Network Optimization

The task for constructing the interpretationally convenient TMK of time series is converged to construction of the temporal network, which contains minimal number of vertices in generalizing and detailing layers. Optimization of the network is performed by analyzing the set of numeric features, which reflect potential usefulness of particular temporal relation implementation for generalizing and detailing.

The frequency of occurrence for positive samples and the frequency of non-occurrence for negative ones are the most valuable characteristics for generalizing features in respect of interpretational fitness. This compromise can be reached utilizing the next criterion:

$$K(\underset{j}{\&}\varphi_j) = \frac{|E^+(\&\varphi_j)|}{|E^+|}\left(1 - \frac{|E^-(\&\varphi_j)|}{|E^-|}\right), \tag{7}$$

where $\&\phi_i$ represents the conjunctive temporal relations group, which corresponds the generalizing feature g; $E^+(\&\varphi_i) = \{e \in E^+ \mid \varphi_i \in e\}$; $E^-(\&\varphi_i) = \{e \in E^- \mid \varphi_i \in e\}$.

Another important characteristic of generalizing feature is a total number of variables used in description of feature. It should be noted that maximal value of temporal index (or temporal depth) is also characteristic, which is important in contexts of generalizing feature determination.

Total number of variables belonging to generalizing feature may be decreased by incorporating optional temporal generalizations with interval relations $g\tau_n^k$:

$$g\tau_n^k \Leftrightarrow g\tau^k \,\&\, g\tau^{k-1} \,\&\, \cdots \,\&\, g\tau^{k-n}.$$

This interval relation means “Symbol g with duration of n is obtained k steps before”.

Algorithm for generalizing features construction is based on the heuristic analysis of above mentioned numeric characteristics. Algorithm represents iterative procedure for construction of temporal relations group, which maximizes generalizing criterion (7) and covers all positive samples of the constructed class. Minimization of temporal depth for generated descriptions can be reached by choosing temporal relations $g\tau^k$ with minimal temporal indexes k in a process of search.

Detailing features describe space subregions belonging to intersection of different various description classes. The main aim of their usage is to eliminate the generalizing features impact for these areas. Because of this, each generalizing feature forms its own subregion of detailing features.

Let g be the generalizing feature, $E^-(g)$ be the set of negative samples and $E^-(g)$ be the set of positive ones. According to thresholding condition (6), each positive sample $e^+ \in E^+(g)$ has relation $\varphi_{i_j} \in T(d)$, which belongs to e^- and does not belongs to e^+, i.e. it belongs to subtraction. Thus, thresholding condition is satisfied by the following disjunction:

$$\varphi_{i_{j1}} \vee \varphi_{i_{j2}} \vee \cdots \vee \varphi_{i_{j|E^-(g)-E^+(g)|}} = \vee_k \varphi_{i_{jk}}, \tag{8}$$

The condition (8) should be fulfilled for each positive sample $e_j^+ \in E^+(g)$, i.e.:

$$\left(\vee_k \varphi_{i_{1k}}\right) \& \left(\vee_k \varphi_{i_{1k}}\right) \& \cdots \& \left(\varphi_{i|E^+(g)|k}\right), \tag{9}$$

Obviously, each implicant of (9) satisfies to the thresholding condition (6) with respect to a sample e_i^- and, therefore, it can be used as detailing feature for ith negative sample e_i^-. If Eq. (9) is revealed and the law of absorption is used the expression characterized the set of all minimal implicants (features with minimal number of relations) is obtained:

$$J_{i_1} \vee J_{i_2} \vee \cdots \vee J_{i_n} = \vee_k J_{i_k}, \tag{10}$$

where J_{i_j} is a jth implicant of (9).

The expression characterizing a complete set of detailing features for initial generalizing feature g may be obtained by making conjunction of the expressions (10) for all negative samples:

$$(\vee_k J_{1_k}) \& (\vee_n J_{2_n}) \& \cdots \& (\vee_m J_{z_m}), \tag{11}$$

where z is the number of negative samples.

If Eq. (11) is revealed and the law of absorption is used, then the expression for minimal detailing feature subregions is obtained:

$$(\underset{k}{\&} J_{1_k}) \vee (\underset{n}{\&} J_{2_n}) \vee \cdots \vee (\underset{m}{\&} J_{z_m}). \tag{12}$$

It is clear that optimal set of detailing features is correspondent by one of the conjuncts containing, for instance, minimal number of belonging implicants or minimal number of all belonging relations in respect of chosen criterion of the interpretational fitness for (12). This set can also be correspondent by the conjunct contained the relations with minimal temporal indexes. Obviously, the various combinations of above mentioned criteria are allowed.

Above given reasoning is considered as the basis of algorithm for detailing features optimal set construction. Initial data of the algorithm is represented in form of rectangular matrix M_d, rows of which correspond to temporal relations φ_{i_j}. Columns of the matrix are combined into z groups, which correspond to negative samples e_i^-. Each negative group contains m columns correspondent to positive samples. Intersection of nth row and jth column (which is from ith group) has the value of 1 if nth feature belongs to the subset $e_i^- - e_j^+$. Algorithm of detailing features construction converges to procedure for searching the optimal row coverage of combined matrix.

6 Fuzzy Interpretation of Temporal Network

Fuzzy interpretations of TMK are aimed to describe semi-structured processes, which are characterized by various non-factors [11]. These generalizations reflect an approximate expert reasoning about time intervals between events.

Fuzzy generalization of temporal relation can be produced, if parameter k from (2) is replaced by fuzzy value. In particular, if fuzzy value z ≈ 1 "APPROXIMATELY ONE" is taken instead of k (the membership function is $\mu_{z\approx 1}(n) \quad (n \in N)$) then fuzzy relation φ: "IN IMMEDIATE PAST" is produced. The fuzzy relation is defined by membership function μ_φ within pair set $S \times q \quad (q = s(t_i)$. In this case, the membership of pair $(g\ ,\ q)\ \left(g = g(t_j)\right)$ is computed as follows:

$$\mu_\phi(g\ \varphi\ q) = \mu_{\approx 1}(i - j). \tag{13}$$

Temporal relation τ:"IN SEVERAL PREVIOUS STEPS" is produced if k is replaced by the fuzzy interval value [3–5], which reflects intuitive ideas of human experts about the qualitative value "SEVERAL STEPS".

In general case, system $\mathfrak{T} = \{\tau_{z1}, \tau_{z2}, \ldots, \tau_{zk}\} \quad (z_i \in N)$ may be used. Each element from $\mathfrak{T}$ characterizes the time interval "APPROXIMATELY z_j". The membership function is produced for each fuzzy relation τ_{zj} and computed by the following equation:

$$\mu_{zj}(g,\ q) = \mu_{\approx zj}(i - j) \quad (g = x(j),\ q = x(i)\ ), \tag{14}$$

where $\mu_{\approx zj}(\cdot)$ is the membership function for fuzzy value "APPROXIMATELY z_j".

Fuzzy model of temporal scenario, which reflects the past evolution of an observed process relatively to the state $q = x_t$, can be computed as follows:

$$\Phi = (s_{j1}\tau_{zj1}\ q)\ \&\ (s_{j2}\tau_{zj2}\ q)\ \&\ \cdots\ \&\ (s_{jk}\tau_{zjk}\ q). \tag{15}$$

Before the Eq. (15) is applied, the procedure of preprocessing should be provided by human expert to emphasize on interesting events in past and eliminate insignificant conjuncts.

7 Conclusions and Future Work

Proposed class of temporal networks is suitable for temporal model representation in form of temporal rule base. The novel approach has a set of advantages in respect to previous ones. In particular, the sets of associative vertices are limited (in comparison with the sets of conceptors from growing pyramidal networks) that leads to reducing the procedures of training and model adaptation. Moreover, detailing feature usage brings new properties to structure of associated temporal rules. In particular, the opportunity for constructing a new class of self-organizing temporal system arises. This system is considered as fuzzy temporal data processing system, which adapts their parameters according to parameters of negation operators. This fact produces the possibility for implementation of new strategies for training adaptive temporal systems based on biological processes imitation, where the negation principles are used. Specifically, based on theory of artificial immune systems, complementary features formation and negative selection can be applied.

Considered approach for generalized descriptions of dynamic objects can find the implementation in various domains such as automatic construction of temporal databases for intelligent systems of dynamic type, intelligent analysis of dynamic data, temporal shape recognition and anomaly detection and prediction in time series.

References

1. Leskovec, J., Rajaraman, A., Ullman, J.D.: Mining of massive datasets. Cambridge University Press (2014)
2. Grabocka, J., Nanopoulos, A., Schmidt-Thieme, L.: Invariant time-series classification. In: Machine Learning and Knowledge Discovery in Databases, pp. 725–740. Springer, Heidelberg (2012)
3. Lin, J., Keogh, E., Lonardi, S., Chiu, B.: A symbolic representation of time series, with implications for streaming algorithms. In: Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, pp. 2–11. ACM (2003)
4. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing SAX: a novel symbolic representation of time series. Data Min. Knowl. Disc. **15**(2), 107–144 (2007)
5. Mörchen, F.: Time series knowledge mining. Citeseer (2006)
6. Sripada, S.G., Reiter, E., Hunter, J., Yu, J.: Generating English summaries of time series data using the Gricean maxims. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 187–196. ACM (2003)

7. Vagin, V.N., et al.: Metody teorii priblizhennykh mnozhestv v reshenii zadach obobshcheniya ponyatiy [The methods of rough sets theory for solving the problems of generalization of concepts]. Izvestiya RAN. Teoriya i sistemy upravleniya. **6**, 52–66 (2004)
8. Gladun, V.P.: Planirovaniye resheniy (Decision planning). Naukova Dumka, Kiev (1987)
9. Hoa, N.S., Son, N.H.: Some efficient algorithms for rough set methods. In: Proceedings of the Conference of Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp. 1451–1456. Citeseer (1996)
10. Kovalev, S.M.: Formirovaniye temporal'nykh baz znaniy na osnove apparata rastushchikh piramidal'nykh setey (Temporal knowledge formation for growing pyramidal networks). Integrirovannyye modeli i myagkiye vychisleniya v iskusstvennom intellekte. Proceedings of 3rd International scientific-practical seminar, pp. 351–357. Fizmatlit, Moscow (2005)
11. Narinyani, A.S.: Nedoopredelennost' v sisteme predstavleniya i obrabotki znaniy [underdetermination in system for knowledge representation and processing]. Informatsionnyye technologii. **7**, 3–28 (1986)

Methods and Algorithms of Anomaly Searching in Collections of Time Series

Marina Fomina, Sergey Antipov and Vadim Vagin

Abstract One of the basic problems arising under processing temporal dependencies is the analysis of time series. The various approaches to processing of temporal data are considered. The problem of anomaly detection among sets of time series is setting up. The algorithm TS-ADEEP-Multi for anomaly detection in time series sets for the case when the learning set contains examples of several classes is proposed. The method for improving the accuracy of anomaly detection, due to "compression" of these time series is used. Modelling results for anomaly detection in time series are produced.

Keywords Clustering · Time series · Anomaly · Classification · Detection

1 Introduction

The development of up-to-date complex intelligent systems is closely connected with designing most perfective their representatives to which decision support systems are referred. An intelligent system can be viewed as a computer system to solve problem classes that cannot be solved by a person in real time or the decision requires automated support. The solution afforded by an intelligent system should give results comparable with decisions obtained by a person who is a specialist in a

M. Fomina (✉) · S. Antipov · V. Vagin
National Research University "Moscow Power Engineering Institute",
Krasnokazarmennaya 14, 111250 Moscow, Russia
e-mail: m_fomina2000@mail.ru

S. Antipov
e-mail: antisergey@gmail.com

V. Vagin
e-mail: vagin@appmat.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 450, DOI 10.1007/978-3-319-33609-1_6

certain domain. Characterization of a computer system as an intelligent one will be incomplete if the nature of solved problems and means of their decisions implemented due to some computer system architecture were not taken into account [1].

One of the most problem class, decision of which demands intelligent support of computer systems, is a manage problem of complex technical objects. The main feature of such manage objects is that they are dynamic ones and have ability for developing. The states of such objects and systems may change over time. Therefore one needs to develop methods and algorithms allowing to take into consideration a time factor under analysis of behavior of these objects.

Ones of the basic problems arising when processing temporal dependencies are problems of clustering and classification. Classification problems are solved on the example of anomaly search in collections of time series.

The paper is structured as follows. In Sect. 2, the problems arising under processing of time series are considered. Section 3 is devoted to a problem of anomaly detection and consists of the following sections: in Sect. 3.1, there is given the statement of the anomaly detection problem in sets of time series; a normalized representation of time series is described in Sect. 3.2; the algorithm of anomaly detection in sets of time series is given in Sect 3.3. Modeling results of the process of anomaly detection are presented in Sect. 4. Finally, conclusions are given in the last section.

2 The Problems of Processing Time Series

Because the analysis of complex technical objects behavior requires the consideration of a time factor, there is a need to work with data that explicitly (or implicitly) contain time. In this regard, one has to deal with the problem of temporal data mining [2, 3]. The most common case of this analysis is time series mining [4]. Time series are used in many different areas (technics, economics, medicine, banking, etc.) and describe different processes that occur over time.

The problem of time series mining is important for solving the following tasks of process analysis:

- to give forecast of a future process state depending on the qualitative assessment of a current or previous state;
- to bring out the presence of typical and abnormal event sorts;
- to bring about qualitative changes of an analyzed process on the basis of time series analysis and to find some time series trends in changing of the researched process states.

The following classes of methods are used to solve these problems: Associations, Sequence, Classification and Clustering [5].

In combination with other methods of data mining, *prediction* or *forecasting* involves trend analysis, classification and model matching. The basis for all kinds of forecasting is the historical information stored in the database in the form of time series. If one can build or find patterns, that adequately reflect the object behavior, it is likely that they can be used to predict the behavior of a system in the future.

The problems of detecting trends, their qualitative assessment and forecast based on analysis of time series are of particular relevance due to the continuous growth of real-time data from the specific and complex technical objects, for example, sensors whose values change over time.

Consider the case where the object's behavior is evaluated on the basis of particular parameter values observations. In general, the time series TS is an ordered sequence of values $TS = \langle ts_1, ts_2, ts_i, \ldots, ts_m \rangle$ describing the flow of a long process, where the index i corresponds to a time mark. ts_i values can be sensor indications, product prices, exchange rates and so on.

3 Anomaly Detection Problem

The anomaly detection problem [6] is set up as the task of searching for patterns in data sets that do not satisfy some typical behaviors. The ability to find anomalies in a data set is important in a variety of subject areas: in the complex technical system analysis (e.g. satellite telemetry), in network traffic analysis, in medicine (analysis of MRT images) in the banking industry (fraud detection) and etc.

The anomaly, or "outlier" is defined as an element that stands out from the data set which it belongs to and differs significantly from the other elements of the sample. Informally, the problem of anomaly detection in time series sets is formulated as follows. There is a collection of time series describing some processes. This collection is used to describe normal processes. It is required to construct a model on the basis of the available data, that is a generalized the description of normal processes and allows to distinguish between normal and abnormal processes.

The problem of obtaining accurate and representative data for learning and testing is usually very important. The boundaries between normal and abnormal behavior of an object are often not precise, and moreover, an exact notion of outlier is varied for different application domains.

3.1 Problem Statement

Let $TSStudy = \langle TSstudy_1, TSstudy_2, \ldots, TSstudy_m \rangle$ be a set of objects where each object is time series. We call it a study set. Each of the time series in a study set represents some "normal" behavior or a process flow. Based on the analysis of *TSStudy* one needs to build a model to distinguish the instances of time series from $TSTEST = \langle TStest_1, TStest_2, \ldots, TStest_n \rangle$ to "normal" or " abnormal".

This problem is divided into two cases [6]: single-class and multi-class anomaly detection, these are the cases when a study set contains instances of a single class of time series or instances of several classes of time series. In the first case, we need just to determine whether time series from a test set are anomalies or not; in the second case, we should also find out whether an object belongs to a particular class.

There are several classes of methods that solve the anomaly detection problem. Classification-based methods are used to learn a model from data assigned to different classes (learning stage), and to refer instances to one of the existing classes using the resulting model (test stage). Classification-based anomaly detection methods assume that if a classifier can be learned in an existing feature space, it can separate the normal and abnormal objects. Most widely used methods are neural networks [7]; bayesian belief networks [8]; support vector machines [9]; production rules [10]. The advantage of classification-based anomaly detection methods includes the ability to use a lot of machine learning techniques and algorithms, especially for the case when the study set contains examples of several classes. Further, the test stage is fast in comparison with other classes of methods, because of a previously constructed model (classifier).

Let us consider this problem with the simple example. Let *TSStudy* is a study set that contains six time series (Fig. 1).

The test set *TSTEST* consists of three time series (Fig. 2).

Based on the above problem statement it is clear that the time series (1), (2) and (5) (Fig. 1) are highly similar to each other, and therefore they are members of the same class, let it be “Class 1”. Time series (3), (4) and (6) are also similar, but belong to another class, let it be “Class 2”. The test set (Fig. 2) shows that the time series (1) is likely to be a member of “Class 2”, a time series (2) is a member of “Class 1”. The third time series is significantly different from the previous two, and apparently “not similar” to any in the study set. This implies that the process by

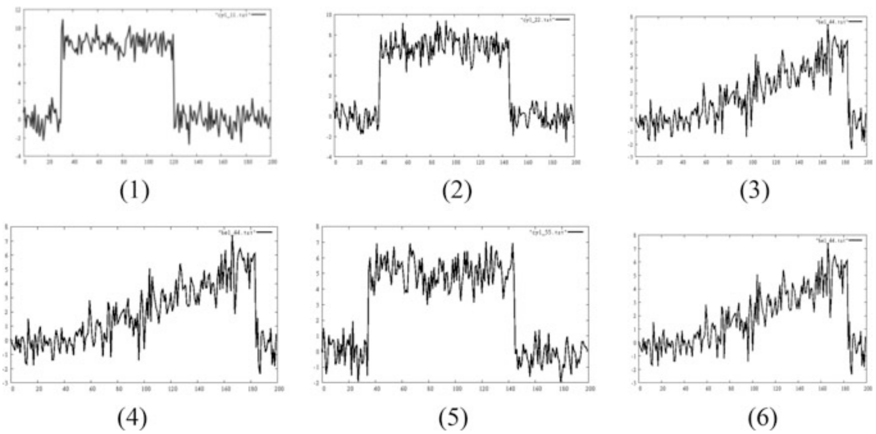


Fig. 1 The example of a study set

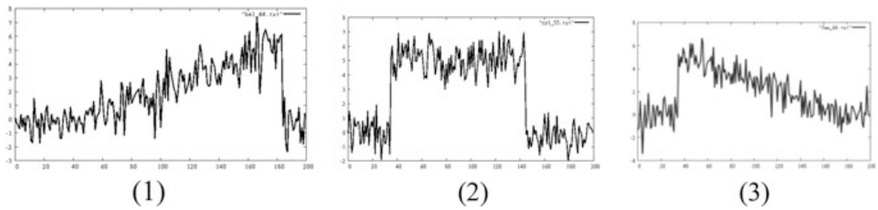


Fig. 2 The example of a test set

which time series (3) from the test sample was received is different from processes, by which time series from the study set were obtained. On the contrary, the time series (1) and (2) from the test set (Fig. 2) are not anomalies, since their shapes are very “similar” to the individual time series in the study set.

3.2 The Representation of Time Series

To create the generalization algorithm of the information provided by time series, we require some kind of time series preprocessing. It is required to transform time series, that represent data from different application domains, in different units to some forms convenient for further analysis. To work with time series, it is proposed to use two methods of their representation: a normalized and a symbolic representation [11]. Normalization is called a process of time series conversion so that its mean value would be equal to zero and a standard deviation would be one; this transformation is a necessary step of the data preprocessing [11]. Examples of original and normalized time series are shown in Table 1.

Symbolic representation of time series can be obtained from normalized one using the algorithm Symbolic Aggregate approXimation [11]. To perform this conversion, the alphabet $\boldsymbol{A}$—finite set of characters—is introduced: $\boldsymbol{A} = \{a_1, a_2, \ldots, a_{|A|-1}\}$ where some value interval of time series corresponds to one of an alphabet symbol.

Table 1 Various ways of time series representation

Time t	1	2	3	4	5	6	7	8	9
Current values	512	1448	88	1448	1448	1448	1448	1024	512
Normalized values	−1.04	0.748	−1.85	0.748	0.748	0.748	0.748	−0.063	−0.042
Symbolic representation	*C*	*P*	*A*	*C*	*C*	*C*	*C*	*J*	*C*

The example of symbolic representation for the time series is given in Table 1. Here the alphabet $\boldsymbol{A}$ consists of 20 characters, $\boldsymbol{A} = \{A, B, C, \ldots, T\}$.

3.3 The Algorithm of Anomaly Detection in Sets of Time Series

We propose a method for the anomaly detection in the sets of time series, This method is a modification of an "exact exception problem" [12] that is described as follows: for the given set of objects I, one needs to get an exclusion-set I_x. To do this, there are introduced:

- the function of dissimilarity $D\ (I_j)$, $I_j \subset I$: defined on $P\ (I)$, the set of all subsets for I_x and receiving positive real values;
- the cardinality function $C\ (I_j)$: $I_j \subset I$, defined on $P\ (I)$ and receiving positive real values such that for any $I_1 \subset I$, $I_2 \subset I$, $I_1 \subset I_2 \Rightarrow C\ (I_1) < C\ (I_2)$;
- the smoothing factor $SF\ (I_j) = C\ (I\backslash I_j) \cdot (D\ (I) - D\ (I\backslash I_j))$, which is calculated for each $I_j \subseteq I$.

Then $I_x \subset I$ will be considered an exclusion-set for I with respect to D(I), and $C(I)$, if its smoothing factor $SF\ (Ix)$ is maximal [12].

Informally, an exception-set is the smallest subset of I, that makes the largest contribution to its dissimilarity. A smoothing factor shows how much dissimilarity of a set I can be reduced, if from it to exclude a subset I_j. A dissimilarity function can be any function that returns a low value if elements of a set are similar to each other and a higher value if elements are dissimilar.

The algorithm TS-ADEEP that is based on this method was adapted for the anomaly detection problem in sets of time series. As a set I, we use $TSStudy \cup \{ts\_test_j\}$ for each $ts\_test_j \in TSTEST$. A dissimilarity function for time series is set up as follows:

$$D(I_j) = \frac{1}{N} \cdot \sum_{a \in I_j} \left|a - \overline{I_j}\right|^2 \text{ where } \overline{I_j} = \sum\nolimits_{a \in I_j} \frac{a}{|I_j|}$$

First, the average for the time series of I_j is calculated. The dissimilarity function is calculated as the sum of squared distances between the mean and vectors of I_j. The cardinality function is given by the formula $C\ (I - I_j) = 1/|\ I_j\ | + 1$. The formula for calculating the smoothing factor is $SF(I_j) = C\ (I - I_j) \cdot (D\ (I) - D\ (I_j))$.

If an exception set for $I = TSStudy \cup \{ts\_test_j\}$ contains ts_test_j, then ts_test_j is an anomaly.

Based on this method, the algorithm TS-ADEEP for anomaly detection in sets of time series was introduced [13].

In this paper we propose the algorithm TS-ADEEP-Multi that is a generalization of the algorithm TS-ADEEP for the case of a study set contains several classes of time series. The generalization is quite obvious: splitting study set to single class subsets and consequently applying the TS-ADEEP algorithm, we can determine whether the considered time series is an anomaly. If time series is an anomaly for each subset or time series is not an anomaly for the only subset of the study set, the answer is quite obvious. However, there is a case where the time series is not an anomaly for several study set subsets.

The algorithm TS-ADEEP-Multi is shown below:

The algorithm **TS-ADEEP-Multi**
input: (*TS Study*: study set that contains time series of several classes;
TSTEST: test set)
output: *TSAnom_O* - a set of anomaly time series of on the "optimistic" assessment
TSAnom_P - a set of anomaly time series on the "pessimistic" assessment
begin
  $TSAnom\_O = \varnothing$; $TSAnom\_P = \varnothing$
  Let *N* be a number of classes containing in the learning set
    $TSStudy\_C = \{TSStudy\_C_1, TSStudy\_C_2, .., TSStudy\_C_N\}$ is a partition of
    TSStudy such that $TSStudy\_C_k$ contains only examples of class *k*, $k = 1..N$
  for *j* from 1 to | *TSTest* |
    begin choose $TSTest_j$ of *TSTest*
     for *k* from *1* to *N*
       begin $I = TSStudy\_C_k \cup TSTest_j$; Find the exclusion-set I_x in *I*
       If the $TSTest_j \in I_x$ is an anomaly for the class *k* (it does not belong to him)
       then break
       end
    If $TSTest_j$ *does* not belong one from classes $TSStudy\_C_k$, $k = 1..N$ then
      $TSAnom\_O = TSAnom\_O \cup TSTest_j$;
      $TSAnom\_P = TSAnom\_P \cup TSTest_j$;
    end if
      If *TSTestj* belongs to a single class $TSStudy\_C_k$ then it is not an anomaly
      If *TSTestj* belongs to several classes $TSStudy\_C_k$ then
      $TSAnom\_P = TSAnom\_P \cup TSTest_j$
    end
    print *TSAnom_O, TSAnom_P;*
end

4 Experimental Results for Anomaly Detection in Time Series

A simulation of anomaly detection process was conducted on widely used datasets "cylinder-bell-funnel" [14] and "control chart" [15]. "Cylinder-bell-funnel" [14] contains three different classes—"cylinder", "bell", "funnel". "Control chart" [15] contains six different classes that describe the trends may be presented in the process: normal, cyclic, increasing trend, decreasing trend, upward shift, downward shift.

In order to determine how well the proposed algorithm deals with anomaly detection in collections of time series, several experiments were conducted. For the "cylinder-bell-funnel" dataset instances of two classes were used as study set, instances of all three classes were used as a test set. For the "control chart" dataset instances of 2, 3, 4 and 5 classes were used as a study set, while all 6 classes were used as a test set.

The experiment showed that the proposed algorithm for anomaly detection does not always show good results for original time series: for some pairs of classes only a little more than half of the time series are correctly assigned to anomalies.

To improve this situation, it is proposed to preprocess the original time series by reducing or compressing it on normalization stage. This makes it possible to discard irrelevant details and to get rid from the noise. Example of time series compression is shown in Fig. 3 (each ten points of original time series were assigned to a single point of new time series). The thin lines in the figure connects points of original time series. The thick line shows compressed time series: horizontal segments correspond to the 10 points of the original time series.

Applying the TS-ADEEP-Multi algorithm to preprocessed time series with reduced size showed significantly better results than applying this anomaly

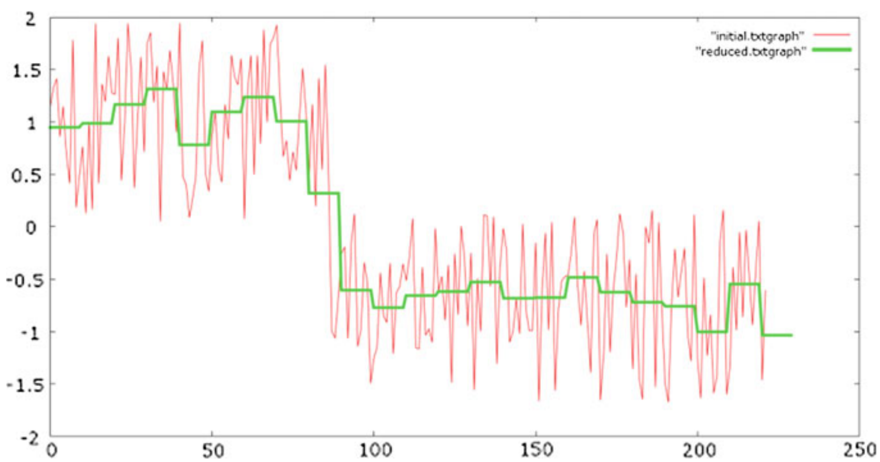


Fig. 3 The example of time series compression

detection algorithm to the original data without compression. Figure 4 shows the comparison of anomaly detection results for original and compressed data for "cylinder-bell-funnel" data set. There are four cases: the numerical time series

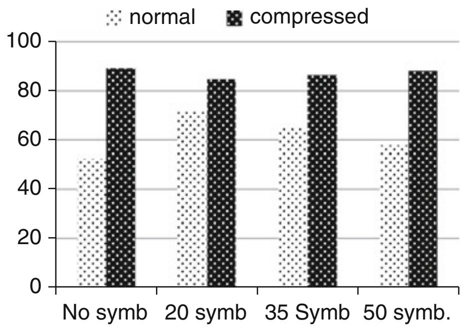


Fig. 4 Success enhancing an anomaly detection by compression of data sets "cylinder-bell-funnel"

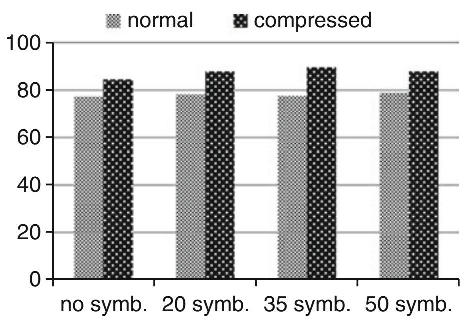


Fig. 5 Success enhancing an anomaly detection by compression of data sets "control chart" in the case of using 2 classes

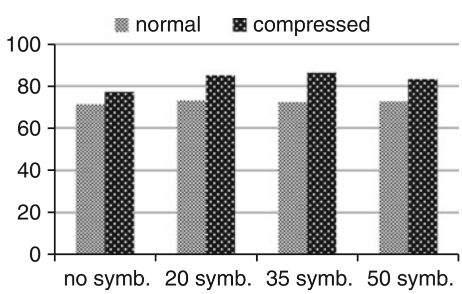


Fig. 6 Success enhancing an anomaly detection by compression of data sets "control chart" in the case of using 3 classes

representation, the symbolic row presentation with varied alphabet size: 20, 35 and 50 symbols. On the *Y*-axis, the percent of successfully recognized time series (normal or abnormal) is shown.

Figures 5 and 6 show similar results for the “control chart” data set.

5 Conclusion

The problem of anomaly search among time series is viewed. Non-parametric algorithm TS-ADEEP-Multi for defining anomalies in time series sets for the case when a learning set contains examples of several classes is offered. Simulation of anomaly detection for widely used data sets—“cylinder-bell-funnel” and “control chart”—are presented. Appropriate preprocessing of original time series—reducing dimensionality—increases accuracy of anomaly detection due to getting rid of inessential details and decreasing a noise. Further it is proposed to modify the suggested anomaly detection algorithm for the case when the classes of time series are a priori unknown.

This work was supported by the Russian Foundation for Basic Research (project nos. 14-07-00862 and 15-01-05567), and by the state project no. 2.737.2014 K.

References

1. Finn, V.K.: The synthesis of cognitive procedures and the problem of induction. In: NIT [Moscow, Russia: VINITI] Series **2**(1–2), 8–44 (1999) (In Russian)
2. Roddick, J., Spiliopoulou, M.: A bibliography of temporal, spatial and spatio-temporal data mining research. SIGKDD Explor. Newsl. **1**(1), 34–38 (1999)
3. Antunes, M., Oliveira, A.: Temporal data mining: an overview. In: Eleventh International Workshop on the Principles of Diagnosis (2001)
4. Perfilieva, L., Yarushkina, N., Afanasieva, T., Romanov, A.: Time series analysis using soft computing methods. Int. J. Gen. Syst. **42**(6), 687–705 (2013)
5. Vagin, V., Golovina, E., Zagoryanskaya, A., Fomina, M.: Exact and plausible inference in intelligent systems. In: Vagin, V., Pospelov, D. (eds.). FizMatLit, Moscow (2008) (in Russian)
6. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection—a surey. ACM Comput. Surv. **41** (3), 1–72 (2009)
7. Hawkins, S., He, H., Williams, G., Baxter, R.: Outlier detection using replicator neural networks. In: Proceedings of the Fifth International Conference and Data Warehousing and Knowledge Discovery (DaWaK 02. 2002), pp. 170–180 (2002)
8. Barbara, D., Wu, N., Jajodia, S.: Detecting novel network intrusions using Bayes estimators. In: Proc. SIAM Intl. Conf. Data Mining. (2001)
9. Ratsch, G., Mika, S., Scholkopf, B., Muller, K.: Constructing boosting algorithms from SVMs: an application to one-class classification. IEEE Trans. Pattern Anal. Mach. Intell. **24** (9), 1184–1199 (2002)
10. Fan, W., Miller, M., Stolfo, S., et al.: Using artificial anomalies to detect unknown and known network intrusions. Knowl. Inf. Syst. **6**(5), 507–527 (2004)

11. Lin, J., Keogh, E., Lonardi, S., Chiu, B.: A symbolic representation of time series, with implications for streaming algorithms. In: Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, pp. 2–11 (2003)
12. Arning, A., Agrawal, R., Raghavan, P.: A linear method for deviation detection in large databases. In: Proceedings of KDD'1996, pp. 164–169 (1996)
13. Antipov, S., Fomina, M.: Problem of anomalies detection in time series sets. J. Prog. Prod. Syst. **2**, 78–82 (2012). (in Russian)
14. Saito, N.: Local feature extraction and its application using a library of bases. Ph.D. thesis, Yale University (1994)
15. Pham, D.T., Chan, A.B.: Control chart pattern recognition using a new type of self organizing neural network. In: Proceedings of the Institution of Mechanical Engineers, vol. 212(1), pp. 115–127 (1998)

Web-Based System for Enterprise Performance Analysis on the Basis of Time Series Data Mining

I.G. Perfilieva, N.G. Yarushkina, T.V. Afanasieva and A.A. Romanov

Abstract The development strategy of a complex enterprise depends on the financial performance evaluation. The latter can be made by time series analysis and forecasting of financial indicators. The paper is focused on a web-based system, which integrates fuzzy time series forecasting techniques with a web service to support enterprise managers in making plan for next period. In the paper the web service with three main components proposed: time series forecasting system using F-transform and fuzzy tendency forecasting techniques, system for linguistic description of predicted fuzzy tendency of each financial indicators and expert system for deriving the linguistic summarization of enterprise financial state and recommendation for planning for next period. As the proposed web service was successfully examined by ten Russian enterprises the example of financial performance evaluation was showed.

Keywords Fuzzy time series · Data mining · F-transform · Enterprise performance evaluation · Fuzzy tendency · Web service · Forecasting

The authors acknowledge that this paper was partially supported by the project no. 2014/232 of the Ministry of Education and Science of Russian Federation "Development of New Approach to the Intellectual Analysis of Information Resources" and by the project no. 16-07-00535 "Development and research of data mining algorithms for organizational and technical systems based on fuzzy models" of the Russian Foundation of Basic Research.

I.G. Perfilieva
Division OU, Centre of Excellence IT4Innovations, Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, 70103 Ostrava, Czech Republic
e-mail: irina.perfilieva@osu.cz

N.G. Yarushkina (✉) · T.V. Afanasieva · A.A. Romanov
Information Systems Department, Ulyanovsk State Technical University, Severny Venec, 32, 432027 Ulyanovsk, Russia
e-mail: jng@ulstu.ru

T.V. Afanasieva
e-mail: tv.afanaseva@mail.ru

A.A. Romanov
e-mail: romanov73@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 450, DOI 10.1007/978-3-319-33609-1_7

1 Introduction

Among the new tasks that have appeared in this field, an enterprise performance evaluation of economic indicators and finance ratios, organized in time series play a leading role. The past decade has witnessed an obvious proliferation of software tools available for business managers and assessors such as knowledge bases, expert systems, forecasting analysis tools, decision-making and support systems, cost estimation tools and others. This increased complexity of the whole process. Local and particular analyses do not increase the efficiency of the business; actually, the opposite effect occurs without predictive analytics, based on data mining. Moreover, various types of data: precise and vague, short-term and medium-term time series, qualitative and quantitative are usually involved. All these arguments put obstacles in the process of placing e-Business on a stable footing. Many ratios are usually used for defining financial and economic performance of the enterprises, such as liquidity ratios, current assets to equity ratio, current assets coverage, autonomy, debt load, etc. Correct prediction of above-mentioned ratios on the base of time series is very important for the enterprise because it confirms its credit worthiness and efficiency. As a result, e-Business managers, auditors, performance assessors and consulting experts are now keenly looking for software tools that are easy to use and can integrate various data mining techniques to obtain predictive image of the enterprise in next period. The solution can be based on an web service that combines various approaches, tools and technologies within a unified web based assessment environment. There are various models of vagueness that are used for the time series data mining with vague observations. Depending on a chosen model, the following techniques were proposed for time series analysis and forecasting: statistical [1, 2], fuzzy IF–THEN rules [3–9], decision rules over rough sets [10, 11], measuring of fuzzy-roughness in order to establish a similarity between a part and the whole of a time series [10]. This branch is developed in the works of a number of scientists, including the considerable works of Song and Chissom [4, 5], Novak et al. [6], Perfilieva [7, 8, 12], and Afanasieva et al. [13]. In 1993, Song and Chissom [4] proposed models of stationary and non-stationary first-order fuzzy time series. They applied these models to predict the number of Alabama University students being registered. Crisp time series were fuzzified by experts. In 2004, Afanasieva et al. [13] defined the concept of fuzzy tendency for fuzzy time series and determined research tasks. In 2006, Perfilieva [7] applied the fuzzy modeling of time series trends based on the F-transform. The structure of this contribution is as follows: fist we briefly describe the steps for the predictive analysis of enterprise performance evaluation; then time series decomposition by F-transform and time series modeling and fuzzy tendency forecasting are proposed as the tools for enterprise performance evaluation on the base of financial indicators; the last part shows the structure of an Web-based data mining system for managers and presents the test result.

2 Enterprise Performance Assessment

The authors developed the web service for the predictive analytics of financial indicators based on time series fuzzy modeling and fuzzy tendency forecasting (http://tsas.ulstu.ru). It uses numerical data of public reports and outputs recommendations in a linguistic form. The input data for analysis are public financial statements (balance sheets and income statements) arranged into groups: liquidity and solvency ratios, profitability ratios, and activity ratios. It is assumed that analyzed statements are available for the defined period, which can be expressed as a time series—a sequence of values stated at equally spaced time moments (the total number of values must be more than seven).

The source data are crisp and processed into fuzzy (qualitative) conclusions, forecasts and recommendations. We copy the structure of the real-world expert conclusions that are up to 80 % comprised of qualitative judgments or estimates of data trends, and the rest of recommendations. The workflow of the Internet service is presented in Table 1.

Table 1 The data mining procedure in steps

Step content	The question to be answered
Step 1. Transformation of the crisp time series into the fuzzy time series (FTS)—Fuzzification	Find parameters that characterize a current state of the estimated object (enterprise)
Step 2. Estimation of the FTS trends	Extract trends of the time series of chosen parameters
Step 3. Choose a structure of an expert summary of a state of the estimated object (enterprise)	In which terms the object can be characterized?
Step 4. Forecasting trends of the chosen time series	In which linguistic terms the forecasted trend of parameters can be expressed?
Step 5. Planning a future	Which recommendations can be made in order to improve the state of the estimated object (enterprise)?

3 Time Series Analysis by the F-Transform

As follows from Table 1, we will be focused on time series of financial indicators. To be realistic, we assume that the analyzed time series are non-stationary and can be decomposed as follows:

$$x_t = f(t) + y_t. \tag{1}$$

In (1), $f(t)$ is a deterministic part, which is usually called a *trend*, and y_t is a random part, which is additionally assumed to be stationary with a zero mean value and a constant variance.

A general model of a stationary time series y_t can be represented in the form of auto-regression [2]:

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \varepsilon_t, \tag{2}$$

where the coefficients α_i satisfy additional requirements and ε_t is a white noise. Expression (2) is the simplest representation of the time series y_t, which can be used for its forecast.

We assume that a time series $\{x_t,\ t \in 1, \ldots, T\}$ can be decomposed in accordance with (1). Therefore, our first task is to extract its trend. For this purpose, we propose the technique called the F-transform ("F" stands for fuzzy) because it has many successful applications in data and image processing [6–8].

Generally, the (discrete) F-transform of a function $f: P \to \mathbb{R}$ is a vector in which the components can be considered weighted local mean values of f. Throughout this paper, $\mathbb{R}$ denotes the set of real numbers, $[a, b] \subseteq \mathbb{R}$ and $P = \{p_1, \ldots, p_l\}$, $n < l$, denotes a finite set of points such that $P \subseteq [a, b]$. A function $f: P \to \mathbb{R}$, defined on the set P, is called discrete.

The first step in defining the F-transform of f is selecting a *fuzzy partition* of the interval $[a, b]$ using a finite number $n \geq 3$ of fuzzy sets $A_1, \ldots, A_n$. In [7], we used five axioms to characterize a fuzzy partition. In [9], the number of axioms was reduced to four, and a fuzzy partition was called relaxed. Below, we repeat the last definition.

Definition 1 [12]. Let $[a, b]$ be an interval on $\mathbb{R}$, $n \geq 3$, and $x_0, x_1, \ldots, x_n, x_{n+1}$ nodes such that $a = x_0 \leq x_1 < \cdots < x_n \leq x_{n+1} = b$. Let $P \subseteq [a, b]$ be a finite set of points such that $P = \{p_1, \ldots, p_l\}$, $l > n + 2$. We say that the fuzzy sets $A_1, \ldots, A_n: [a, b] \to [0, 1]$, identified with their membership functions, constitute a *fuzzy partition* of both sets $[a, b]$ and P, if the following conditions are satisfied:

(1) (*locality*)—for every $k = 1, \ldots, n,\ A_k(x) = 0$ *if* $x \in [a, b] \setminus (x_{k-1}, x_{k+1})$;
(2) (*continuity*)—for every $k = 1, \ldots, n,\ A_k$ *is continuous on* $[x_{k-1}, x_{k+1}]$;
(3) (*density*)—for every $k = 1, \ldots, n,\ \sum_{j=1}^{l} A_k(p_j) > 0$;

(4) (*covering*)—for every $j = 1, \ldots, l, \sum_{k=1}^{n} A_k(p_j) > 0$.

A fuzzy partition is called *uniform* if the fuzzy set A_1 is symmetrical (with respect to the axis $x = x_1$) and $A_2, \ldots, A_n$, are shifted copies of A_1, i.e. $A_k = A_1(x - x_k)$, $k = 2, \ldots, n$ (the associated details can be found in [9]. The membership functions $A_1, \ldots, A_n$, in the fuzzy partition are called basic functions. We say that the basic function A_k *covers* a point p_j if $A_k(p_j) > 0$.

The formal expression of a triangular shaped A_k, $k = 2, \ldots, n-1$, is given below:

$$A_k(x) = \begin{cases} 1 - \frac{|x - x_k|}{h}, & x \in [x_{k-1}, x_{k+1}], \\ 0, & otherwise. \end{cases} \tag{3}$$

In what follows, we fix the interval $[a, b]$, finite set of points $P \subseteq [a, b]$ and fuzzy partition $A_1, \ldots, A_n$ of $[a, b]$. We define [7] the (direct) *F-transform* of a discrete function $f: P \to \mathbb{R}$ as the vector $F_n(f) = (F_1, \ldots, F_n)^T$, where the k-th *component* F_k equals

$$F_k = \frac{\sum_{j=1}^{l} f(p_j) \cdot \leq A_k(p_j)}{\sum_{j=1}^{l} A_k(p_j)}, \; k = 1, \ldots, n. \tag{4}$$

To stress that the *F-transform components* $F_1, \ldots, F_n$, depend on $A_1, \ldots, A_n$, we say that the F-transform is taken with respect to $A_1, \ldots, A_n$.

It is easy to see that, when the basic functions $A_1, \ldots, A_n$, are fixed, the trend f in the representation (1) is determined by the F-transform components $F_1, \ldots, F_n$. The sequence $\{F_k, k \in 1, \ldots, n\}$ is considered to be a new time series with observations F_k. We call it a *time series with F-transform components* and use it to forecast the original time series $\{x_t, t \in 1, \ldots, T\}$. Below, we see that, to forecast $\{x_t, t \in 1, \ldots, T\}$, it is sufficient to forecast two corresponding time series: the series with F-transform components and the other series that has residual vectors.

The forecast of the time series with the F-transform components is based on the assumption that the series $\{F_k, k \in 1, \ldots, n\}$ is autoregressive and thus obeys

$$F_k = G(F_{k-1}, \ldots, F_{k-p}), \; k = p + 1, \ldots, n, \tag{5}$$

where p is the order of regression and $G: \mathbb{R}_p \to \mathbb{R}$ is some function. Three models for G are used in our experiment: linear autoregression

$$F_k = \alpha_1 F_{k-1} + \cdots + \alpha_p F_{k-p}, \tag{6}$$

an artificial neural network and a fuzzy relation model. The first two models are not discussed in our paper, because they are well known and described in the literature.

4 Time Series Decomposition

In this section, we show how a time series can be decomposed into two new series: one series with the F-transform components and the other series with residual vectors. This decomposition is further used for forecasting.

Assume that $\{x_t,\ t \in 1, \ldots, T\},\ T \geq 3$, is a time series in which the observations x_t are real numbers. We consider x_t to be a value of the discrete function x, which is defined on the set $P_T = \{1, \ldots, T\}$ of time moments. In what follows, we do not distinguish between the function x and the time series $\{x_t,\ t \in 1, \ldots, T\}$, and we use the latter in both meanings. Let $A_1, \ldots, A_n,\ 3 \leq n < T$, be basic functions that constitute a fuzzy partition of the interval $[1,\ T]$. Denote $P_k,\ k = 1, \ldots, n$, the subset of P_T that consists of points covered by A_k. Because of the *density* condition, P_k is not empty, and because of the *covering* condition,

$$\bigcup_{k=1}^{n} P_k = P_T.$$

Let vector $(X_1, \ldots, X_n)$ be the F-transform of a time series $\{x_t,\ t \in 1, \ldots, T\}$ with respect to $A_1, \ldots, A_n$. We say that $r_k = \{x_t - X_k \mid t \in P_k\}$ is the $|P_k|$-dimensional *residual vector* of $\{x_t,\ t \in 1, T.\}$ with respect to $A_k,\ k = 1, \ldots, n$. All vectors $r_k,\ k = 1, \ldots, n$, have the same dimension, provided that fuzzy partition $A_1, \ldots, A_n$. is uniform.

Let us extend r_k to the full dimensional vector $\tilde{r}_k = (\tilde{r}_{1k}, \ldots \tilde{r}_{Tk})$ by

$$\tilde{r}_{tk} = \begin{cases} x_t - X_k, & \textit{if } t \in P_k \\ -\infty, & \textit{otherwise} \end{cases}.$$

By the following proposition [12], we know that the original time series can be reconstructed from the vector of its F-transform components and the series $\{\tilde{r}_k,\ k = 1, \ldots, n\}$ of extended residual vectors.

Proposition 1 *Let* $\{x_t,\ t \in 1, \ldots, T\}$ *be a time series and* $\mathrm{A}_1, \ldots, \mathrm{A}_\mathrm{n}$, *be a fuzzy partition of* $[1, T]$. *Assume that* $(X_1, \ldots, X_n)$ *is the F-transform of* $\{x_t,\ t \in 1, \ldots, T\}$ *with respect to* $A_1, \ldots, A_n$, *and* $\{\tilde{r}_k,\ k = 1, \ldots, n\}$ *are extended residual vectors. Then, every element* $x_t,\ t = 1, \ldots, T$, *can be represented as follows:*

$$x_t = \bigvee_{k=1}^{n} (X_k + \tilde{r}_{tk}), \tag{7}$$

where $\vee$ *denotes the operation of maximum.*

5 Fuzzy Tendencies Forecasting

In this section, a new method for the time series modeling and forecasting is introduced. The method is based on the notion of the *fuzzy tendency* of a time series [13].

Assume that $\{x_t,\ t \in 1, \ldots, T\},\ T \geq 3$, is a time series that has real observations x_t and that the set of these observations is inside an interval $[c, d] \in \mathbb{R}$. We adopt the approach of Song and Chissom [4] and associate with the time series $\{x_t,\ t \in 1, \ldots, T\}$ a dynamic process with fuzzy values. In practice, this procedure means that we use partition of $[c, d]$ using fuzzy sets $Y_1, \ldots, Y_l$ (Definition 1), with the result that the partition is uniform and $Y_1, \ldots, Y_l$ have triangular shapes (3). Moreover, the distance between any two neighboring nodes of the partition is chosen in accordance with the required accuracy, e.g., $\delta > 0$. With each observation x_t, we associate the fuzzy set Y_{it} such that

$$Y_{it}(x_t) = max\{Y_i(x_t) \mid Y_i(x_t) > 0,\ i = 1, \ldots, k\}. \tag{8}$$

In a certain sense, Y_{it} gives the best characterization of x_t. Thus, the initial time series $\{x_t,\ t \in 1, \ldots, T\}$ can be replaced by the time series $\{Y_{it},\ t \in 1, \ldots, T\}$ with fuzzy observations.

If $\{Y_{it},\ t \in 1, \ldots, T\}$ is *time-invariant* then it can be represented in the form of autoregression, as shown in [4]. Below, we propose weaker conditions of the similar claim. They are based on another type of decomposition of $\{Y_{it},\ t \in 1, \ldots, T\}$ compared to that in [4]. Let us explain the details.

Let x_{t-1}, x_t be two neighboring values of the time series $\{x_t,\ t \in 1, \ldots, T\}$, and let Y_{it-1} and Y_{it} be the respective fuzzy sets that satisfy (9). We denote $d_t = |i_t - i_{t-1}|$ and $v_t = sgn(i_t - i_{t-1})$, where

$$sgn(x) = \begin{cases} -1, & \text{if } x < 0, \\ 0, & \text{if } x = 0, \\ +1, & \text{if } x > 0. \end{cases}$$

Then, $\{d_t,\ t \in 2, \ldots, T\}$ and $\{v_t,\ t \in 2, \ldots, T\}$ are two new time series that correspond to $\{Y_{it},\ t \in 1, \ldots, T\}$. Their values are inside the respective intervals $[0,\ l-1]$ and $[-1, 1]$ of real numbers. Let the interval $[0, l-1]$ be partitioned by fuzzy sets $D_1, \ldots, D_m,\ m \geq 2$, and let the interval $[-1, 1]$ be partitioned by fuzzy sets *V*1, *V*2, *V*3, where *V*1 represents "increase", *V*2 represents "stabilization" and *V*3 represents "decrease".

Similar to (8), the values d_t and v_t uniquely determine the respective fuzzy sets D_t and V_t, $t = 2, \ldots, T$, where

$$D_t(d_t) = max\{D_i(d_t) \mid D_i(d_t) > 0,\ i = 1, \ldots, m\}, \tag{9}$$

$$V_t(v_t) = max\{V_i(v_t) \mid V_i(v_t) > 0,\ i = 1, 2, 3\}, \tag{10}$$

such that the fuzzy time series $\{D_t,\ t \in 2, \ldots, T\}$ and $\{V_t,\ t \in 2, \ldots, T\}$ are associated with the original time series $\{X_t,\ t \in 1, \ldots, T\}$.

The fuzzy set V_t characterizes a "type" of dynamic behavior for x_t at the moment $t-1$ in terms of the "increase", "decrease", and "stabilization". The fuzzy set D_t characterizes the "intensity" of the dynamic behavior of x_t at the moment $t-1$, which is measured using fuzzy numbers. The pair (V_t, D_t) characterizes a "fuzzy tendency" of a time series at the moment t [13].

Below, we give a description of the proposed technique (see also [12]), which focuses on the representation of autoregression models for $\{D_t,\ t \in 2, \ldots, T\}$ and $\{V_t,\ t \in 2, \ldots, T\}$, in the form of first-order fuzzy relational models. We start with

$$D_t = D_{t-1} \circ R_D(t-1, t), \tag{11}$$

$$V_t = V_{t-1} \circ R_V(t-1, t), \tag{12}$$

where $R_D(t-1, t)$ and $R_V(t-1, t)$ are respective (unknown) fuzzy relations on $[0, l-1]$ and $\{V_t,\ t \in 2, \ldots, T\}$ belongs to some fixed collection of fuzzy sets (that does not depend on t). Then the fuzzy time series $\{D_t,\ t \in 2, \ldots, T\}$ $(\{V_t,\ t \in 2, \ldots, T\})$ is time invariant and fulfills $R_D(t-1, t) = R_D(R_V(t-1, t) = R_V)$. In Proposition 2, we show conditions that guarantee that both series $\{D_t,\ t \in 2, \ldots, T\}$ and $\{V_t,\ t \in 2, \ldots, T\}$ are time-invariant (see [12]). As a consequence, we represent unknown fuzzy relations $R_D(t-1, t)$ and $R_V(t-1, t)$ by (13) and (14).

Proposition 2 *Assume that* $\{x_t,\ t \in 1, \ldots, T\} \subseteq [c, d]$ *is a time series such that its first-order differences are bounded, i.e. there exists* $M > 0$ *such that, for all* $t = 2, \ldots, T,\ |x_t - x_{t-1}| \leq M$.

Then there exists partition $Y_1, \ldots, Y_l$ *of* $[c, d]$ *such that the corresponding fuzzy time series* $\{D_t,\ t \in 2, \ldots, T\}$ *and* $\{V_t,\ t \in 2, \ldots, T\}$ *are time-invariant.*

According to Proposition 2, the fuzzy relations R_D and R_V in (12) and (13) can be taken as

$$R_D = \bigvee_{t=2}^{T} (D_{t-1} \wedge D_t), \tag{13}$$

$$R_V = \bigvee_{t=2}^{T} (V_{t-1} \wedge V_t), \tag{14}$$

where $\wedge$ denotes the min operation. It is easy to see that, once we know the fuzzy sets D_{t-1} and V_{t-1} and the fuzzy relations R_D and R_V, we can apply (11) and (12) and obtain the fuzzy sets D_t and V_t for any $t = 2, \ldots, T$. To conclude, the original time series $\{x_t,\ t \in 1, \ldots, T\}$ satisfies the following autoregression scheme:

$$x_t = Def(x_t - 1) + Def(V_t) \cdot Def(D_t), \tag{15}$$

where $Def(\cdot)$ is the defuzzified value of a respective fuzzy set. In our approach, we use the center of gravity defuzzificaton, i.e. for an abstract fuzzy set A on a finite set X, the defuzzified value $Def(A) \in X$ is

$$Def(A) = \frac{\sum_{x \in X} xA(x)}{\sum_{x \in X} A(x)}.$$

Below, the scheme (15) is used to forecast a time series [12].

6 Structure of Internet Service for Enterprise Performance Assessment

The development of a new software product—an Internet service for data mining of financial indicators—employed the system approach and the component-based architecture. The developed Internet service is a Web-based decision support system for managers, consisting of the following structural components: a control system, two specialized Internet services (or web services) and a database. From the viewpoint of a user, the developed Internet service is a multipage web site with authorized access.

The Internet service is a component-based system whose main components are:

1. Expert System (SES) containing:
   - Fuzzy Logic Inference System (FLIS).
   - Relational Data Server (RDS).
2. Time series Forecasting system using F-transform and fuzzy tendency forecasting techniques.
3. Visualization systems for time series and linguistic summarizations.

One of the fundamental features of the system is that any experienced consulting expert or auditing firm can develop and propose their own components. Any component is viewed as a wrapped tool and other components exclusively through its inputs/outputs (formal procedure parameters). Thus, Internet service is an extendable system capable of self-organization and enlargement of its functions. Its component-based nature is perfectly suited for such an assessing session in which multiple independent components need to be utilized in various ways throughout an expert diagnosis procedure.

The expert diagnosis procedure is realized by the Expert System (SES), the intellectual heart of the system, while the others are realized as traditional data server and web-server.

The expert system is a fuzzy expert system possessing the following features:

- The SES uses statistical data interpreting them as input for rules based inference.
- The SES represents knowledge in the form of linguistic variables (or membership functions), fuzzy rules.
- The SES operates as a system of interacting components.

The FLIS is implemented as fuzzy rules base inference system to predict behaviour of a time series and by virtue of this ability, to forecast the future state of object. Expert system combines the traditional inference mechanism and time series forecasting. The RDS has been developed as a set of triggers and procedures for data server to save membership functions, linguistic estimates and the data base containing fuzzy rules, crisp time series and fuzzy time series, and intermediate assessment results together with estimated states and trends.

Experience lead us to the conclusion that although computer programs are developed by skilled programmers, a good software can be developed only with the help of consulting experts from an application area. With their help the following options: “Generating linguistic estimates of the basic financial parameter values”, “Evaluation of linguistic estimates of the analytic parameter values” and “Generating a linguistic description of data trend diagrams and their analysis” were

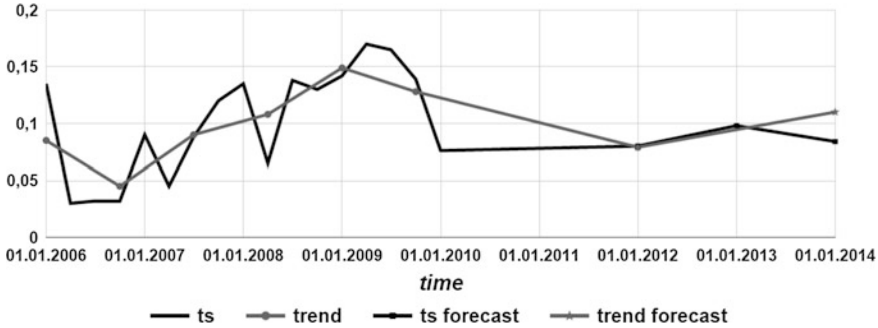
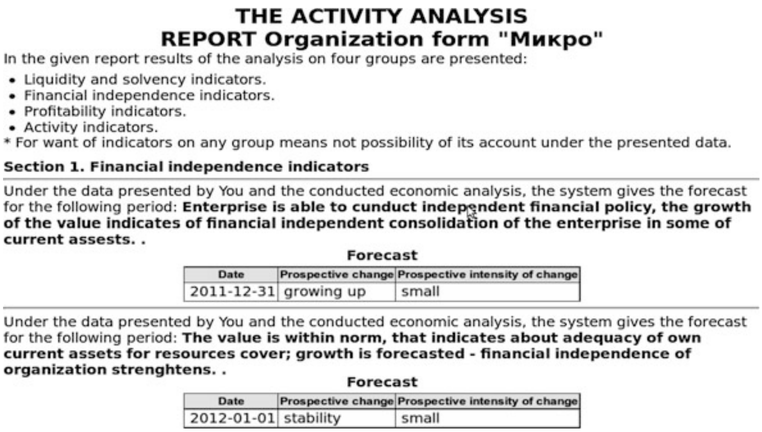


Fig. 1 Fuzzy tendencies forecasting of the indicator ‘Product Profitability’ (19 points)



THE ACTIVITY ANALYSIS
REPORT Organization form "Микро"

In the given report results of the analysis on four groups are presented:

- Liquidity and solvency indicators.
- Financial independence indicators.
- Profitability indicators.
- Activity indicators.

* For want of indicators on any group means not possibility of its account under the presented data.

Section 1. Financial independence indicators

Under the data presented by You and the conducted economic analysis, the system gives the forecast for the following period: **Enterprise is able to cunduct independent financial policy, the growth of the value indicates of financial independent consolidation of the enterprise in some of current assests. .**

Forecast

Date	Prospective change	Prospective intensity of change
2011-12-31	growing up	small

Under the data presented by You and the conducted economic analysis, the system gives the forecast for the following period: **The value is within norm, that indicates about adequacy of own current assets for resources cover; growth is forecasted - financial independence of organization strenghtens. .**

Forecast

Date	Prospective change	Prospective intensity of change
2012-01-01	stability	small

Fig. 2 Linguistic summarization of enterprise performance on the base of forecasted financial indicators

developed. We used two types of testing data: abstract and real economic time series. The option "Visualization of the analytical parameter diagrams" is very important for creating the understandable results to any managers.

The results of the data mining analytics via the web service are forecasted values and tendencies of financial indicators and linguistic summarizations of enterprise performance evaluation. These values are presented in a graphical form (see Fig. 1) and are explained in a natural language as recommendations (see Fig. 2). To illustrate, let us give a fragment of the produced results of data mining and forecasting of enterprise performance for the small Russian enterprise "Micro" (Figs. 1 and 2).

7 Conclusion

In this contribution, we describe a new web-based system that was elaborated using F-transform and fuzzy tendencies forecasting techniques as the main data mining tools. The results demonstrate that the fuzzy modeling of time series is efficient for economic non-stationary time series of short length (typically, from 7 to 40 points).

The important conclusion is that a complicated technology and architecture of a web-based system should not be visible to a user, so that he/she should benefit from its intelligence, linguistic summarization and graphical support. Moreover, the essential role in providing an expert knowledge support should be played by experts from an application domain.

References

1. Box, G., Jenkins, G.: Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco (1970)
2. Wold, H.: A Study in the Analysis of Stationary Time Series. Almqvist and Wiksel, Stockholm (1938)
3. Hwang, J.R., Chen, S.M., Lee, C.H.: Handling forecasting problems using fuzzy time series. Fuzzy Sets Syst. **100**, 217–228 (1998)
4. Song, Q., Chissom, B.: Fuzzy time series and its models. Fuzzy Sets Syst. **54**, 269–277 (1993)
5. Song, Q., Chissom, B.: Forecasting enrollments with fuzzy time series. Part I. Fuzzy Sets Syst. **54**, 1–9 (1993)
6. Novak, V., Stepnicka, M., Dvorak, A., Perfilieva, I., Pavliska, V.: Analysis of seasonal time series using fuzzy approach. Int. J. Gen Syst. **39**, 305–328 (2010)
7. Perflieva, I.: Fuzzy transforms: theory and applications. Fuzzy Sets Syst. **157**, 993–1023 (2006)
8. Perfilieva, I.: Fuzzy transforms: a challenge to conventional transforms. In: Hawkes, P.W. (ed.) Advances in Images and Electron Physics, 147, pp. 137–196. Elsevier Academic Press, San Diego (2006)
9. Perfilieva, I., Danková, M., Bede, B.: Towards a higher degree F-transform. Fuzzy Sets Syst. **180**, 3–19 (2011)
10. Sarkar, M.: Ruggedness measures of medical time series using fuzzy-rough sets and fractals. Pattern Recogn. Lett. Arch. **27**, 447–454 (2006)

11. Yao, J., Herbert, J.P.: Financial time-series analysis with rough sets. Appl. Soft Comput. **9**, 1000–1007 (2009)
12. Perfilieva, I., Yarushkina, N., Afanasieva, T., Romanov, A.: Time series analysis using soft computing methods. Int. J. Gen. Syst. **42**(6), 687–705 (2013)
13. Afanasieva, T., Yarushkina, N., et al.: Time series forecasting using fuzzy techniques. In: Proceedings of International Joint Conferece IFSA-EUSFLAT, pp. 1068–1075 (2015)

Feature Selection Using Semi Discrete Decomposition and Singular Value Decompositions

Intisar Hussien, Sara Omer, Nour E. Oweis and Václav Snášel

Abstract Nowadays, a large amount of digital data is available due to new technologies and different sources of data such as social networks, sensors, etc. There is a challenge to deal with this high dimensional data because query performance degrades as dimensionality increases. However, most of this data are redundant. Hence, it can be reduced to the smaller number of attributes without significant loss of information. The dimensionality reduction and feature selection techniques can be applied for that. In this paper, we compare two techniques Semi-Discrete Decomposition (SDD) and Singular Value Decomposition (SVD) to select significant features from Hepatitis dataset. We found that SVD is more appropriate than SDD in terms of accuracy and acceptable training time.

Keywords Dimensionality reduction · Feature selection · Semi-discrete decomposition · Singular value decomposition

1 Introduction

In recent years, data domain contains a huge amount of information and a large number of features "high dimensions" due to individualized intelligent systems [1]. Examples are multimedia images with thousands of pixels, documents that

I. Hussien
Arab Open University, Al-Ardiya, Kuwait
e-mail: iothman1@hotmail.com

S. Omer
Sudan University of Science and Technology, Khartoum, Sudan
e-mail: saraer.omer@gmail.com

N.E. Oweis (✉) · V. Snášel
Department of Computer Science, VSB-Technical University of Ostrava, Ostrava, Czech Republic
e-mail: oweis.nour@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 450, DOI 10.1007/978-3-319-33609-1_8

contain thousands of words, Genomics with thousands of genes. The analysis and query performance degrade rapidly with an increase in dimension. This leads to difficulty in interpretation and visualization of the data and also computation cost. There is a need to reduce dimensionality to be able to apply data mining algorithms to them and achieve fast computation and reduced storage.

Dimensionality reduction can be achieved using features extraction by transforming from feature space to lower dimension space. This leads to a generalization of feature selection so individual features are no longer recognizable [2]. Another possibility is to use features selection approach, which is done by finding a minimum set of features by looking at correlations between the different features and remove the redundant ones such that accuracy and performance of the resulting data are as close to the original data as possible.

Singular value decomposition was already used in dimensionality reduction. However, it is not suitable for large datasets due to the complexity and requires large storage [3]. The researchers proposed an alternative technique called Semi-Discrete Decomposition (SDD). It provides accurate approximation and less storage compared to other methods such as truncated Singular Value Decomposition (SVD).

In this experiment, a Multilayer Perceptron Neural Network has been chosen to be used for the classification part of the task. It is feed forward neural networks trained with the standard backpropagation algorithm. It is used to approximate any input-output map in difficult problems using attributes weight. This classifier implemented in Weka data mining tool.

The advantages of using Multi-Layer Perceptron Neural Network are easy to use, and they can approximate any input-output map with high accuracy. The key disadvantages are that they train slowly, and require lots of training data [4]. The dataset used in this experiment is hepatitis dataset from UCI Machine Learning Repository datasets.

The following diagram shows the general proposed approach for feature selection (Fig. 1):

This paper is organized as follows. Section 2 presents a discussion of different dimensionality reduction techniques. Section 3 revisits related work. Section 4 presents feature selection strategies in both SVD and SDD. The experimental details and results are given in Sect. 5 and finally, Sect. 6 provides our conclusion.

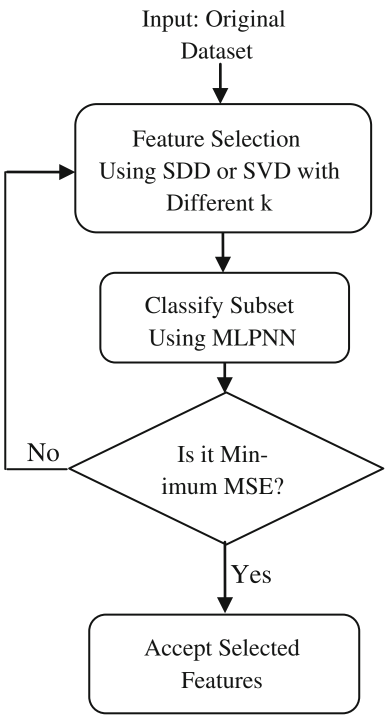


Fig. 1 Diagram shows general proposed approach

2 Dimensionality Reduction Techniques

2.1 Dimensionality Reduction and Feature Selection

Nowadays in a single experiment most of the time there are hundreds of attributes available. Dimensionality reduction is important to allow efficient use of this data because as dimensionality increase, query performance is decreased and computation time increase. Behavioral experiments suggest that humans do enhance query performance and minimize computational procedures by reducing features dimensionality [5]. The problem of finding features in a highly dimensional space is referred to as the dimensionality curse [6].

Dimensionality reduction can be defined as finding suitable low-dimensional space that represents original data. Mathematically dimensionality reduction can be defined as: for any vector $X = (x_1, x_2, \ldots, x_n)^T$ our goal is to find lower dimension $Y = (y_1, y_2, \ldots, y_k)$ where $k < n$ which represents original data.

There are two techniques for dimensionality reduction either by selecting a subset of original attributes or transform existing attributes to new reduced set [2]. The selection of a subset of features from original data is referred to as features selection. This leads to better performance by reducing computation cost, storage, and time.

2.2 Singular Value Decomposition

The singular value decomposition of matrix A with $m \times n$ rows and columns can be defined as follows:

$$A = U \sum V^T \tag{1}$$

where U is $m \times r$, V is $n \times r$ and $\sum$ is $r \times r$ diagonal matrix where the nonnegative elements in diagonal $\sum_{ii} = \sigma\, i$ are called singular values. The columns of U and V are called left and right singular vectors for A. Both SVD and QR factorization can be used to discover the rank of the matrix A.

The singular values decrease quickly so we can sort them in decrease order and take top k $(k < r)$ and use them instead of r in singular value decomposition.

$$A_k = U_k \sum_k V_k^T \tag{2}$$

where U_k is m x k, V_k^T is $k \times n$ and $\sum_k$ is $k \times k$ diagonal matrix. The A_k is the best representation for original data.

Singular value decomposition is not suitable in a large amount of data because it requires large storage and also when there are many updates it is expensive to recalculate SVD when every data is added [7].

2.3 Semi-discrete Decomposition

The semi-discrete decomposition works with large binary data such as Latent Semantic Indexing where there is a relation between objects or not. However it can be used for dimensionality reduction and feature extraction; it provides accurate approximation and less storage compared to other methods such as truncated Singular Value Decomposition.

Semi-Discrete Decomposition for matrix A with n rows and m columns can be used to compute reduced k-rank approximation as follows:

$$A \approx A_k = X_k D_k Y_k^T \tag{3}$$

where k is smaller than the original dimension of A and values of X_k and Y_k are in the set $\{-1, 0, 1\}$, and D_k is a diagonal matrix with positive coordinates. The SDD doesn't reproduce A exactly, even if $k = n$, however, It provides more accurate approximation and less storage of $k(m + n)$ compared to other methods [7].

3 Related Work

Kumar [8] their experiment compares unsupervised dimensionality reduction techniques. They analyze these techniques in term of complexity, approximation error, and retrieval quality. They use Medline, Cranfield, CACM and CISI datasets.

They concluded that singular value decomposition and Fuzzy K-Means Clustering produces better retrieval results than other Dimensionality Reduction techniques. However, Fuzzy K-Means Clustering with less complexity is a better option for deriving the semantic space.

Lin et al. [9] develop new dimensionality reduction technique called sparsified singular value decomposition (SSVD) to identify and remove nonessential features to reduce computation time for better performance. The generated set of features is called S2R.

Their experiment result shows that S2R procedure can reduce the computational cost in high dimensional datasets on feature search used in data mining process without compromise quality of original data. Although S2R is used for better feature selection performance, it could be extended to any data mining algorithm.

Aravindan et al. [10] their experiment develop a system, which automatically extracts the product features from the customers reviews and decide if they have been expressed in a positive or a negative way. Their technique works in two steps, feature extraction using association rule mining and polarity classification using Support Vector Machine.

Their results show that by using feature extraction and Support Vector Machine classifier achieves an accuracy of 79.67 %.

Rodríguez-González et al. [11] develop two algorithms, one for mining frequent similar patterns for similarity functions that fulfill the f-downward closure property and other for any similarity function. Their experiment results have shown that the two algorithms have better computation time and number of similar frequent patterns generated is better than other methods. However, they didn't show any experimental results to prove efficiency.

Wajid et al. [12] applied local Energy based Shape Histogram Feature Extraction Technique (LESH) on mammogram datasets. The extracted features were fed to support vector machine classifiers.

Their results show that selecting a subset of LESH features achieved a higher classification accuracy of 99.00 ± 0.50.

Ba-Alwi et al. [13] they compare different classification algorithms namely, neural network, FT Tree, Naive Bayes updatable, KStar, LMT, J48 and Naive Bayes for analyzing Hepatitis prognostic data. They applied rough set theory for features selection.

They concluded that the best accuracy is achieved by Naïve Bayes 96.52 % and the worst achieved by Neural Network 70.41 %.

4 Features Selection Strategy

4.1 Updated Singular Value Decomposition

The Truncated singular value decomposition for matrix A is decomposed as multiplication of three matrixes U_k, D_k and V_k^T where $(0 < k < n)$ see Fig. 2.

The Truncated SVD is used to reduce dimensionality by removing part of A which considered as noise in original data. Even with applying dimensionality reduction, storage cost may be large since U_k and V_k^T contain many small entries. It is possible to replace part of the small entries in U_k and V_k^T with zeros and remove them without significant effect on query accuracy performance on the dataset.

In this paper reduce dimensionality used to label features with small entries by applying threshold value T on U_k and V_k^T. The small entries in U_k and V_k^T are set to zero when they are smaller than the threshold value and treated as nonessential features, this will result in a sparse matrix.

There are many strategies used to generate sparse matrix using threshold namely single threshold strategy, Column Threshold Strategy, Exponential Threshold Strategy and L_1 regularization [14, 15].

The Exponential Threshold Strategy (ETS) will be used to find a smooth threshold function because it calculates threshold for each column in U_k and V_k, however, the other techniques apply one threshold value to all matrix. The ETS for matrix A with m rows and n columns can be calculated by the equation:

$$T_j = \frac{s}{m} \sum_{i=0}^{m} |a_{ij}| e^{(kj)^{-2}} \tag{4}$$

where ε is scalar factor and k is a number of singular values. The Eq. 1 computes threshold for each column in U_k and V_k^T and adjusted it by scalar ε. And for any u_{ij} in U_k, if $u_{ij} < T$ we set $u_{ij} = 0$ result in $\tilde{U}_k$. The same thing is applied to V_k^T to give $\tilde{V}_k^T$. The resulted sparse matrix will be:

$$\tilde{A}_k = \tilde{U}_k D_k \tilde{V}_k \tag{5}$$

The zero vectors in new matrix $\tilde{A}_k$ can be seen as nonessential features, and can be removed from original data to reduce dimensionality.

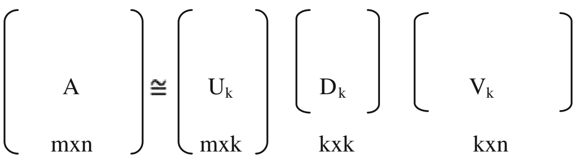


Fig. 2 Truncated singular value decomposition "rank-k"

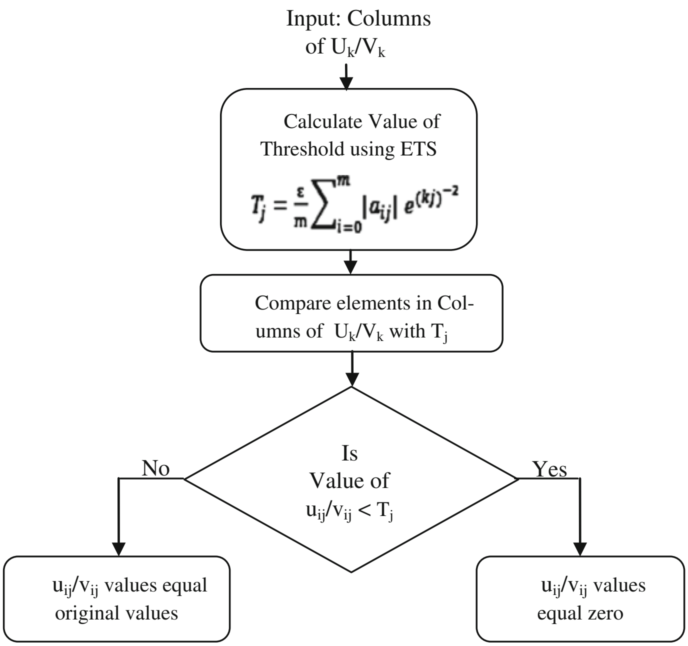


Fig. 3 Features selection strategy using truncated SVD and ETS

The following Fig. 3 shows features selection strategy using Truncated SVD and Exponential Threshold Strategy.

4.2 Semi-discrete Decomposition

Kolda and O'Leary [16] develop semi-discrete decomposition to reduce storage by using only three values {1, 0, −1} to define entries in X_k and Y_k^T. As shown in Fig. 4.

In SDD only A_k has been calculated which result in zero vectors assumed as nonessential features, those zero vectors are removed from original data to reduce

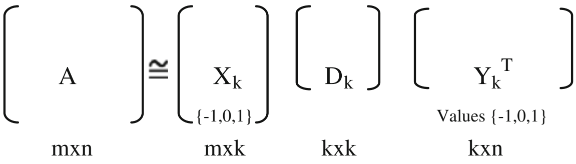


Fig. 4 Semi-discrete decomposition "Rank-k"

dimensionality and the remaining nonzero features selected as the new dataset features that were used in the training of multilayer perceptron Neural Network.

The evaluation parameters to compare results between applying SVD and SDD on original dataset and set of selected features using MLPNN on hepatitis dataset are accuracy in term of a number of correctly classified instances, reduction of storage and training time.

5 Experimental Results

5.1 Dataset and Preprocessing

For testing of semi-discrete decomposition and singular value decomposition as dimensionality reduction and features selection, we use the hepatitis dataset downloaded from UCI Machine Learning Repository [17]. The dataset total number of instances 155 and 20 attributes including Class (die and live), age, sex, steroid, antivirals, fatigue, malaise, anorexia, liver big, liver firm, spleen palpable, spiders, ascites, varices, bilirubin, alk phosphate, sgot, albumin, protime and histology.

Preprocessing stage applied on hepatitis dataset by cleaning it from missing values, noise "include errors, outlier/extreme values". Hepatitis dataset contains 167 missing values; they are replaced by the mean value. Then there is 28 outliers and 5 extreme values are removed. After preprocessing the remaining instances are 122 and 20 attributes.

5.2 Applying Semi-discrete Decomposition

First the semi-discrete decomposition is applied on the hepatitis dataset to produce three matrixes $X_{k,}$ $D_{k,}$ and Y^T_k with different values for k, according to the value of k some features value are set to zero. These zero vectors can be seen as nonessential features, it is difficult to identify them in original data.

Table 1 below shows selected features with different k values, reduction rate, accuracy (number of correctly classified instances) and time needed for training for original dataset and newly selected attributes set:

5.3 Applying Singular Value Decomposition

The truncated singular value decomposition is applied on hepatitis dataset which results in the three matrixes U_k, D_k and V^T_k. after that Exponential Threshold

Table 1 Semi-discrete decomposition experimental result

	Original dataset	Selected features
	K = 7	
No of attributes	20	4
Feature reduction rate	–	80 %
Training time with MLP in sec	0.66	0.1
Accuracy	86 %	83.6 %
Storage reduction rate	–	50 %
	Original dataset	Selected features
	K = 13	
No of attributes	20	5
Feature reduction rate	–	75 %
Training time with MLP	0.66	1.2
Accuracy	86 %	84.4 %
Storage reduction rate	–	37 %

Table 2 Singular value decomposition experimental result

	Original dataset	Selected features			
		K = 7	K = 10	K = 13	K = 16
No of attributes	19	18	15	18	18
Feature reduction rate	–	5 %	25 %	5 %	5 %
Training time—MLP	0.66	0.56	0.47	0.57	0.64
Accuracy	86 %	87 %	86.06 %	87.7 %	87.7 %
Storage reduction rate	–	0 %	0 %	0 %	0 %

Strategy are used to calculate the threshold for each column in U_k and V_k^T with $\varepsilon = 0.04$ for all experiments and different values for each k.

The elements u_{ij} and v_{ij} in each column is compared with column threshold if the value is less than column threshold the value of the element will be set to zero. The zero vectors are treated as a nonessential feature and removed from original data. The remaining set of significant features used in training.

Table 2 show selected features set from hepatitis dataset with different k using SVD, reduction rate, time taken for training using multi-layer perceptron and accuracy.

6 Discussion

From the previous experiments, the accuracy of original hepatitis dataset after preprocessing is 86 % which is good compared to other algorithms in the literature [13]. We tried to improve it using dimensionality reduction techniques.

First we apply semi-discrete decomposition with k = 7, 13 on the dataset. The number of selected features has slightly decreased from 19 to 4, 5 features respectively. Many important features are removed so the number of correctly classified instances (accuracy) is also decreased to 83 and 84. However, the multi-layer perceptron training time and storage reduction rate are significantly improved.

In the second experiment we apply the singular value decomposition with values for k (7, 10, 13, 16) on hepatitis dataset there is a few attributes are treated as nonessential and set to zero and in real they are not important to diagnose hepatitis like (malaise, anorexia, steroid and sgot) when we remove them the accuracy is stable and sometimes increase. However the training time has a slight decrease and there is no reduction in the storage as our dataset is small.

7 Conclusion

In this paper singular value decomposition and semi-discrete decomposition have been applied on hepatitis dataset for dimensionality reduction and features selection by finding nonessential features using Exponential Threshold Strategy and remove them from original data to improve performance. After that multi-layer perceptron for classification has been applied on selected features set. Finally a comparison between two techniques in term of training time, storage and accuracy have been done.

We conclude that in this experiment features selection using Singular Value Decomposition is more appropriate than Semi-Discrete Decomposition because it improves accuracy and has acceptable computation cost.

Acknowledgement This work is partially supported by Grant of SGS No. SP2016/97, VŠB—Technical University of Ostrava, Czech Republic and Arab Open University, Kuwait Branch.

References

1. Hamid, O.H., Alaiwy, F.H., Hussien, I.O.: Uncovering cognitive influences on individualized learning using a hidden Markov models framework. In: 2015 Global Summit on Computer and Information Technology (GSCIT). IEEE (2015)
2. Sorzano, C.O.S., Vargas, J., Pascual Montano, A.: A survey of dimensionality reduction techniques. arXiv preprint arXiv:1403.2877 (2014)
3. Li, M., et al.: Large-scale Nyström kernel matrix approximation using randomized SVD. IEEE Trans. Neural Netw. Learn. Syst. **26**(1), 152–164 (2015)
4. Ivancevic, V.G., Ivancevic, T.T.: Quantum Neural Computation, vol. 40. Springer Science & Business Media (2010)
5. Hamid, O.H., Wendemuth, A., Braun, J.: Temporal context and conditional associative learning. BMC Neurosci. **11**, 45 (2010)

6. Vervliet, N., et al.: Breaking the curse of dimensionality using decompositions of incomplete tensors: tensor-based scientific computing in big data analysis. IEEE Signal Process. Mag. **31** (5), 71–79, (2014)
7. Moravec, P., Snásel, V.: Dimension reduction methods for Iris recognition. DATESO (2009)
8. Kumar, A.C.: Analysis of unsupervised dimensionality reduction techniques. Comput. Sci. Inf. Syst. **6**(2), 217–227 (2009)
9. Lin, P., Zhang, J., An, R.: Data dimensionality reduction approach to improve feature selection performance using sparsified SVD. In: 2014 International Joint Conference on Neural Networks (IJCNN). IEEE (2014)
10. Aravindan, S., Ekbal, A.: Feature extraction and opinion mining in online product reviews. In: 2014 International Conference on Information Technology (ICIT). IEEE (2014)
11. Rodríguez-González, A.Y., et al.: Mining frequent patterns and association rules using similarities. Expert Syst. Appl. **40**(17), 6823–6836 (2013)
12. Wajid, S.K., Hussain, A.: Local energy-based shape histogram feature extraction technique for breast cancer diagnosis. Expert Syst. Appl. (2015)
13. Ba-Alwi, F.M., Hintaya, H.M.: Comparative study for analysis the prognostic in hepatitis data: data mining approach. Spinal Cord **11**, 12 (2013)
14. Gao, J., Zhang, J.: Sparsification strategies in latent semantic indexing. In: Proceedings of the 2003 Text Mining Workshop (2003)
15. Ng, A.Y.: Feature selection, L1 vs. L2 regularization, and rotational invariance. In: Proceedings of the Twenty-First International Conference on Machine Learning. ACM (2004)
16. Kolda, T.G., O'leary, D.P.: A semidiscrete matrix decomposition for latent semantic indexing information retrieval. ACM Trans. Inf. Syst. (TOIS) **16**(4), 322–346 (1998)
17. UCI Machine Learning Repository, https://archive.ics.uci.edu/ml/datasets/Hepatitis

Feature Selection by Principle Component Analysis for Mining Frequent Association Rules

Tayseer M.F. Taha, Eltayeb Shomo, Nour E. Oweis and Vaclav Snasel

Abstract Data mining techniques have been increasingly studied. Extracting the association rules have been the focus of this studies. Recently research have focused on association rules to help uncover relationships between seemingly unrelated data in a relational database or other information repository. The large size of data makes the extraction of association rules hard task. In this paper, we propose a new method for dimension reduction and feature selection based on the Principal Component Analysis, then find the association rules by using the FP-Growth Algorithm. Experimental results reveals that the reduction technique can discover the same rules obtained by the original data.

Keywords Association rules · Feature selection · Frequent pattern mining · Mining association rules · Principal component analysis

1 Introduction

Data mining is a process of extracting knowledge and pattern from large amount of data. Frequent itemset mining is an interesting branch of data mining that focuses on looking at sequences of actions or events [18]. The term Association Rules refers

T.M.F. Taha · E. Shomo
Department of Computer Sciences, Sudan University for Sciences and Technology, Khartoum, Sudan
e-mail: tayseertaha1@gmail.com

E. Shomo
e-mail: tayebshomo@gmail.com

T.M.F. Taha · E. Shomo · N.E. Oweis (✉) · V. Snasel
Faculty of Electrical Engineering and Computer Science,
Department of Computer Science, VSB-Technical University of Ostrava,
Ostrava, Czech Republic
e-mail: oweis.nour@gmail.com

V. Snasel
e-mail: vaclav.snasel@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 450, DOI 10.1007/978-3-319-33609-1_9

to “Finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transactional databases, relational databases, and other information repositories” [5]. Finding these frequent patterns is achieved by using the criteria support and confidence to identify the most important relationships.

However, data mining techniques have to deal with large number of features (data in high dimension) of any type of data, which cause what known as *curse of dimensionality* [20]. Removing any irrelevant or abundant features is key factor to more understandable model, easy visualization for the data as stated in [28].

Feature Selection and Dimension Reduction Methods are two approaches for reducing the number of features in dataset. Feature selection method select a subset of the attributes, while dimension reduction methods in general produce a linear combination of the attributes. A survey of various feature selection and dimension reduction methods can be found in [12, 17].

Association Rules mining plays an important role in Knowledge Discovery in Data. It has been applied in many applications such as marketing [30], fraud detection [27], diagnosing and analysis of diseases [8, 9] etc.

Matrix decomposition techniques as SVD and Non negative matrix factorization (NMF) for reducing dimensions for constructing reduced concept lattice, and for compression size of the formal context were studied in [23–26]. Finding the Association rules for reduced data by NMF and using Formal Concept Analysis (FCA) was addressed in [6, 7].

To the best of our knowledge, mining frequent association rules by the Principal Component Analysis dimension reduction technique have not discussed so far. In this paper we introduce a new feature selection method based on the Principal Component Analysis weights, within the proposed methods features with weights less than a certain pre-defined threshold are eliminated from the data and the FP-Growth algorithm is applied to the reduced data to extract the association rules.

The rest of this paper is organized as follows: Frequent itemsets, Association rules, SVD and PCA background in the next section. Related work of mining association rules SVD and PCA in Sect. 3. Our proposed approach, data collection, implementations described in Sect. 4. Evaluated the experiments and the results are in Sect. 5. Finally a recommendations, and a conclusion in Sect. 6.

2 Background

2.1 Frequent Pattern Mining

Pattern mining is an important data mining task, with the goal of discovering interesting, unexpected useful patterns in the databases. Pattern mining is applied in different types of data among these are transaction databases, sequence databases, streams, graphs, etc. In addition pattern mining algorithms can find different kinds of patterns: subgraph, lattices, sequential patterns, associations, indirect associations, rules, etc.

The interesting patterns are those which appears frequently in databases. The problem of finding frequent pattern mining can be stated as follows: *Given a database D with transactions $T_1, \ldots, T_N$ determine all the patterns P that are present in at least a fraction s of the transactions, s is known as the minimum support* [3]. In another way, each $T_i \in \mathcal{T}, \forall i = \{1, \ldots, n\}$ consists of a set of items $T_i = \{x_1, \ldots, x_n\}$. A set $P \subseteq T_i$ is called an itemset. The number of transactions containing P is referred to the *support* of P. The pattern P is frequent if it is support is at least equal to the some predefined minimum threshold σ [2].

One of the most popular algorithms for Frequent pattern mining is *Frequent Pattern Growth (FP-Growth)* [14] which does not generate candidates unlike the a priori [4], instead it generate FP-tree from the data set. The following steps explain the algorithm in short:

- Build a compact data structure called the FP-tree: By Scanning data and finding support for each item. Then it discard infrequent items and sort frequent items in a decreasing order based on their support.
- FP-Growth extract the frequent itemsets directly from the FP-tree.

The algorithm outputs all the frequent itemsets.

2.2 Association Rules Mining

The role of association rules is to identify the strong rules discovered in data base transaction by using different measures of interestingness [10, 19]. The generation of the rules takes two phases process, the first phase determine all the frequent patterns given the minimum support threshold. The second phase extracts all the rules from these patterns, these rules should satisfy minimum confidence threshold. Once these rules are discovered they are ranked based on some interestingness measure [21]. A rule is defined in the form $X \Rightarrow Y$, where X and Y are set of items and $X \cap Y = \phi$. These set items are called *antecedent* (Left Hand Side or LHS) and *consequence* (Right Hand Side or RHS) of the rule. FP-Growth is known algorithm for mining the association rules. The rule is considered to be interesting if it satisfies a user minimum support and user minimum confidence at the same time. Where the support of an itemset X is defined as the proportion of transactions in the data sets which contain the itemset.

$$Supp(X) = \frac{no.\ of\ transactions\ which\ contain\ the\ itemset\ X}{total\ no.\ of\ transactions} \tag{1}$$

The confidence of a rule is the ratio of the number of transactions that include all items in the Y, as well as the in the itemset X (the support) to the number of transactions that include all itemset in the X.

$$Conf(X \Rightarrow Y) = \frac{Supp(X \cup Y)}{Supp(X)} \tag{2}$$

There are too many association rules satisfying the support and the confidence measurements, a further interest measurement have been used to rank these rules which is *lift*. Is defined as s the deviation of the support of the whole rule from the support expected under independence given the supports of the LHS and the RHS.

$$lift(X \Rightarrow Y) = \frac{Supp(X \cup Y)}{Supp(Y) * Supp(X)} \tag{3}$$

2.3 Singular Value Decomposition (SVD)

Generally if $A$is the covariance matrix that is extracted from a given data set D. The elements of the covariance matrix are the covariance's between the random variables representing the data attributes (or features). The variance of the random variables lie in the diagonal of A and their sum is the total variability.

The SVD of matrix C as a product of three matrices: U, Σ, and V. Since matrix A is an $m \times m$ matrix, then the SVD components of matrix A are given by:

$$A = U\Sigma V \tag{4}$$

where U and V are orthogonal $m \times m$ matrices, and Σ is a $m \times m$ diagonal matrix. The diagonal elements of Σ are the eigenvalues of the covariance matrix A, ordered according to their magnitudes from bigger to smaller. The columns of U (or$V$) are the eigenvectors of matrix A that correspond the eigenvalues of A. Columns of U, defines the principal components, and reducing the dimension is done through selecting the first k principal components [29].

2.4 Principle Component Analysis (PCA)

The principal component analysis (PCA) is a multivariate statistical technique aiming at extracting the features that represent most of the information in the given data and eliminating the least features with least information. Therefore, the main purpose from using the principal component analysis is to reduce the dimensionality of the data without seriously affecting the structure of the data. The PCA works to replace the original random variables with other set of orthonormal set of vectors called the principal components. The first principal component is desired to pass closer to as much possible to data points, and the projection of the data into the space spanned by the first component is the best projection over spaces spanned by other vectors in one dimension. The second principal component is orthogonal to the first principal component and the plain spanned by it together with the first component is the closest plane to the data. The third, fourth, etc. components feature at the same logic. In the new space with the same dimension as the original space, the covariance matrix is a diagonal matrix, and preserves the total variability of the original covariance matrix. The PCA's are obtained by applying the SVD to the covariance matrix of the data attributes. Inan et al. [15] used feature selection method that combines the a priori algorithm and the PCA method with an artificial neural network classifier for detecting breast cancer. PCA based on association rules have been used as feature selection in the neuroimage field for early diagnosis of Alzaheimer's disease by Chaves et al. [9], then a classification was performed by Support Vector Machines.

3 Related Work

In this section we will discuss some recent research work for mining frequent association rules and the SVD as dimension reduction technique.

Das et al. [11] presented three techniques for dimension reduction by feature selection, dimension reduction based on frequency count and dimension reduction for association rules algorithms which used only for single objective association rules to a desired level and lead to a loss of some interesting rules. Thus a reduction of dimension based on multi-objective for discovering rules is proposed as well. The affect of matrix decomposition techniques SVD and Non negative Matrix Factorization (NMF) in reducing dimensions for constructing reduced concept lattice studied by Snasel et al. [23, 24]. In another work, the application of SVD and NMF methods for compression size of the formal context investigated by Snasel et al. [25] where it leads to a control of the size of implications. The SVD has been used as method of decomposition on web searching problems by Abdullah et al. [1]. The experiments show that it works as a good solution for search results clustering. Aswani et al. discussed mining association rule by applying FCA on reduced dimension based on SVD application of Mining association rues conducted in healthcare datasets [6]. Aswani proposed the NMF as dimension reduction method for finding the Association rules by using Formal Concept Analysis (FCA) [7]. An incremental model building technique for generation SVD based on folding-in for recommendation systems proposed by Sarwar et al. [22]. The folding-in technique requires less time and storage space but can result in the loss of quality due to the non-orthogonality of the incremental SVD space. Zhou et al. [31] initiated an incremental algorithm based on singular value decomposition (SVD), which combines the Incremental SVD algorithm with the Approximating the Singular Value Decomposition (ApproSVD) algorithm, called the Incremental ApproSVD. A comparison of the prediction accuracy and running time between the Incremental ApproSVD algorithm and the Incremental SVD algorithm on the MovieLens dataset and Flixster dataset was presented as well. A latent semantic analysis (LAS) via SVD reduction technique for mining rules is proposed in [16]. The proposed method of reduction association rule enhanced the speed and the reduction rate, also credibility and support. More over it finds out the relations between items set such as inverse and equivalence.

4 Methodology

In this section we will explain data collection, data conversion, dimension reduction and feature extraction by PCA, and the proposed algorithm.

4.1 Data Collections

We conduct the experiment on the Groceries data set of real-world point-of-sale transaction data from a typical local grocery outlet. The data set contains 9835 transactions of items with 169 categories [13]. Every transaction contains from one to 32 itemsets.

4.2 Data Conversion

From the original file we created a new datasets in which the attributes are the names of the products (169 itemsets), and the instances are the 9835 transactions, to a shape of a binary matrix D of 9835×169, such that:

$$d_{ij} = \begin{cases} 1 \text{ if itemset j appears in transaction i} \\ 0 \; otherwise \end{cases} \tag{5}$$

4.3 Dimension Reduction and Feature Selection by the PCA

The main purpose from using the principal component analysis is to reduce the high-dimensional data with a dimension m into a lower dimensional data of dimension k, where $k \leq m$. To determine the reduction size we use the attributes weights generated by the components created by the PCA. The attribute weights reflect the relevance of the attributes with respect to the class attribute. The higher the weight of an attribute, the more relevant it is considered. The Results of applying PCA gives a numerical floating numbers which are not compatible with the FP Growth algorithm. Therefore, it has been necessary to convert the data to binominal.

4.4 Proposed Algorithm

Our proposed algorithms consists of 5 phases: The data transformation phase transform the transaction Groceries dataset into binary file, in which each column corresponds to an attribute (itemset), and each row correspond to a transaction. The resulting dataset is 9835 × 169 binary matrix. In the second phase we pass the dataset to the weight by PCA operator which uses the PCA to assign weights to the attributes and sorts them according to their weights in an ascending order. The third phase is the feature selection (dimension reduction) based on the PCA weights. In this phase we chose a weight threshold such that each attribute with a weight less than the threshold is dropped from the dataset, and only the attributes with weights above the threshold are considered for the next phase. Then a new datasets is created subject to the selected attributes. The forth phase is conversion to binominal. In this phase the new datasets obtain from the lase phase is converted to binominal dataset by using the numeric to binominal operator. The fifth phase apply the FP-Growth algorithm to the new datasets to find the minimum frequent itemsets that satisfies the support parameter. Then the FP-Growth passes the output to the create association rule operator to extract the rules.

These five phases are explained in Fig. 1.

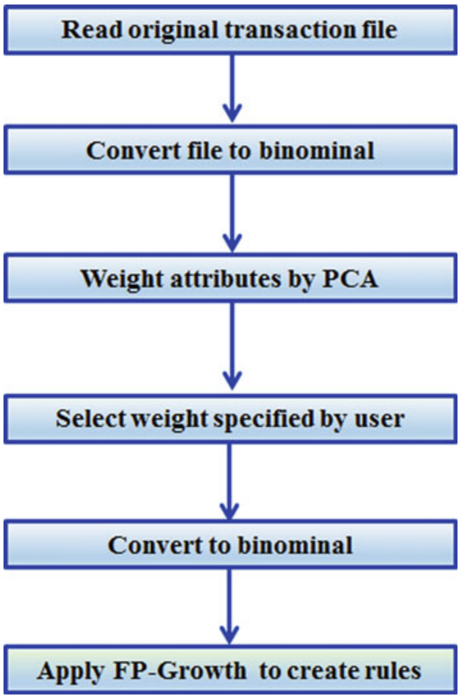


Fig. 1 Association rules extraction flow chart

4.5 Experiment

This section explains the experiment we applied on the Groceries data set to extract the association rule based on the PCA.

We ran the experiment in Rapidminer 6.5. We configured the Weight by PCA operator such that the outputs weights are normalized between [0, 1], and the attributes are sorted in an ascending order subject to their PCA weights. The attribute with the maximum weight was the whole milk with weight 1, followed by other vegetables, yoghurt, root vegetables as 0.75, 0.47, 0.42 respectively. The (sound storage medium) was the attribute with the minimum PCA weight valued at 0.

We adjusted the selected by weight operator to select only the attributes with PCA weights ≥ 0.01. This resulted in selecting 107 attributes(itemsets).

For mining the association rules we used the FP Growth operator with minimum support value set to 0.0018 to find the minimum frequent itemsets by building FP-tree, then we used the create association rules operator which takes these frequent itemsets and find the rules that satisfies the confidence parameter, we set it to 0.95.

5 Results

The following Fig. 2, Table 1 describes the Association Rules found by applying the FP-Growth algorithm, with confidence of 1.000 and lift value as 3.91 for the first four rules and 5.17 for the fifth rule.

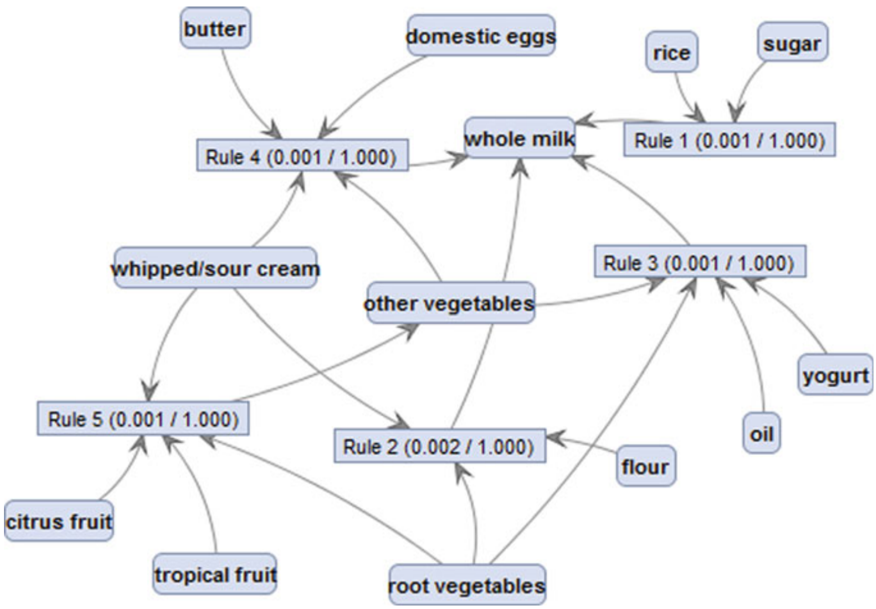


Fig. 2 Association rules extracted by PCA

Table 1 Association rules extracted by PCA

Consequence (LHS)	Antecedent (RHS)	Lift
Sugar, rice	Whole milk	3.91
Root vegetables, whipped/sour cream, flour	Whole milk	3.91
Other vegetables, yogurt, root vegetables, oil	Whole milk	3.91
Other vegetables, whipped/sour cream, domestic eggs, butter	Whole milk	3.91
Root vegetables, tropical fruit, citrus fruit, whipped/sour cream	Other vegetables	5.17

6 Conclusion

In real application that datasets contain of huge number of transactions and itemsets, which makes the extractions of the association rules a hard problem to solve, as a result to the high dimensionality of the data. Not all the transactions have the same level of importance. Therefore eliminating them might not affect the information gained from the transaction dataset.

The main goal from this paper was to introduce feature selection method based on the PCA weight. Where the attributes are ranked according to their weight and the attributes with weights less than a certain pre-determined weighting threshold are eliminated from the data, giving new data with less dimensions.

Our algorithm shows that with this data reduction technique one can obtain the same association rules that are obtained by the original data.

To the best of our knowledge the proposed technique is pioneering work that uses this approach in mining the association rules. In this paper we used the principle component analysis as dimension reduction and for feature extraction technique to beat the curse of dimensionality. The weight by PCA were used.The higher the weight of an attribute, the more relevant it is considered. The dimension was reduced without loosing the information. FP-Growth algorithm to find the frequent itemsets in order to discover the useful rules.

Acknowledgments This work is partially supported by Grant of SGS No. SP2016/97, VB—Technical University of Ostrava, Czech Republic.

References

1. Abdulla, H.D., Abdelrahman, A.S., Snášel, V., Aldosari, H.: Using singular value decomposition as a solution for search result clustering. In: Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014, pp. 53–61. Springer (2014)
2. Aggarwal, C.C., Bhuiyan, M., Al Hasan, M.: Frequent pattern mining algorithms: a survey. In: Frequent Pattern Mining, pp. 19–64 (2014)
3. Aggarwal, C.C., Han, J. (eds.): Frequent Pattern Mining. Springer (2014)
4. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. SIGMOD Rec. **22**(2), 207–216 (1993)
5. Amir, A., Feldman, R., Kashi, R.: A new and versatile method for association generation. Inf. Syst. **22**(6/7), 333–347 (1997)
6. Aswani Kumar, Ch., Srinivas, S.: Mining associations in health care data using formal concept analysis and singular value decomposition. J. Biol. Syst. **18**(04), 787–807 (2010)
7. Aswani Kumar, Ch.: Mining association rules using non-negative matrix factorization and formal concept analysis. In: Venugopal, K.R., Patnaik, L.M. (eds.) Computer Networks and Intelligent Computing. Communications in Computer and Information Science, vol. 157, pp. 31–39. Springer, Berlin (2011)
8. Chaves, R., Ramrez, J., Grriz, J.M.: Integrating discretization and association rule-based classification for alzheimers disease diagnosis. Expert Syst. Appl. **40**(5), 1571–1578 (2013)
9. Chaves, R., Ramírez, J., Górriz, J.M., Puntonet, C.G., et al.: Alzheimers disease, initiative, neuroimaging, association rule-based feature selection method for alzheimers disease diagnosis. Expert Syst. Appl. **39**(14), 11766–11774 (2012)
10. Chen, Y., Li, F., Fan, J.: Mining association rules in big data with NGEP. Cluster Comput. **18**(2), 577–585 (2015)
11. Das, S., Nath, B.: Dimesionality reduction using association rule mining. In: IEEE Region 10 and the Third International Conference on Industrial and Information Systems. ICIIS 2008, pp. 1–6, Dec 2008
12. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
13. Hahsler, M., Hornik, K., Reutterer, T.: Implications of probabilistic data modeling for mining association rules. In: Spiliopoulou, M., Kruse, R., Borgelt, C., Nrnberger, A., Gaul, W. (eds.) From Data and Information Analysis to Knowledge Engineering. Studies in Classification, Data Analysis, and Knowledge Organization, pp. 598–605. Springer, Berlin (2006)
14. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. SIGMOD Rec. **29**(2), 1–12 (2000)
15. Inan, O., Uzer, M.S., Yılmaz, N.: A new hybrid feature selection method based on association rules and PCA for detection of breast cancer. Int. J. Innov. Comput. Inf. Control **9**(2), 727–729 (2013)
16. Kim, K.T., Seol, W.S., Kim, U.M., Youn, H.Y.: Latent semantic analysis for mining rules in big data environment. In: 2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 193–200. IEEE (2014)
17. Molina, L.C., Belanche, L., Nebot, À.: Feature selection algorithms: a survey and experimental evaluation. In: Proceedings of the 2002 IEEE International Conference on Data Mining, ICDM '02, pp. 306–, Washington, DC, USA. IEEE Computer Society (2002)
18. Oweis, N.E., Owais, S.S., George, W., Suliman, M.G., Snášel, V.: A survey on big data, mining: (tools, techniques, applications and notable uses). In: Intelligent Data Analysis and Applications, pp. 109–119. Springer (2015)
19. Piatetsky-Shapiro, G.: Discovery, analysis and presentation of strong rules. In: Piatetsky-Shapiro, G., Frawley, W.J. (eds.) Knowledge Discovery in Databases, pp. 229–248. AAAI Press (1991)
20. Powell, W.B.: Approximate Dynamic Programming: Solving the Curses of Dimensionality (Wiley Series in Probability and Statistics). Wiley-Interscience (2007)

21. Soysal, Ö.M.: Association rule mining with mostly associated sequential patterns. Expert Syst. Appl. **42**(5), 2582–2592 (2015)
22. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Incremental singular value decomposition algorithms for highly scalable recommender systems. In: Proceedings of the 5th International Conference in Computers and Information Technology (2002)
23. Snasel, V., Abdulla, H.M.D., Polovincak, M.: Behavior of the concept lattice reduction to visualizing data after using matrix decompositions. In: 4th International Conference on Innovations in Information Technology. IIT '07, pp. 392–396, Nov 2007
24. Snasel, V., Gajdos, P., Dahwa Abdulla, H.M., Polovincak, M.: Using matrix decompositions in formal concept analysis. In: The 10th Intenational Connference of Information Systems Implementations and Modeling (2007)
25. Snasel, V., Polovincak, M., Dahwa Abdulla, H.M., Horak, Z.: On concept lattices and implication bases from reduced contexts. In: Eklund, P.W., Haemmerl, O. (eds.) ICCS Supplement, CEUR Workshop Proceedings, vol. 354, pp. 83–90. CEUR-WS.org (2008)
26. Snasel, V., Polovincak, M., Dahwa Abdulla, H.M., Horak, Z.: On knowledge structures reduction. In: Snsel, V., Abraham, A., Saeed, K., Pokorn, J. (eds.) CISIM, pp. 33–37. IEEE Computer Society (2008)
27. Snchez, D., Vila, M.A., Cerda, L., Serrano, J.M.: Association rules applied to credit card fraud detection. Expert Syst. Appl. **36**(2, Part 2), 3630–3640 (2009)
28. Tan, P.-N., Steinbach, M., Kumar, V.: Introduction to Data Mining, 1st edn. Addison-Wesley Longman Publishing Co., Inc., Boston (2005)
29. Thomasian, A.: Singular value decomposition, clustering, and indexing for similarity search for large data sets in high-dimensional spaces (2015)
30. Yunyan, L., Juan, C.: Application of association rules mining in marketing decision-making based on rough set. In: 2010 International Conference on E-Business and E-Government (ICEE), pp. 3749–3752, May 2010
31. Zhou, X., He, J., Huang, G., Zhang, Y.: SVD-based incremental approaches for recommender systems. J. Comput. Syst. Sci. **81**(4), 717–733 (2015)

Time Series Forecasting Based on Hybrid Neural Networks and Multiple Regression

Alexey Averkin, Sergey Yarushev, Igor Dolgy and Andrey Sukhanov

Abstract This paper presents research in the field of hybrid methods of time series forecasting, including a detailed review of the latest researches in the field of forecasting. The paper includes detailed review of studies what compared the performance of multiple regression methods and neural networks. It is also consider a hybrid method of time series prediction based on ANFIS. In addition, showed the results of time series forecasting based on ANFIS model and compared with results of forecasting based on multiple regression.

Keywords Time series forecasting · Neural networks · Regression models · ANFIS time series prediction

---

This work was supported by the Russian Foundation for Basic Research (Grant No. 14-07-00603).

A. Averkin (✉)
Dorodnicyn Computing Centre of RAS, Moscow, Russia
e-mail: averkin2003@inbox.ru; iiti16@rgups.ru

S. Yarushev
Dubna State University, Moscow Oblast, Russia
e-mail: sergey.yarushev@icloud.com

I. Dolgy · A. Sukhanov
Rostov State Transport University, Rostov-on-Don, Russia
e-mail: drewnia@rambler.ru

A. Sukhanov
VŠB—Technical University of Ostrava, Ostrava, Czech Republic
e-mail: andrey.sukhanov.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 450, DOI 10.1007/978-3-319-33609-1_10

1 Introduction

The recent surge in research of artificial neural networks (ANN) showed that neural networks have a strong capability in predicting and classification problems. ANN successfully used for various tasks in many areas of business, industry and science [1].

Such high interest in neural networks caused by the rapid growth in the number of articles published in scientific journals in various disciplines. It suffices to consider several large databases to understand the huge number of articles published during the year on the theme of the study of neural networks, it is thousands of articles.

A neural network is able to work parallel with input variables and consequently handle large sets of data quickly. The main advantage of neural networks is the ability to find patterns [2]. ANN is a promising alternative in the toolbox professionals involved in forecasting. In fact, the non-linear structure of the neural networks is partially useful to identify complex relationships in most real-world problems.

Neural networks are perhaps the universal method of forecasting in connection with those that they cannot only find the non-linear structures in problems, they can also simulate the processes of linear processes. For example, the possibility of neural networks in the modeling of linear time-series line were studied and confirmed by a number of researchers [3–5].

One of the main applications of ANN is forecasting. In recent years, it was seen increasing interest in forecasting using neural networks. Forecasting has a long history, and its importance reflected in the application in a variety of disciplines from business to engineering.

The ability to predict the future is fundamental to many decision-making processes in planning, developing strategies, building policy, as well as in the management of supply and stock prices. As such, forecasting is an area in which a lot of effort has invested in the past. In addition, it remains an important and active area of human activity in the present and will continue to evolve in the future. Review of the research needs in the prediction presented to the Armstrong [6].

For several decades in forecasting dominated by linear methods. Linear methods are simple in design and use, and they are easy to understand and interpret. However, linear models have significant limitations, owing to which they cannot discern any nonlinear relationships in data. Approximation of linear models to complex are not linear relationships do not always give a positive result. Earlier in 1980, there have been large-scale competition for forecasting, in which most widely used linear methods were tested on more than 1,000 real-time series [7]. Mixed results showed that none of the linear model did not show the best results worldwide, which can be interpreted as a failure of the linear models in the field of accounting with a certain degree of non-linearity, which is common for the real world.

Predicting financial markets is one of the most important trends in research due to their commercial appeal [8]. Unfortunately, the financial markets are dynamic, non-linear, complex, non-parametric and chaotic by nature [9]. Time series of multi-stationary, noisy, casual, and have frequent structural breaks [10]. In addition, the financial markets also affects a large number of macroeconomic factors [11, 12], such as political developments, global economic developments, bank rating, the policy of large corporations, exchange rates, investment expectations, and events in other stock markets, and even psychological factors.

Artificial neural networks are one of the technologies that have received significant progress in the study of stock markets. In general, the value of the shares is a random sequence with some noise, in turn; artificial neural networks are powerful parallel processors of nonlinear systems, which depend on their own internal relations. Development of techniques and methods that can approximate any nonlinear continuous function without a priori notions about the nature of the process itself seen in the work of Pino [13]. It is obvious that a number of factors demonstrate sufficient efficacy in the forecast prices, and most importantly a weak point in this is that they all contain a few limitations in forecasting stock prices and use linear methods, the relative of this fact, although previous investigation revealed the problem to some extent, none of them provides a comprehensive model for the valuation of shares. If we evaluate the cost and provide a model in order to remove the uncertainty, it is largely can help to increase the investment attractiveness of the stock exchanges. Conduct research to get the best method of forecasting financial time series is currently the most popular and promising task.

2 Latest Researches in Hybrid Time Series Prediction

One approach to solving the difficult-solvable problems of the real world is a hybridization of artificial intelligence and statistical methods to combine the strengths and eliminate weaknesses in the case of each method separately [14].

Bisoi and Dash used a hybrid dynamic neural network (DNN) trained by sliding mode algorithm and differential evolution (DE) [15]. They used this model to predict stock price indices and stock return volatilities of two important Indian stock markets, namely the Reliance Industries Limited (RIL) from one day ahead to one month in advance. The DNN comprises a set of first order IIR filters for processing the past inputs and their functional expansions and its weights are adjusted using a sliding mode strategy known for its fast convergence and robustness with respect to chaotic variations in the inputs. Extensive computer simulations are carried out to predict simultaneously the stock market indices and return volatilities and it is observed that the simple IIR–based DNN–FLANN model hybridised with DE produces better forecasting accuracies in comparison to the more complicated neural architectures.

Kumar proposed Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis and bootstrap method [16]. The aim of this study is to develop an ensemble modeling approach based on wavelet analysis, bootstrap resampling and neural networks (BWANN) for reservoir inflow forecasting. In this study, performance of BWANN model is also compared with wavelet based ANN (WANN), wavelet based MLR (WMLR), bootstrap and wavelet analysis based on multiple linear regression models (BWMLR), standard ANN, and standard multiple linear regression (MLR) models for inflow forecasting. This study demonstrated the effectiveness of proper selection of wavelet functions and appropriate methodology for wavelet based model development. Moreover, performance of BWANN models is found better than BWMLR model for uncertainty assessment, and is found that instead of point predictions, range of forecast will be more reliable, accurate and can be very helpful for operational inflow forecasting.

Wang proposed a Hybrid forecasting model-based data mining and genetic algorithm-adaptive particle swarm optimization [17]. This study proposes a hybrid-forecasting model that can effectively provide preprocessing for the original data and improve forecasting accuracy. The developed model applies a genetic algorithm-adaptive particle swarm optimization algorithm to optimize the parameters of the wavelet neural network (WNN) model. The proposed hybrid method is subsequently examined in regard to the wind farms of eastern China. The forecasting performance demonstrates that the developed model is better than some traditional models (for example, back propagation, WNN, fuzzy neural network, and support vector machine), and its applicability is further verified by the paired-sample T tests.

Singh suggest Big Data Time Series Forecasting Model—a novel big data time series forecasting model which is based on the hybridization of two soft computing (SC) techniques, viz., fuzzy set and artificial neural network [18]. The proposed model is explained with the stock index price data set of State Bank of India (SBI). The performance of the model is verified with different factors, viz., two-factors, three-factors, and M-factors. Various statistical analyzes signify that the proposed model can take far better decision with the M-factors data set.

3 Neural Networks Versus Multiple Regression

In their work Cripps and Engin [19] compared the effectiveness of forecasts made by multiple regression and an artificial neural network using Backpropagation. The comparison is made on the example of predicting the value of residential real estate. Two models were compared using different datasets, functional specifications and comparative criteria. The same specifications of both methods and a real information can explain why other studies, which have compared the ANN and multiple regression, yielded different results.

For an objective comparison of models it is necessary to identify possible problems of each model, which could distort the performance of the method. Research shows that multiple regression and ANN are determined. In addition, some studies on models based on multiple regression are also determined and applied to the data set used in this study.

A comparison used a standard feedforward neural network with training on the basis of back-propagation. The same experiments were performed with many variations in methods of training neural networks. Different neural network architectures, such as ARTMAP (adaptive resonance theory), GAUSSIAN, neural-regression, were also studied. After hundreds of experiments and changes of architecture, a standard method of reverse spread showed a better performance than other architectures of neural networks. Several other studies on comparison of neural networks and multiple regression obtained different results. Thus, ironically, the main issue and the purpose of this study is to determine the causes why some researchers got the best result using multiple regression, while others have come to the conclusion that using neural network is better.

Some studies show the superiority of the ANN multiple regression in solving the problem of real estate market forecasting [20, 21]. Other studies [22], however, showed that ANN are not always superior to regression. Due to the ability to train ANN to learn and recognize complex images without being programmed by certain rules, they can easily be used in a small set of statistical data. In contrast, the regression analysis, neural network does not need to form predetermined functional based determinants. This function of ANN is important because several studies [23, 24] found that the property age has a nonlinear dependence on its value (for a set of data used in their studies). Other studies have shown that in addition to age, the living area also has a non-linear relationship with the value of [25]. Based on the results of previous studies and theoretical capacity of ANN, one would expect that ANN have a better performance than a multiple regression.

When using multiple regression, it is necessary to solve methodological problems of the functional form because of incorrect specification, linearity, and multicollinearity heteroscedasticity. The possibility of a nonlinear functional form in most cases can be translated to a linear-nonlinear relationship before we proceed to using a regression analysis [26]. As noted earlier, some studies found that the age and living area have a nonlinear relationship with the value of the property. Multicollinearity does not affect the predictive capabilities of multiple regression, as well as at ANN [27] because the conclusions are drawn together in a certain area of observation. Multicollinearity, however, makes it impossible to affect separation of supposedly independent variables. Heteroscedasticity arises when using cross the intersection data. In addition to the model of methodological problems, the lack of relevant explanatory variable is another source of error when using multiple regression and ANN. This is often due to a lack of data.

When using a neural network of direct distribution with training on the basis of back propagation of errors, it is necessary to solve the following methodological problems: the number of hidden layers, the number of neurons in each hidden layer, the sample of training data, the size of the sample, the sample of test data and the corresponding size of the sample as well as overtraining. As a general rule, the level of education and the number of hidden neurons affects the storage and generality predicted by the model. The more widely training and the bigger the number of hidden neurons used, the better the production of correct predictions on the training set in the model. On the other hand, ANN is less likely to predict a new data (summary), i.e. ANN ability to generalize weakens when there is overtraining, which can occur when the dimension of the hidden layer is too big. To avoid overtraining, it is advisable to use the heuristic method described in Hecht-Nielsen's article [28]. Despite limitations, there is some theoretical basis to facilitate the determination of the number of hidden layers and neurons in use. In most cases there is no way to determine the best number of hidden neurons without training and evaluation of multiple networks generalization errors. If ANN have only a few hidden neurons, then an error and training error of generalization will be high due to thigh statistical accuracy. If a neural network has too many hidden neurons, there will be a few errors while training, but the error of generalization will be high because of the re-education and high dispersion [29]. If the training sample is not a representative set of data (statistics), there is no basis for ANN training. Typically, a representative training set is generated using a random sample of the dataset. When a training data set is too small, then INS will have a tendency to memorize that training models are too specific and extreme points (noise) will have an extraordinary impact on the quality model. This can be correct, however, with the help of K-fold cross-validation method of teaching.

The authors compared multiple regression and ANN. An attempt made to conduct valid comparisons of predictive abilities for both models. Multiple comparative experiments were carried out with different sets of training data, the functional specifications and in different periods of the forecast. 108 trials were conducted. For comparison, two criteria: MAPE (mean absolute percentage error) and PE (prediction error). Based on the results of prediction, both of these criteria may differ for different samples of a data model. Thus, must treated with caution to each indicator for which data are used criteria to assess forecasting accuracy, as well as to the size of the sample used. When using the sample size from moderate to large, ANN performs better in both criteria with respect to the multiple regression. For these purposes, the size of data samples from 506 to 1,506 cases (a total of 3906 cases) for ANN outperformed multiple regression (using both criteria). In general, a complication of the functional specification ANN training sample size should be increase to ensure that ANN have worked better than a multiple regression. A multiple regression shows better results (using the criterion of the mean absolute relative error) than ANN when using the small size of the data for training. For each of the functional specification model performance multiple regression to some extent remains constant at different sample sizes, while the performance is significantly improves ANN with increasing size of training data.

Fluctuations in the performance of an ANN model are associated with the large number of possible options and the lack of a methodical approach to the choice of the best parameters. For example, experiments must be conduct to determine the best way to represent the data model specification, the number of hidden layers, the number of neurons in each hidden layer, learning rate and the number of training cycles. All of these actions are intended to identify the best model of a neural network. Failure to conduct such experiments can lead to a badly specified ANN model.

If other input variables, such as a fireplace, floors, finishing materials, lot size, connected communication and the type of funding are included, the results may vary.

Research findings provide an explanation why previous studies give different results when comparing multiple regression and ANN for prediction. Prognostic efficiency depends on the evaluation criteria (MAPE and FE) used in conjunction with the size and parameters of the model training. Fluctuations of an ANN model performance may be due to the large number of settings selected by experimentation and depend on the size of the training sample.

4 Hybrid Time Series Forecasting with Using ANFIS

ANFIS (adaptive neuro-fuzzy system) [30] is a multilayer feed forward network. This architecture has five layers such as fuzzy layer, product layer, normalized layer, de-fuzzy layer and total output. The fixed nodes are represented by circle and the nodes represented by square are the adapted nodes. ANFIS gives the advantages of the mixture of neural network and fuzzy logic. The aim of mixing fuzzy logic and neural networks is to design an architecture, which uses a fuzzy logic to show knowledge in fantastic way, while the learning nature of neural network to maximize its parameters. ANFIS put forward by Jang in 1993 integrate the advantages of both neural network and fuzzy systems, which not only have good learning capability, but can be interpreted easily also. ANFIS has been used in many applications in many areas, such as function approximation, intelligent control and time series prediction (Fig. 1).

In this work, we create adaptive neuro-fuzzy system for time series prediction. For example, we used data from Russian government company Rosstat—personnel engaged in research and development in the sector of non-profit organizations. We used system, based on multiple regression for forecasting this indicator and comparison results (Fig. 2) with results (Fig. 3) what we made from ANFIS forecasting system.

From presented figures, we can see what ANFIS model has better performance then multiple regression model.

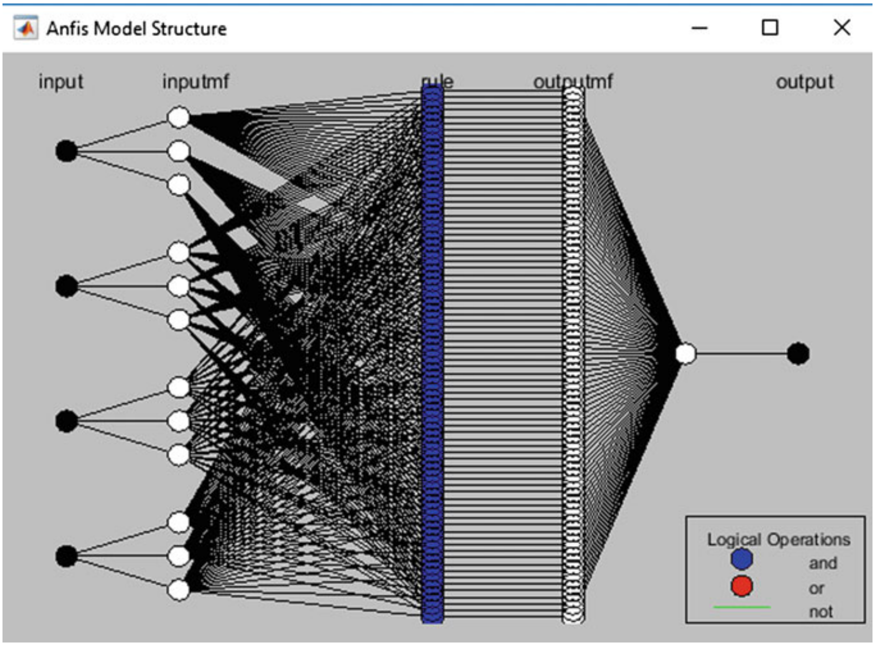


Fig. 1 Structure of ANFIS Model for time series forecasting

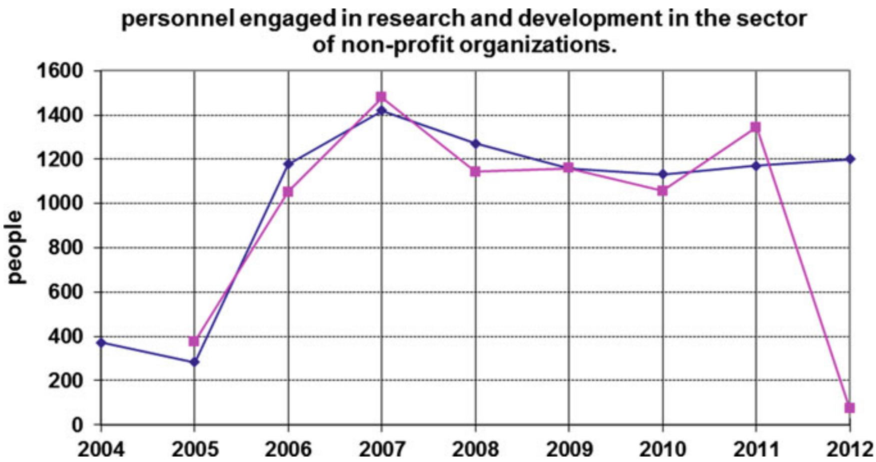


Fig. 2 Forecasting results with Multiple regression model

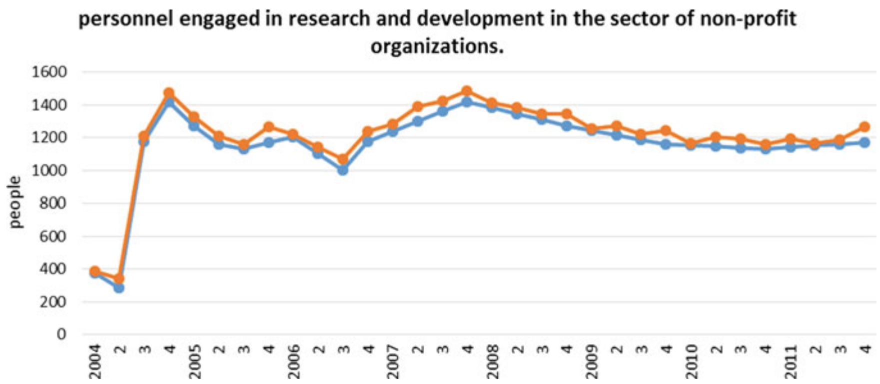


Fig. 3 Forecasting results with ANFIS model

5 Conclusions and Future Work

In this research, we make detailed review of models for time series forecasting such as multiple regression and based on artificial neural networks. A thorough analysis of studies was conducted in this field of research. And the results of studies to identify the qualitative method to make predictions. Particular attention was given to methods based on comparison of multiple regression and artificial neural networks. It was described in detail an example of a study in which conducted a practical comparison of predictive capability of ANN and multiple regression for example predicting the value of residential real estate. Result of this study demonstrating the superiority of neural networks for prediction quality obtained in comparison with the multiple regression. There were also identified some difficulties encountered in predicting both methods. Thus, based on the research results, it can be concluded that in order to get the most superiority as predictions using ANN over multiple regression should be administered as soon as possible largest training data [31]. The larger the training set, the qualitative forecast making the neural network.

In conclusion, it should be noted that in order to solve the problem of predicting and get the best results should develop hybrid methods of time series prediction based on neural networks and statistical techniques used. Specifically, based on theory of artificial immune systems, complementary features formation and negative selection can be applied.

References

1. Widrow, B., Rumelhart, D., Lehr, M.A.: Neural Networks: Applications in Industry, Business and Science. Stanford University (Commun. ACM **37**(3)) (1994)
2. Chang, P.C., Wang, Y.W., Liu, C.H.: The development of a weighted evolving fuzzy neural network for PCB sales forecasting. Expert Syst. Appl. **32**, 86–96 (2007)
3. Hwang, H.B.: Insights into neural-network forecasting of time series corresponding to ARMA (p, q) structures. Omega **29**, 273–289 (2001)
4. Medeiros, M.C., Pedreira, C.E.: What are the effects of forecasting linear time series with neural networks? Eng. Intell. Syst. 237–424 (2001)
5. Zhang, G.P.: An investigation of neural networks for linear time-series forecasting. Comput. Operat. Res. **28**, 1183–1202 (2001)
6. Armstrong, J.S.: Research needs in forecasting. Int. J. Forecast. **4**, 449–465 (1988)
7. Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibdon, M,. Lewandowski, R., Newton, J., Parzen, E., Winkler, R.: The accuracy of extrapolation (time series) methods: results of a forecasting competition. J. Forecast. **1**(2), 111–153 (1982)
8. Majhi, R., Panda, G.: Stock market prediction of S&P 500 and DJIA using bacterial foraging optimization technique. In: 2007 IEEE Congress on Evolutionary Computation, 2007, pp. 2569–2579
9. Tan, T.Z., Quek, C., Ng, G.S.: Brain inspired genetic complimentary learning for stock market prediction. In: IEEE Congress on Evolutionary Computation, 2–5th Sept 2005, vol 3, pp 2653–2660
10. Oh, K.J., Kim, K.-J.: Analyzing stock market tick data using piecewise non linear model. Expert Syst. Appl. (3), 249–255 (2002)
11. Miao, K., Chen, F., Zhao, Z.G.: Stock price forecast based on bacterial colony RBF neural network. J. QingDao Univ. (20), 50–54 (2007) (in Chinese)
12. Wang, Y.: Mining stock prices using fuzzy rough set system. Expert Syst. Appl. (1), 13–23 (2003)
13. Pino, R., Parreno, J., Gomez, A., Priore, P.: Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks. Eng. Appl. Artif. Intell. (21), 53–62 (2008)
14. Averkin A.N., Yarushev, S.A.: Hybrid methods of time series predicton in financial markets. In: International Conference on Soft Computing and Measurements. St. Petersburg: Vol. Saint Petersburg State Electrotechnical University “LETI” name V. I. Ulyanov (Lenin), T. 1. Section 1–3, pp 332–335 (2015)
15. Bisoi, R., Dash, P.K.: Prediction of financial time series and its volatility using a hybrid dynamic neural network trained by sliding mode algorithm and differential evolution. Int. J. Inf. Decis. Sci. **7**(2), 166–191 (2015)
16. Kumar, S. et al.: Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis and bootstrap method. Water Resour. Manag. **29**(13), 4863–4883 (2015)
17. Wang, J. et al.: Hybrid forecasting model-based data mining and genetic algorithm-adaptive particle swarm optimization: a case study of wind speed time series. In: IET Renewable Power Generation (2015)
18. Singh, P.: Big data time series forecasting model: a fuzzy-neuro hybridize approach. In: Computational Intelligence for Big Data Analysis, pp 55–72 (2015)
19. Nguyen, N., Cripps, A.: Predicting housing value: a comparison of multiple regression analysis and artificial neural networks. JRER (3), 314–336 (2001)
20. Tsukuda, J., Baba, S.-I.: Predicting Japanese corporate bankruptcy in terms of financial data using neural networks. Comput. Ind. Eng. **27**, 1–4, 445–448 (1994)
21. Do, Q., Grudnitski, G.: A neural network approach to residential property appraisal. The Real Estate Appraiser **58**, 38–45 (1992)

22. Allen, W.C., Zumwalt, J.K.: Neural Networks: A Word of Caution, Unpublished Working Paper, Colorado State University (1994)
23. Grether, D., Mieszkowski, P.: Determinants of real values. J. Urban Econ. **1**(2), 127–145 (1974)
24. Jones, W., Ferri, M., McGee, L.: A competitive testing approach to models of depreciation in housing. J. Econ. Bus. **33**(3), 202–211 (1981)
25. Goodman, A.C., Thibodeau, T.G.: Age-related Heteroskedasticity in Hedonic House price equations. J. Hous. Res. **6**, 25–42 (1995)
26. Kmenta, J.: Elements of Econometrics. Macmillan Publishing, New York (1971)
27. Neter, J., Wasserman, W., Kutner, M.H.: Applied Linear Statistical Models, 3rd edn. McGraw-Hill, Boston (1990)
28. Hecht-Nielsen, R.: Kolmogorov's mapping neural network existence theorem. Paper presented at IEEE First International Conference on Neural Networks, San Diego, CA (1987)
29. Geman, S., Bienenstock, E., Doursat, R.: Neural networks and the bias/variance dilemma. Neural Comput. **4**, 1–58 (1992)
30. Jang, J.S.R.: ANFIS: Adaptive-network-based fuzzy inference systems. IEEE Trans. Syst. Man Cybern. **23**, 665–685 (1992)
31. Yarushev S.A., Ephremova N.A.: Hybrid methods of time series prediction. In: Hybrid and Synergistic Intelligent Systems: Proceedings of the 2nd International Symposium of Pospelov. BFU Kant, Kaliningrad (2014), pp. 381–388

Power Load Daily-Similarity and Time-Series Prediction Using the Selective Differential Polynomial Network

Ladislav Zjavka and Václav Snášel

Abstract Cooperation on the electricity grid requires from all providers to foresee the demands within a sufficient accuracy. Short-term electric energy estimations of a future demand are needful for the planning of generating electricity in regional grids and operating power systems. An over-estimating of a future load results in an unused spinning reserve. Under-estimating a future load is equally detrimental because buying at the last minute from other suppliers is obviously too expensive. Differential polynomial neural network is a new neural network type, which forms and solves a selective general partial differential equation of an approximation of a searched function, described by observations, with combination sum series of convergent relative polynomial derivative terms. An ordinary differential equation, which can model 1-variable function real time-series, is analogously substituted with partial derivatives of selected time-point variables. A new method of the short-term power demand forecasting, based on similarity relations of subsequent day progress cycles at the same time-series points is presented and tested on 2 datasets. Comparisons were done with the artificial neural network using the same prediction method.

Keywords Power demand prediction • Week load cycle • Differential polynomial neural network • Sum relative derivative term

L. Zjavka (✉) · V. Snášel
IT4innovations, VŠB-Technical University of Ostrava, Ostrava, Czech Republic
e-mail: ladislav.zjavka@vsb.cz

V. Snášel
Department of Computer Science, VŠB-Technical University of Ostrava, Ostrava, Czech Republic
e-mail: vaclav.snasel@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 450, DOI 10.1007/978-3-319-33609-1_11

1 Introduction

Power load forecasting is important for economically efficient and effective control of power systems and enables to plan the operation of generating units. Day-ahead load forecasting techniques can involve autoregressive integrated moving average (ARIMA) [3], chaotic dynamic non-linear models with evolutionary hybrid computation [9], exponential smoothing or interval time-series [4]. 1-variable real power function time-series progress is difficult to predict using deterministic methods as weather conditions as well as other extraneous casual factors can fairly influence it. The power demand model can describe time-points data relations of several subsequent days in previous weeks, as the daily power cycles of each following week are of a similar nature. The daily trained models may form the predictions with respect to actual data of few last days with the same denomination and should be updated to take into account a dynamic character of the applied time-series. Artificial neural networks (ANN) are able to model simple functions of systems, which exact solutions are problematic or impossible to get using standard regression techniques.

$$Y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots$$
$$n - \mathit{number\ of\ input\ variables}\ x(x_1, x_2, \ldots, x_n) \qquad a(a_1, a_2, \ldots, a_n), \ldots - \mathit{vectors\ of\ parameters} \tag{1}$$

Differential polynomial neural network (D-PNN) is a new neural network type, which results from the GMDH (Group Method of Data Handling) polynomial neural network (PNN), created by a Ukrainian scientist Aleksey Ivakhnenko in 1968. The Volterra functional series, a discrete analogue of which is Kolmogorov-Gabor polynomial (1) can express a general connection between input and output variables. This polynomial can approximate any stationary random sequence of observations and can be computed by either adaptive methods or system of Gaussian normal equations. The GMDH decomposes the complexity of a system or process into many simpler relationships each described by the low order polynomial (2) for every pair of the input values. A typical GMDH network maps a vector input $\boldsymbol{x}$ to a scalar output y, which is an estimate of the true function $f(\boldsymbol{x}) = y^t$ [8].

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{2}$$

The D-PNN can combine the PNN multi-layer architecture with some mathematical techniques of differential equation (DE) substitutions. Its partial DE solutions are possible to extract from the backward structure. D-PNN defines and solves a selective form of the general partial DE in a searched function model. Nodes in all network layers produce equivalent convergent series of fractional derivative terms, which sum combinations can substitute for the general DE, decomposing a system model into plenty of partial derivative relations (analogous to the GMDH general connection polynomial decomposition). In contrast with the ANN, each neuron (i.e. substitution DE term) may be added in the total network output, calculated by the sum of the active neuron output values [11].

2 Forecasting Electrical Energy Demands

The potential benefits of the energy demand prediction are obvious useful in the automatic power dispatch, load scheduling and energy control. Electricity demand accurate forecasts are necessary in the operation of electric power grid systems; producers should develop strategies to maximize the profits and minimize risks. The purpose of the short-term electricity demand forecasting is to forecast in advance the system load, represented by the sum of all consumers load at the same time. A precise load forecasting is required to avoid high generation cost and the spinning reserve capacity. Under-prediction of the demands leads to an insufficient reserve capacity preparation and can threaten the system stability, on the other hand, over-prediction leads to an unnecessarily large reserve that leads to a high cost preparations. The nature of parameters that affect this problem includes many uncertainties. The accuracy of a dispatching system is influenced by various conditional input parameters (weather, time, historical data and random disturbances), which a prediction model can involve, applying e.g. fuzzy logic [6]. ANN is able to model the non-linear nature of dynamic processes, reproduce an empirical relationship between some inputs and one or more outputs. It is applied for such purpose regarding to its approximation capability of any continuous nonlinear function with arbitrary accuracy that offer an effective alternative to more traditional statistical techniques. The load at a given hour is dependent not only on the load at the previous hour, but also on the load at the same hour on the previous day, and on the load at the same hour on the day with the same denomination in the previous week. There are also many important exogenous variables that should be considered, especially weather-related variables [5].

The proposed method keynote of the power demand forecasting (using only 1-variable function time-series) is to train a neural network model with daily cycle similarity relations of previous weeks, concerning several consecutive days with the same denomination, foregoing the 24-h prediction. 3 power series of the 3 consequent days in previous weeks at the same time points form the input vector while the following day 24-h series define desired network outputs of the training data set. After training the network can estimate the following day power progress, using 9 input vector variables of 24-h shifted 3 time-series of the current week last 3 days (Fig. 1). The model does not allow for weather or other disturbing effects, as these

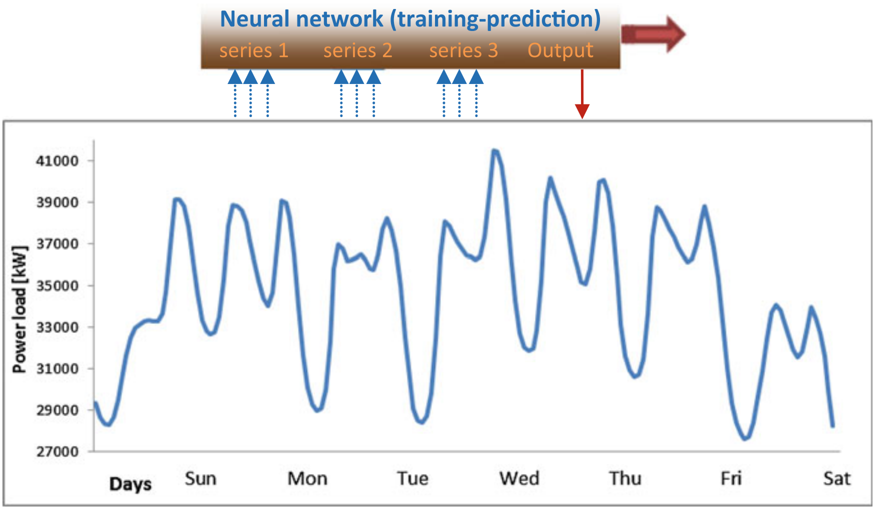


Fig. 1 A typical week cycle of power loads (February) [10]

are not at disposal in the majority of cases. The power demand day cycle progress is much more variable in winter than summer months, which is influenced largely by using heating systems and temperature conditions [10].

3 General Differential Equation Substitution

The D-PNN decomposes and substitutes for the general partial differential Eq. (3), in which an exact definition is not known in advance and which can generally describe a system model with a sum of relative multi-variable polynomial derivative convergent term series (4).

$$a + bu + \sum_{i=1}^{n} c_i \frac{\partial u}{\partial x_i} + \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \cdots = 0 \qquad u = \sum_{k=1}^{\infty} u_k$$
$$u = f(x_1, x_2, \ldots, x_n) - \textit{searched function of all input variables}$$
$$a, B(b_1, b_2,, \ldots, b_n), C(c_{11}, c_{12,}, \ldots) - \textit{polynomial parameters} \tag{3}$$

Substitution partial DE terms are formed according to the adapted method of integral analogues, which is a part of the similarity dimensional analysis. It replaces mathematical operators and symbols of a DE by the ratio of the corresponding values. Derivatives are replaced by their integral analogues, i.e. derivative operators are removed and simultaneously along with all operators are replaced by similarly proportion signs in equations to form dimensionless groups of variables [2].

$$u_i = \frac{\left(a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2 + \cdots\right)^{m/n}}{b_0 + b_1x_1 + \cdots} = \frac{\partial^m f(x_1, \ldots, x_n)}{\partial x_1 \partial x_2 \ldots \partial x_m} \tag{4}$$
$$\begin{aligned}&-\mathit{combination\ degree\ of\ a\ complete\ polynomial\ of\ n-variables}\\&-\mathit{combination\ degree\ of\ denominator\ variables}\end{aligned}$$

The designed fraction substitution DE terms (4), describe partial mutual relative derivative polynomial changes of *n*-input variables. The numerator (4) is a complete polynomial of *n*-input variables (*n*-degree) and partly defines an unknown function *u* of Eq. (3). The denominator applies incomplete polynomials of the competent derivative combination of input variables. The root function of the numerator decreases a combination degree of the input polynomial (4), in order to get the dimensionless units [2]. An ordinary differential Eq. (5) with only time derivatives can describe 1-variable function time-series and may be solved analogously by means of partial DE term (3) substitutions, which apply relations of several time-point variables [12].

$$a + bf + \sum_{i=1}^{m} c_i \frac{df(x_i)}{dt} + \sum_{i=1}^{m}\sum_{j=1}^{m} d_{ij} \frac{df(x_i, x_j)}{dt} + \cdots + \sum_{i=1}^{m} cc_i \frac{d^2 f(x_i)}{dt^2} + \cdots = 0 \tag{5}$$
$$f(x) - \mathit{function\ of\ independent\ time\ observations}\ x(x_1, x_2, \ldots, x_m)$$

Blocks of the D-PNN (Fig. 2) form derivative neurons with the same inputs, one for each fractional polynomial derivative combination, so each neuron represents a substitution DE term (4). Each block contains a single output GMDH polynomial (2), without derivative part. Neurons do not affect the block output but can participate directly in the total network output sum calculation of a DE composition. Each block has *1* and neuron *2* vectors of adjustable parameters ***a***, and ***a***, ***b*** resp.

$$F\left(x_1, x_2, u, \frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \frac{\partial^2 u}{\partial x_1^2}, \frac{\partial^2 u}{\partial x_1 \partial x_2}, \frac{\partial^2 u}{\partial x_2^2}\right) = 0 \tag{6}$$
$$\mathit{where}\ F(x_1, x_2, u, p, q, r, s, t)\ \mathit{is\ a\ function\ of}\ 8\ \mathit{variables}$$

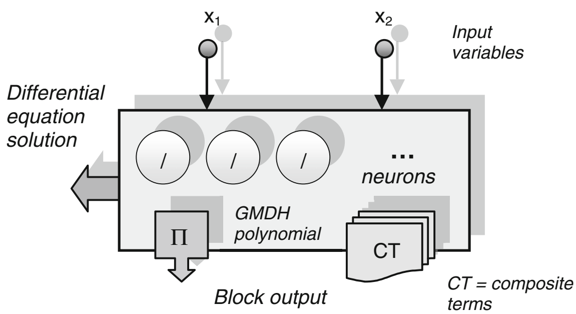


Fig. 2 The D-PNN block involves simple and composite neurons (DE terms)

While using 2 input variables the 2nd order partial DE may be expressed in the form (6), which involves derivative terms formed in respect of all the GMDH polynomial (2) variables. Each D-PNN block forms *5* corresponding simple derivative neurons of single x_1, x_2 (7) square x_1^2, x_2^2 (8) and combination x_1x_2 (9) derivative variables, which can directly solve and substitute for the 2nd order partial DE (6), most often used to model physical or natural system non-linearities.

$$y_1 = \frac{\partial f(x_1, x_2)}{1} = w_1 \frac{\left(a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2\right)^{1/2}}{b_0 + b_1x_1} \tag{7}$$

$$y_3 = \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} = w_3 \frac{a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2}{b_0 + b_1x_2 + b_2x_2^2} \tag{8}$$

$$y_5 = \frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2} = w_5 \frac{a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2}{b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2} \tag{9}$$

4 Backward Selective Differential Polynomial Neural Network

Multi-layer networks form composite functions (Fig. 3). Composite terms (CT), i.e. derivatives in respect of variables of previous layer blocks, are calculated according to the composite function partial derivation rules (10), (11), formed by products of the partial derivatives of the external and internal functions.

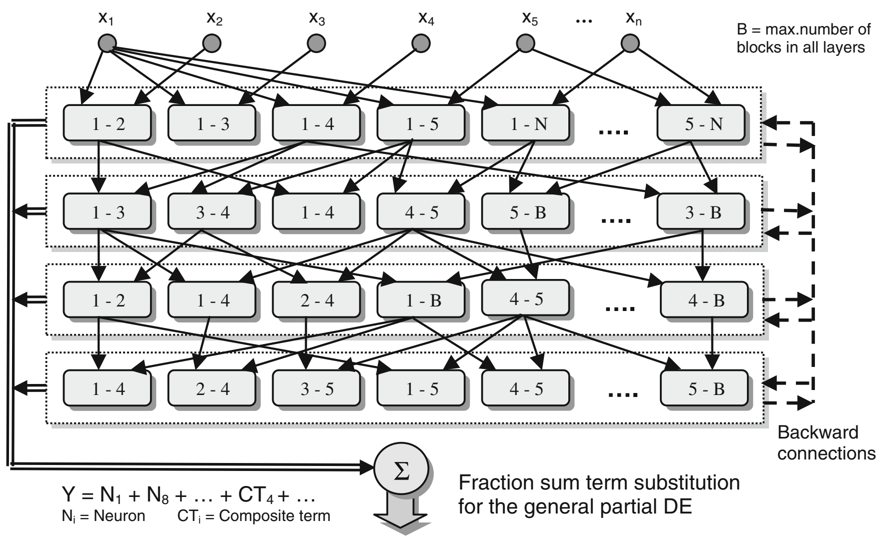


Fig. 3 N-variable D-PNN selects from 2-variable combination blocks in each hidden layer

$$F(x_1, x_2, \ldots, x_n) = f(y_1, y_2, \ldots, y_m) = f(\phi_1(X), \phi_2(X), \ldots, \phi_m(X)) \qquad (10)$$

$$\frac{\partial F}{\partial x_k} = \sum_{i=1}^{m} \frac{\partial f(y_1, y_2, \ldots, y_m)}{\partial y_i} \cdot \frac{\partial \varphi_i(X)}{\partial x_k} \qquad k = 1, \ldots, n \qquad (11)$$

Each D-PNN block can form 5 simple neurons, additionally the blocks of the 2nd and following hidden layers produce CT (extended neurons) using composite substitution derivatives with respect to the output and input variables of the back connected previous layer blocks, e.g. 3rd layer blocks can form linear CT in respect of 2nd (12) and 1st layer (13). The number of neurons in blocks, which include composite function derivatives, doubles each previous back-connected layer [11].

$$y_2 = \frac{\partial f(x_{21}, x_{22})}{\partial x_{11}} = w_2 \frac{(a_0 + a_1 x_{21} + a_2 x_{22} + a_3 x_{21} x_{22} + a_4 x_{21}^2 + a_5 x_{22}^2)^{1/2}}{x_{22}} \cdot \frac{x_{21}}{b_0 + b_1 x_{11}} \qquad (12)$$

$$y_3 = \frac{\partial f(x_{21}, x_{22})}{\partial x_1} = w_3 \frac{(a_0 + a_1 x_{21} + a_2 x_{22} + a_3 x_{21} x_{22} + a_4 x_{21}^2 + a_5 x_{22}^2)^{1/2}}{x_{22}} \cdot \frac{(x_{21})^{1/2}}{x_{12}} \cdot \frac{x_{11}}{b_0 + b_1 x_1} \qquad (13)$$

The square and combination derivative terms are also calculated according to the composite function derivation rules (11). D-PNN is trained with only a small set of input-output data samples, in a similar way to the GMDH algorithm [8]. The number of network hidden layers should coincide with the number of input variables. However the D-PNN with an increased number of input variables (more than 3) usually need not define all the possible DE terms (using a deep multi-layer structure), e.g. 4–5 hidden layers can form a function model for 9 input variables (Fig. 3).

Only some of all the potential combination DE terms (neurons) may be included in the DE solution, in despite of they have an adjustable term weight (w_i). A proper neuron combination, which substitutes for a DE solution, is not able to accept a disturbing effect of the rest of the neurons (possible to form other solutions) in the parameter optimization process. D-PNN total output Y is the arithmetic mean of all the active neuron output values so as to prevent a changeable number of neurons (of a combination) from influencing the total network output value (14) [11].

$$Y = \frac{\sum_{i=1}^{k} y_i}{k} \qquad k = actual\ number\ of\ active\ neurons \qquad (14)$$

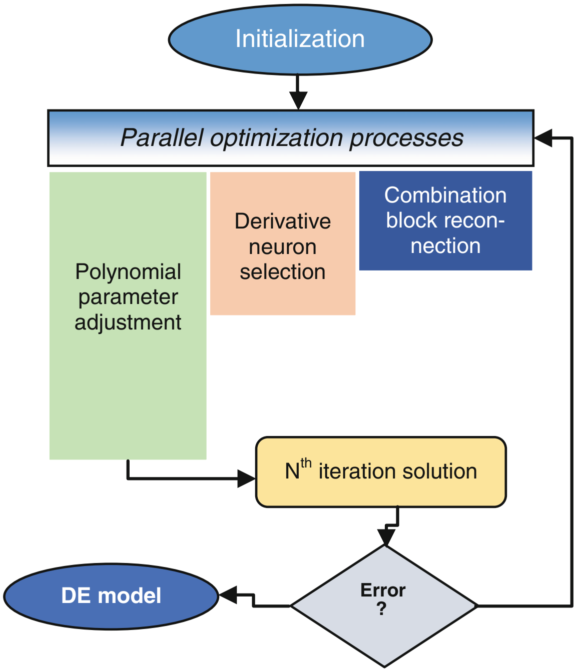


Fig. 4 3-parallel gradually finishing processes of the parameter and selection adjustment

The selection of a fit neuron combination is the principal part of the DE composition, performed simultaneously with block combination reconnections and a polynomial parameter adjustment in the initial composing phase (Fig. 4). It may apply the binary particle swarm optimization, able to solve binary (1, 0) combinatorial problems. The D-PNN (with more than 3 input variables) must select from all the possible 2-combination blocks (nodes) in each hidden layer (Fig. 3), analogous to the original GMDH faces the combinatorial explosion. It can use the simulated annealing, which employs a random search which not only accepts changes that decrease the objective function (assuming a minimization problem) but also some changes that increase it.

The error function, calculated as the root mean squared error (RMSE) (15), requires a minimization with respect to the polynomial parameters, which could be done by means of the gradient steepest descent method [7] supplied with sufficient random mutations to prevent from to be trapped to a local error depression.

$$E = \sqrt{\frac{\sum_{i=1}^{M} \left(Y_i^d - Y_i\right)^2}{M}} \rightarrow \min$$

$Y_i^d = desired\ output$ $\quad Y_i = estimated\ output\ for\ ith\ training\ vector$ $\quad M = number\ of\ samples$ (15)

5 Power Demand Forecast Model Experiments

The presented D-PNN (Fig. 3) applied the described daily-similarity prediction method (Fig. 1), based on day cycles data relations of the power consumption 1-variable historical time-series [10]. A power demand daily period consists typically of 2 peak-hs, morning (midday) and evening peak consumptions (Fig. 5). 3 power time-series of 3 consequent days (foregoing the forecasting day denomination) of several previous weeks at the same time points form the network input vector while the 24-h shifted values define desired scalar outputs in the training data set (Fig. 1). After training the network can predict the following day power values in respect of the current week last 3 day series in a uniform time.

Comparisons were done with 1 or 2-layer ANN with 20 neurons in hidden layers using the sigmoidal activation function using the same similarity prediction method as the D-PNN (Fig. 7). Some models can embody higher prediction errors, influenced likely by a rude break in the weather conditions or some extraneous factors. Sundays, Saturdays and legal holidays power models result largely in higher inaccuracies, induced likely by more variable demands of weekends and holidays (Fig. 6).

The 2nd data set comprises measured electric load profiles for Canadian detached houses [12], involving much more variable power progress of the average electric draw of the house (in Watts), evoked likely by a lower quantity in the customer demands and some more exogenous factors (Fig. 8). The Sundays and holidays are mostly peak consumption days in this case.

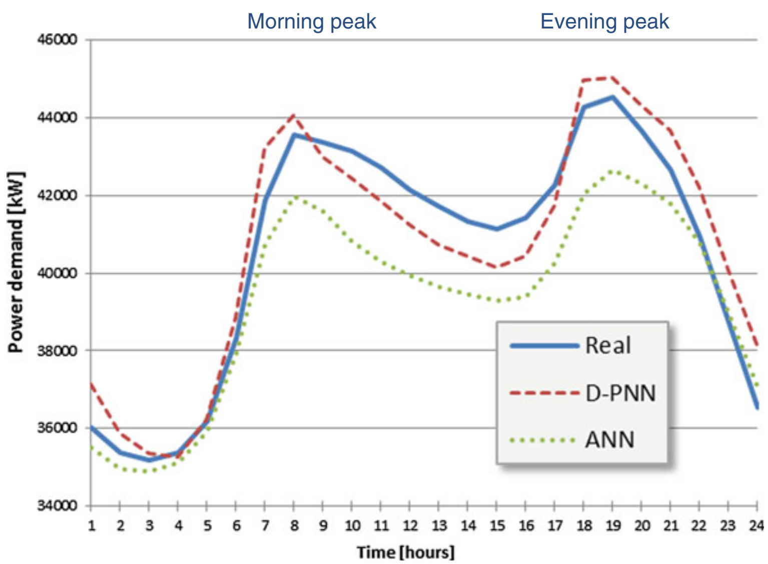


Fig. 5 PJM, 25.1.2013 (Friday), $\text{D-PNN}_{\text{RMSE}} = 875.4$, $\text{ANN}_{\text{RMSE}} = 1508.3$

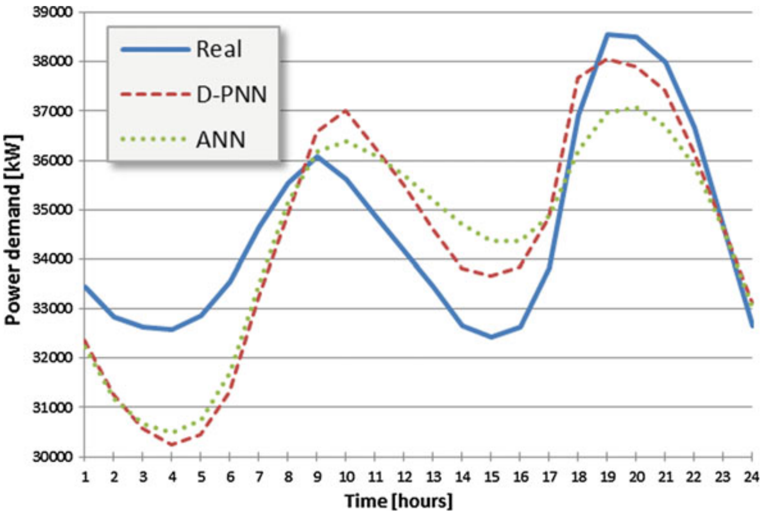


Fig. 6 PJM, 27.1.2013 (Sunday), D-PNN_{RMSE} = 1310.3, ANN_{RMSE} = 1424.8

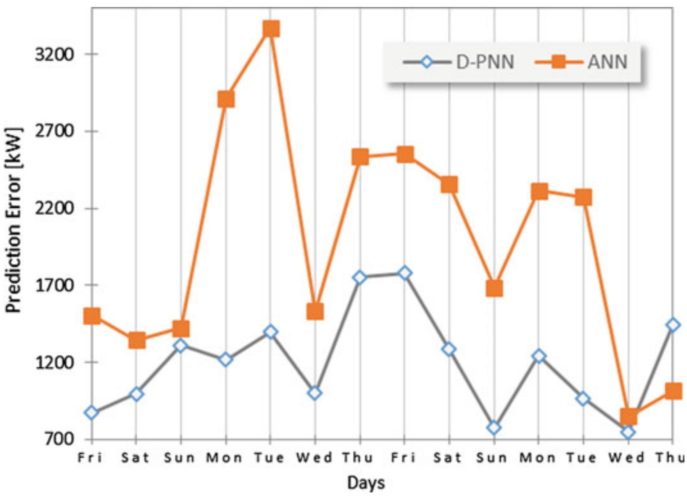


Fig. 7 2-week daily power load prediction errors (25.1.2013 to 7.2.2013), Average RMSE: $DPNN_{RMSE}$ = 1200.4, ANN_{RMSE} = 1978.6 [10]

The real models could be backward tested for corresponding days of the last not trained week. 3-week training period seems to define an optimal learning scheme, it is necessary to apply less or more week-cycles (2–6) in some cases. The valid number of training week periods might be estimated in practice with respect to previous day(s) best testing results, as the progress of the subsequent daily cycles is

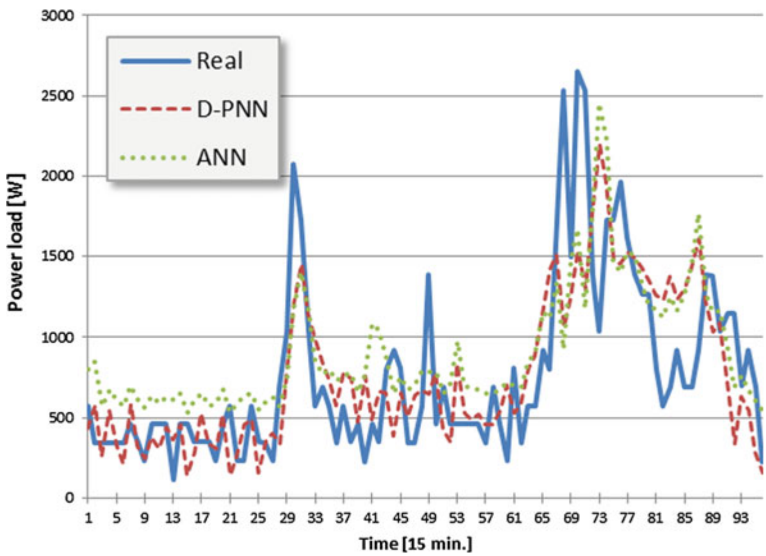


Fig. 8 ANEX, 1.2.1995 (Wednesday), $\text{D-PNN}_{\text{RMSE}} = 405.5$, $\text{ANN}_{\text{RMSE}} = 437.5$ [1]

quite of a uniform character. There is necessary to consider also legal holidays, which corresponding week Sundays (or working days) might replace in the learning scheme.

6 Conclusions

The power daily affinities, foregoing to the current week next day denomination, primarily influence the model solution accuracy, as no other considerable effects were taken into account especially weather conditions, which an optimal prediction model should apply combined with some input exogenous factors. Several previous weeks 3 + 1 day hourly series form a learning scheme of the neural network model, which after training gradually applies the current week corresponding last 3-day power series to estimate following 24-h load values. A new neural network type D-PNN extends the GMDH polynomial network structure. 1-variable time-series function of daily observations is modeled by means of an ordinary differential equation solution, formed with producing sum series of polynomial derivative terms.

Acknowledgements This work was supported by The Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II) project “IT4Innovations excellence in science—LQ1602” and is partially supported by Grant of SGS No. SP2016/97, VŠB—Technical University of Ostrava, Czech Republic.

References

1. ANEX Canadian Occupant-driven Load Profiles. www.iea-annex54.org/annex42/data.html
2. Chan, K., Chau, W.Y.: Mathematical theory of reduction of physical parameters and similarity analysis. Int. J. Theor. Phys. **18**, 835–844 (1979)
3. Darbellay, G.A., Slama, M.: Forecasting the short-term demand for electricity. Int. J. Forecast. **16**, 71–83 (2000)
4. Garcıa-Ascanio, K., Mate, C.: Electric power demand forecasting using interval time series. Energy Policy **38**, 715–725 (2010)
5. Hippert, H.S., Pedreira, C.E., Souza, R.C.: Neural networks for short-term load forecasting: a review and evaluation. IEEE Trans. Power Syst. **16**(1) (2001)
6. Mamlook, R., Badran, O., Abdulhadi, E.: A fuzzy inference model for short-term load forecasting. Energy Policy **37**, 1239–1248 (2009)
7. Nikolaev, N.Y., Iba, H.: Polynomial harmonic GMDH learning networks for time series modelling. Neural Netw. **16**, 1527–1540 (2003)
8. Nikolaev, N.Y., Iba, H.: Adaptive Learning of Polynomial Networks. Springer, New York (2006)
9. PJM Interconnection, regional transmission organization. www.pjm.com/markets-and-operations/energy/real-time/loadhryr.aspx
10. Unsihuay-Vila, C., Zambroni, A.C., Marangon-Lima, J.W., Balestrassi, P.P.: Electricity demand and spot price forecasting using evolutionary computation combined with chaotic nonlinear dynamic model. Electr. Power Energy Syst. **32**, 108–116 (2010)
11. Zjavka, L., Pedrycz, W.: Constructing general partial differential equations using polynomial and neural network. Neural Netw. **73**, 58–69 (2016)
12. Zjavka, L., Snášel, V.: Constructing ordinary sum differential equations using polynomial networks. Inf. Sci. **281**, 462–477 (2014)

An EEG Classification Approach Based on Intrinsic Signal Properties and Wavelets

Petr Gajdoš, Pavel Dohnálek, Michal Čerbák and Jitka Mohylová

Abstract In both medicine and technology, human brainwave recognition is a very much discussed topic because of its wide applications. These range from disease diagnostic to brain-computer interfaces used to control various devices. In this paper, we present a novel combined approach to epileptic spike recognition from recorded electroencephalograms. The highest accuracy obtained is 99.22 % when measured with the 10-fold cross-validation of a single patient's data. This is achieved by combining a continuous wavelet transform to provide classification marks and classifiers to assign them. Results are also collected for the scenario of training the classifiers on one record of the patient and testing on another, taken at a different time. This provides an insight on the classifier's ability to generalize for a given patient.

Keywords Electroencephalograhy · Wavelet · Classification · Feature extraction

1 Introduction

Epileptic seizures has been known to mankind throughout its entire history, although it wasn't relatively recently that they have been recognized as illness symptoms. Around 400 B.C., Hippocrates referred to an epileptic seizure as "the sacred disease". It wasn't until the 19th century that the first medicine used to counteract a seizure was invented. As such, the problem of epileptic seizures has been under research for

P. Gajdoš · P. Dohnálek (✉) · M. Čerbák
Department of Computer Science, FEECS, VŠB-Technical University of Ostrava, Ostrava, Czech Republic
e-mail: pavel.dohnalek@vsb.cz

P. Gajdoš
IT4Innovations, National Supercomputing Center, Ostrava, Czech Republic

J. Mohylová
Department of General Electrical Engineering
VŠB-Technical University of Ostrava, 17. Listopadu 15,
708 33 Ostrava, 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 450, DOI 10.1007/978-3-319-33609-1_12

a very long time, but as of the writing of this paper still no reliable cure exists as it is still unknown what really causes them.

One of the standard ways of detecting epileptic seizures before treating them is via an electroencephalograph (EEG). Nowadays it is widely documented that a seizure has a very distinct manifestation in an EEG record. This leads to the idea of automatically detecting and classifying seizures from these records. Given the nature of the human brain, recognizing anything with good reliability is a challenging task. This include seizure detection.

In this paper, we propose a combined approach to epileptic seizure detection that is shown to perform well when detecting seizure episodes in a patient. This approach combines the Continuous Wavelet Transform (CWT) used to mark the places where a seizure occurs and a classifier to classify the seizure. Real EEG records with real epileptic seizure spikes were used when the algorithm was tested, providing a solid background for real-world application of the approach.

2 Dataset

In order to test the suggested approach, real-world data obtained from an epileptic patient was used. This data represents a various recording of variable length, ranging from several minutes to several hours, of the same patient. The records were taken with a Brain-Quick device, using 19 leads places according to the international 10–20 system. The sampling frequency was 128 Hz. The signal was filtered to remove the frequencies below 0.4 Hz and above 70 Hz, taking into account the standard "operational" frequencies of the human brain and providing a reasonable margin. The sensitivity was 100 µv per 100 mm.

Several recordings were taken from a single epileptic patient, each taken at a different recording session, providing variable recording conditions. The three recordings used in this research were labeled as PAT01–1, PAT01–2 and PAT01–3. PAT01–1 was the shortest with the length of less than one minute, but containing several epileptic spikes. PAT01–2 provided 8 min of signals, while the PAT01–3 was the longest, almost 19 min (Fig. 1).

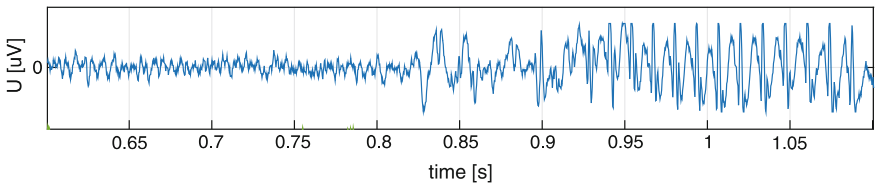


Fig. 1 An example of epileptic seizure manifestation on an EEG record. The seizure is visible in the second half of the graph, with rather regular spike-wave formations in the last third of the figure

3 Continuous Wavelet Transform

Wavelet transform provides a time-frequency analysis of the signal. Historically older Fourier transform provides information about obtained frequencies, however, it isn't so indicative of their position at the time, so it is only suitable for stationary signals.

In the Fourier transform, the analyzing functions are complex exponentials, the $e^{j\omega t}$. The resulting transform is a function of a single variable, ω [1]. In the short-time Fourier transform, the analyzing functions are windowed complex exponentials, $w(t) = e^{j\omega t}$, and the result is a function of two variables. The STFT coefficients $F(\omega, \tau)$ represent the match between the signal and a sinusoid with angular frequency ω in an interval of a specified length centered at τ [2].

The continuous wavelet transform (CWT) uses inner products to measure the similarity between a signal and an analyzing function. In the CWT, the analyzing function is a wavelet ψ. The CWT compares the signal to shifted and compressed or stretched versions of a wavelet. Stretching or compressing a function is collectively referred to as dilation or scaling and corresponds to the physical notion of scale. The result is a function of two variables. The two-dimensional representation of a one-dimensional signal is redundant. If the wavelet is complex-valued, the CWT is a complex-valued function of scale and position. Analogously, if the signal

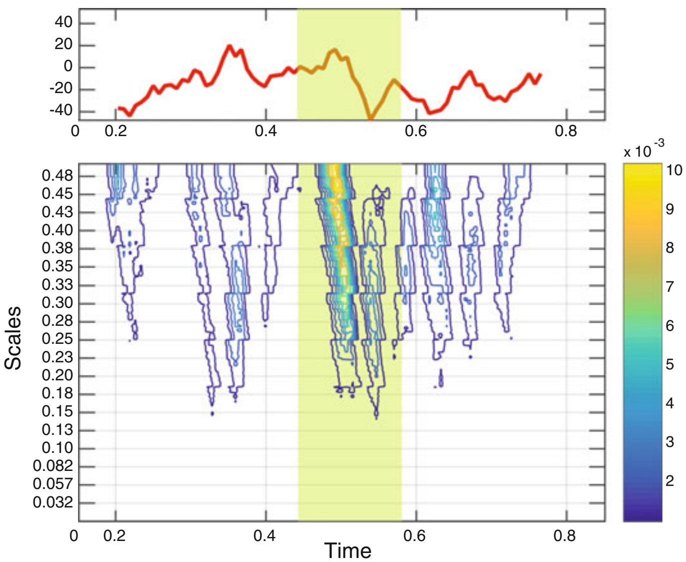


Fig. 2 An example of epileptic spike detection via CWT. The *upper plot* represents the analyzed signal, the *lower plot* shows the percentage of energy for each wavelet coefficient at a given scale and time

is real-valued, the CWT is a real-valued. Complex values are also used in medical applications [3]. Figure 2 provides an example of an epileptic spike and it's manifestation in a CWT scalogram.

4 Wavelet Marker

The aim is to prepare data for further processing and classification. The input is dataset that contains the EEG data. These data are composed of discrete signals recorded on multiple channels. These are represented by a matrix where the rows are data measured by one sensor, and each column represents one of the sensors. CWT detects spikes in data by comparing them to mother wavelets. Mother wavelets are created from short selected EEG data, containing real spikes.

The marker creates data packets that contain prepared data for classifiers. Packages include EEG data recorded at intervals across all channels. For each channel, the presence of a spike is determined by CWT and 9 scalar statistics that describe the characteristics of the EEG signal are computed.

4.1 EEG Data Preprocessing

Division to windows To divide the data to windows, a rectangular window function is used. The result is a series of EEG signals from all channels for a given interval. Division into fixed-length windows can represent a problem when determining the presence of a spike. If a potential spike occurs at the end of window, CWT cannot detect it. Therefore, division to overlapping windows of fixed length can avoid such problems. The overlap value indicates how many values from current window are overlapped by the very next window. The size of the overlap is adjusted to be greater than the size of a spike. A spike located exactly at the end of the window and that doesn't end at it will be obtained by the next window being computed. This ensures that all spikes can be detected.

Preparation of mother wavelets The principle for designing a new wavelet for CWT is to approximate a given pattern using the least squares optimization under constraints leading to an admissible wavelet well suited for the detection pattern using the CWT. To achieve that, epileptic spikes were manually selected and used as a template. A custom mother wavelet was then created from the template and used in CWT detection of epileptic spikes.

CWT calculation Calculation of coefficients returns the CWT of the real-valued signal. The wavelet transform is computed for the specified scales using the analyzing wavelets. Scales are 1-D vectors with positive elements. Scale values determine the degree to which the wavelet is compressed or stretched. Low scale values compress the wavelet and correlate better with high frequency. The low scale CWT coefficients represent the fine-scale features in the input signal vector. High scale values stretch the wavelet and correlate better with the low frequency content of the signal. The high scale CWT coefficients represent the coarse-scale features in the input signal. Coefficients are obtained by computing the product of the signal with the shifted and scaled analyzing wavelet and integrating the result [4, 5].

These factors reflect the degree of similarity between the test wavelets and these signals. Input signal is contaminated with noise. Therefore, it contains coefficients that do not reflect real spikes but only noise that is similar to the analyzing wavelet. These non-zero-valued coefficients are filtered out using a threshold. Coefficients that are higher than the threshold are considered to be spikes. Its necessary to determine the number of spikes in a given window. Each spike is identified by a set of non-zero coefficients of different values that are higher in the center and lower by the edges. These places are the scalogram local maxima, so their count is equal to the number of local maxima in coefficients [6].

Statistics As mentioned above, statistics were computed from the EEG signal. These statistics are then used as the input for classifiers. For the best description of EEG segments the following set of signs describing the frequency and time graphoelements parameters were selected.

1. Amplitude variability in the segment with length L. The average amplitude of the signal segment consists of unidirectional (DC) and alternating (AC) amplitude

$$\overline{A} = A_{DC} + \overline{A}_{AC} \tag{1}$$

2. The DC part can be computed as

$$A_{DC} = \frac{\sum_{i=1}^{L} x_i}{L} \tag{2}$$

3. The DC component reflects the crustal activity and as such should be removed. The average AC amplitude is calculated by

$$\overline{A}_{AC} = \sqrt{\frac{\sum_{i=1}^{L} x_i}{L} - \overline{A}_{DC}^2} = \sigma \tag{3}$$

4. Amplitude variability can be computed by

$$A_{VAR} = \frac{L \sum_{i=1}^{L} x_i}{[\sum_{i=1}^{L} |x_i|][\sum_{i=1}^{L} |x_i|] + \epsilon} - 1 \tag{4}$$

5. The AC amplitude variability can be computed by

$$AC_{VAR} = \frac{L(\sum_{i=1}^{L} x_i - A_{DC}^2)}{[\sum_{i=1}^{L} |x_i| - LA_{DC}][\sum_{i=1}^{L} |x_i| - LA_{DC}] + \epsilon} - 1 \quad (5)$$

6. Difference between the maximum positive and minimum negative amplitude:

$$A_{+max} = A_{max} - A_{DC} \quad (6)$$

$$A_{-min} = A_{min} - A_{DC} \quad (7)$$

7. The maximum value of the first derivative of signal segment approximated by this relationship:

$$\frac{dx}{dt} = x_{i+1} - x_i \quad (8)$$

8. The maximum value of the second derivative of signal segment is also computed:

$$\frac{d^2x}{dt} = x_{i+4} - 2x_{i+2} + x_i \quad (9)$$

5 Classifiers Being Evaluated

To provide an overview of learning algorithms with application to human brainwave recognition, five distinct classifiers were tested. The following subsections provide a brief informal description for each of the classifiers.

k-Nearest Neighbors k-NN is a non-parametric algorithm, meaning that it makes no assumptions about the structure or distribution of the underlying data, thus being suitable for real-world problems that usually do not follow the theoretical models exactly. The method is also considered to be a lazy learning algorithm as it performs little to no training during computation. As a result, the method uses the whole training dataset during classification. k-NN is well known for its simplicity, speed and generally good classification results in applications like bioinformatics [7, 8], image processing [9], audio processing [10] and many others.

Classification and Regression Tree This algorithm classifies a sample according to groups of other samples with similar properties. During training, the training data is continuously divided into smaller subsets (tree nodes). When the divisions are finished, the samples are clustered together according to their properties. Testing samples are then evaluated against certain conditions in each node and propagated throughout the tree. When the sample reaches a leaf node, it is then assigned the class to which the samples in that node belong. CARTs are still under extensive research and can be used as a standalone classifier [11] or as a part of larger algorithmic structures [12].

Self-organizing maps The self-organizing map (SOM) is one of the common approaches on how to represent and visualize data and how to map the original dimensionality and structure of the input space onto another—usually lower-dimensional—structure in the output space. We can imagine the input data to be transformed to vectors, which are recorded in neural network. Only the adjacent neurons are interconnected. Besides of the input layer, SOM has only the output (competitive) layer. The number of inputs is equal to the dimension of input space. Every input is connected with each neuron in the grid, which is also an output (each neuron in grid is a component in output vector). With growing number of output neurons, the quality coverage of input space grows, but so does computation time. SOM can be used as a classification or clustering tool that can find clusters of input data which are closer to each other [13].

Random Forest RF is an ensemble algorithm consisting of a number of CARTs combined together into a single voting system. Each CART receives a dataset with only a subset of parameters, giving it a perspective and knowledge about the problem unique from any other CART. These parameters are chosen randomly, but each parameter set is unique. Its downside is both time and spatial complexity inherent to its nature. However, the algorithm has been shown to be very robust in performance in many research areas [14, 15], making it a suitable alternative for consideration in both classification and regression problems.

Nearest Centroid Classifier An extremely fast classifier. The approach is similar to that of k-NN, but instead of k closest training samples the method picks the label of the class whose training samples' mean (centroid) is the closest to the signal query. The speed and simplicity of the algorithm is compensated by low classification performance. Therefore NCC is usually coupled with one or more data preprocessing techniques. In many implementations the method has been successfully used to create pattern recognition systems in human activity recognition [16] or bioinformatics [17, 18].

6 Experiments and Results

In this section, the summary of the experiments performed is provided, along with the result for individual tasks set to achieve.

6.1 Tasks

To evaluate the suggested approach, we set 2 tasks to accomplish. The first one is the 10-fold cross-validation of the different recordings. During the cross-validation, classifier abilities are measured in the scope of a single dataset. The second task is to train the classifiers on one records and testing on another. This tests a classifier's ability to generalize for a given patient. The resulting data can be seen in the tables and are described in the Results section below.

6.2 Experimental Settings

Since the recorded signals went through a number of preprocessing techniques, it is important to note the settings in use. For segmentation, a Haar window with the length of 32 was used. Five different 32 sample spike waveforms were manually collected from each channel (95 waveforms in total) for mother wavelets. Then they were averaged to obtain an average representation of the spike sample. The averaged waveform was then used to create a mother wavelet for each channel. During CWT, scales were set as a sequence 1 through 64 (double the length of a mother wavelet) with the increment of 0.1. Increments for the signal and mother wavelet were set to match the sampling frequency of 128 Hz.

The classifier settings were chosen empirically. SOM were set as an 8×8 hexagonal grid computed with 1000 iterations. RF contained 1000 CARTs, OMP was set to compute 10 sparse coefficients, k-NN's k parameter was 3. These are the values that have been shown to provide satisfactory results in previous research.

In each table, confusion matrix data are provided with the standard binary-classification statistics. The True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN) values are related to the spike detection, that is TP means spike expected and detected, FP means spike detected but normal EEG expected and so on.

6.3 Results

As mentioned above, one of the tasks was to perform the 10-fold cross-validation. This was performed on all of the tested recordings. Tables 1, 2 and 3 show the results for the three recordings. It doesn't come as a surprise that RF performs the best, considering it's ensemble paradigm. SOM appears to be a suboptimal classifier in this task, as well as k-NN and OMP. NCC, while commonly used in medicine as a classification procedure, was highly inaccurate.

Table 4 contains the results obtained when the classifiers were trained on the PAT01-2 dataset and tested on a much longer recording, PAT01-3. In terms of

Table 1 10-fold cross-validation for the PAT01-1 recording

Method	TP	FP	FN	TN	Sensitivity	Specificity	Model accuracy
SOM	6	26	1	222	85.71	89.52	89.41
CART	29	3	2	221	93.55	98.66	98.04
RF	30	2	0	223	100.00	99.11	99.22
k-NN	9	23	3	220	75.00	90.53	89.80
NCC	32	0	200	23	13.79	100.00	21.57
OMP	21	11	5	218	80.77	95.20	93.73

Table 2 10-fold cross-validation for the PAT01-2 recording

Method	TP	FP	FN	TN	Sensitivity	Specificity	Model accuracy
SOM	126	79	58	275	68.48	77.68	74.54
CART	174	31	27	306	86.57	90.80	89.22
RF	193	12	27	306	87.73	96.23	92.75
k-NN	143	62	59	274	70.79	81.55	77.51
NCC	205	0	325	8	38.68	100.00	39.59
OMP	136	69	42	291	76.40	80.83	79.37

Table 3 10-fold cross-validation for the PAT01-3 recording

Method	TP	FP	FN	TN	Sensitivity	Specificity	Model accuracy
SOM	14	104	8	382	63.64	78.60	77.95
CART	71	47	46	344	60.68	87.98	81.69
RF	66	52	10	380	86.84	87.96	87.80
k-NN	45	73	35	355	56.25	82.94	78.74
NCC	118	0	381	9	23.65	100.00	25.00
OMP	45	73	42	348	51.72	82.66	77.36

Table 4 When trained on PAT01-2 and tested on PAT01-3

Method	TP	FP	FN	TN	Sensitivity	Specificity	Model accuracy
SOM	87	208	383	4068	18.51	95.14	87.55
CART	130	130	165	4321	44.07	97.08	93.78
RF	96	23	199	4428	32.54	99.48	95.32
k-NN	153	1188	142	3263	51.86	73.31	71.98
NCC	295	4434	0	17	100.00	0.38	6.57
OMP	179	1292	116	3159	60.68	70.97	70.33

classifier performance, the results are virtually the same—RF performed best, followed by CART. NCC provided the least satisfactory results, indicating that samples from different classes are too intertwined to be measured by distance alone. While k-

NN performed better, its results suggest the same. The results show a certain degree of robustness when used on the given data, indicating distinct internal properties of the signal.

Unfortunately, the absence of a standardized epileptic spike dataset makes it difficult to provide a comparison with other results reported in the current literature, because the data is obtained under very different conditions.

7 Conclusion

The results achieved suggest that the statistics computed from the segmented recordings provide a significant measure of the signals' internal properties. They can be used with a notable success to classify epileptic spikes in a patient, with Random Forest achieving the most accurate results. A wavelet-based marking algorithm was also proposed and tested. To further the research, the ability of the classifiers to generalize on untrained patients need to be tested. Also, it's likely that more statistics significant for EEG signal can be described and used as input. This constitutes the future work on this research.

Acknowledgments This work was supported by SGS project, VSB-Technical University of Ostrava, under the grant no. SP2016/97. This work was also supported by The Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II) project IT4Innovations excellence in science—LQ1602.

References

1. Bracewell, R.: The Fourier Transform and its Applications, vol. 5. New York (1965)
2. Schiff, Steven J., Aldroubi, Akram, Unser, Michael, Sato, Susumu: Fast wavelet transformation of EEG. Electroencephalogr. Clin. Neurophysiol. **91**(6), 442–455 (1994)
3. Karvounis, E.C., Papaloukas, C., Fotiadis, D.I., Michalis, L.K.: Fetal heart rate extraction from composite maternal ECG using complex continuous wavelet transform. In: IEEE Computers in Cardiology, pp. 737–740 (2004)
4. Hazarika, N., Chen, J.Z., Tsoi, A.C., Sergejew, A.: Classification of EEG Signals Using the Wavelet Transform, IEEE, pp. 89–92 (1997)
5. Adeli, Hojjat, Zhou, Ziqin, Dadmehr, Nahid: Analysis of EEG records in an epileptic patient using wavelet transform. J. Neurosci. Meth. **123**(1), 69–87 (2003)
6. Subasi, A.: EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst. Appl. **32**(4), 1084–1093 (2007)
7. Horton, Paul, Nakai, Kenta: Better prediction of protein cellular localization sites with the it k nearest neighbors classifier. Ismb **5**, 147–152 (1997)
8. Peterek, T., Gajdos, P., Dohnalek, P., Krohova, J.: Human fetus health classification on cardiotocographic data using random forests. In: Advances in Intelligent Systems and Computing, vol. 298, pp. 189–198 (2014) (cited By 0)
9. Haralick, R.M., Shanmugam, K., Dinstein, IH.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **6**, 610–621 (1973)

10. Fujinaga, I., MacMillan, K.: Realtime recognition of orchestral instruments. In: Proceedings of the International Computer Music Conference, vol. 141, pp. 143 (2000)
11. Hess, K.R., Abbruzzese, M.C., Lenzi, R., Raber, M.N., Abbruzzese, J.L.: Classification and regression tree analysis of 1000 consecutive patients with unknown primary carcinoma. Clin. Cancer Res. **5**(11), 3403–3410 (1999)
12. Liaw, A., Wiener, M.: Classification and regression by randomforest. R news **2**(3), 18–22 (2002)
13. Gajdos, P. Moravec, P. Dohnalek, P., Peterek, T.: Mobile sensor data classification using gm-som. Adv. Intell. Syst. Comput. **210**, 487–496 (2013) (cited By 0)
14. Zhang, Q.Y., Aires-de Sousa, J.: J. Chem. Inf. Model. Random forest prediction of mutagenicity from empirical physicochemical descriptors. **47**(1), 1–8 (2007)
15. Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P.: Random forest: a classification and regression tool for compound classification and qsar modeling. J. Chem. Inf. Comput. Sci. **43**(6), 1947–1958 (2003)
16. Dohnálek, P., Gajdoš, P., Peterek, T.: Tensor modification of orthogonal matching pursuit based classifier in human activity recognition. In: Nostradamus 2013: Prediction, Modeling and Analysis of Complex Systems, of Advances in Intelligent Systems and Computing, vol. 210, pp. 497–505. Springer International Publishing (2013)
17. Levner, I.: Feature selection and nearest centroid classification for protein mass spectrometry. BMC Bioinf. **6**, (2005)
18. Dabney, A.R.: Classification of microarrays to nearest centroids. Bioinformatics **21**(22), 4148–4154 (2005)

Fuzzy c-Means Algorithm in Automatic Classification of EEG

Jitka Mohylova, Vladimir Krajca, Hana Schaabova, Vaclava Sedlmajerova, Svojmil Petranek and Tomas Novak

Abstract The electroencephalogram (EEG) provides markers of brain disturbances in the field of epilepsy. In short duration EEG data recordings, the epileptic graphoelements may not manifest. The visual analysis of lengthy signals is a tedious task. It is necessary to track the activity on the computer screen and to detect the epileptiform graphoelements and the other pathological activity. The automation of the process is suggested. The procedure is based on processing temporal profiles computed by means of multichannel adaptive segmentation and subsequent classification of detected signal graphoelements. The temporal profiles, function of the class membership in the course of time, reflect the dynamic EEG microstructure and may be used for visual indication of abnormal changes in the EEG using different colors. We will show that Fuzzy c-means (FCM) algorithm can be used for correct classification of epileptic pattern, creating homogeneous compact classes of significant EEG segments.

J. Mohylova (✉) · T. Novak
Faculty of Electrical Engineering and Computer Science,
VSB-Technical University of Ostrava, Ostrava, Czech Republic
e-mail: jitka.mohylova@vsb.cz

T. Novak
e-mail: tomas.novak1@vsb.cz

V. Krajca · H. Schaabova · V. Sedlmajerova
Faculty of Biomedical Engineering, Czech Technical University in Prague,
Prague, Czech Republic
e-mail: vladimir.krajca@nudz.cz; krajcvla@fbmi.cvut.cz

H. Schaabova
e-mail: hana.schaabova@fbmi.cvut.cz

V. Sedlmajerova
e-mail: Vaclava.Sedlmajerova@nudz.cz; vaclava.sedlmajerova@fbmi.cvut.cz

V. Krajca · S. Petranek
Department of Neurology, Hospital Na Bulovce, Prague, Czech Republic
e-mail: svojmil.petranek@nudz.cz; svojmil.petranek@bulovka.cz

V. Krajca · V. Sedlmajerova · S. Petranek
National Institute of Mental Health, Prague, 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 450, DOI 10.1007/978-3-319-33609-1_13

Keywords EEG graphoelements · Adaptive segmentation · Feature extraction · Cluster analysis · Fuzzy c-means

1 Introduction

In this paper we would like to present methods for automatic classification of EEG graphoelements which can be used for the detection and pattern recognition of EEG epileptic activity. The analysis and evaluation of long-term EEG recordings have gained increasing importance. One of the problems connected with the evaluation of EEG signals is the necessity of visual checking of such recording performed by a physician. Computer aided EEG analysis is a great help for physician. The present paper describes issue of application of advanced methods to the analysis of EEG signals. The work describes the design and implementation of a system, which performs an automatic analysis of EEG signals based on processing the time profiles computed by multichannel adaptive segmentation, subsequent feature extraction and classification of detected graphoelements of signal. The distribution of classes during a long recorded time shows the temporal profile (the function of class membership depending on time) and may be used for indication of abnormal changes in the EEG structure (seizures, epileptic spikes or focal epilepsy).

Fuzzy set theory has already been used in epileptic EEG processing. The review of the contemporary methods can be found in [1, 2].

The new idea is the combination of the adaptive segmentation for optimal feature extraction, fuzzy c-means algorithm for clusters homogeneity improving (and artefact rejection) and temporal profile for abnormal EEG segments identification.

The main difference compared to the other research teams is the classification of all EEG segments in many classes. We use 10 classes and other research teams (e.g. Acharya et al. [1], Krajca et al. [3]) usually use up to three classes for normal, ictal and interictal EEG. More classes help us distinguishing between different graphoelements.

2 Materials

The recordings were performed using the Brain-Quick (Micromed s.r.l.) digital system with sampling frequency of 128 Hz. We used ambulatory EEG recordings of fifteen adult patients with diagnosis of epilepsy and from two patients we used the records which were repeatedly recorded during the 3 years. Times of these records were last 20 min. The examinations were carried out in an EEG laboratory in Hospital Na Bulovce in Prague in standard conditions from nineteen active electrodes according to the international 10/20 system. The FIR filter (Finite Impulse

Response—Hamming window, order 845) was set as a bandpass (0.4 and 70 Hz) and the sensitivity was 100 μV per 100 mm. Electrode impedances were not higher than 5 kΩ.

3 Methods

The algorithm of signal processing of EEG is the off-line processing and it has five independent steps: (1) adaptive segmentation, (2) cluster analysis, (3) feature extraction, (4) classification and (5) temporal profile.

Signal processing is performed and evaluated in the WaveFinder software [2], basic software was developed by one of the authors (V.K.) and this part is now upgraded and extended with next modules [4].

3.1 Adaptive Segmentation

EEG recording has by its nature a non-stationarity character. Method of adaptive segmentation [5–9] is dividing the signal into quasi-stationarity segments of variable length, depending on the occurrence of non-stationarities in the signal. Technique of cutting the signal into fixed segments (length of segment 2 s) is inconvenient for our signal processing.

The used adaptive segmentation algorithm is based on two connected windows which are moving along the signal together in each channel. The change of stationarity is indicated by means of local maxima of their difference measure. The boundary of segment is set by measure of a difference measure composed of a frequency measure (estimated by the average difference of consecutive signal samples)

$$FDIF = \sum_{i=1}^{WL} |x_i - x_{i-1}| \tag{1}$$

and an amplitude measure of signal—*ADIF* (window length *WL*, x_i signal samples, for $i = 1, \ldots, 256$)

$$ADIF = \sum_{i=1}^{WL} |x_i| \tag{2}$$

The resulting difference measure is

$$DIF = k_A|ADIF_1 - ADIF_2| + k_F|FDIF_1 - FDIF_2| \tag{3}$$

where $ADIF_1$, $ADIF_2$ and $FDIF_1$, $FDIF_2$ are the amplitude and frequency measures of two connected windows computed according to Eqs. (1) and (2). Coefficients

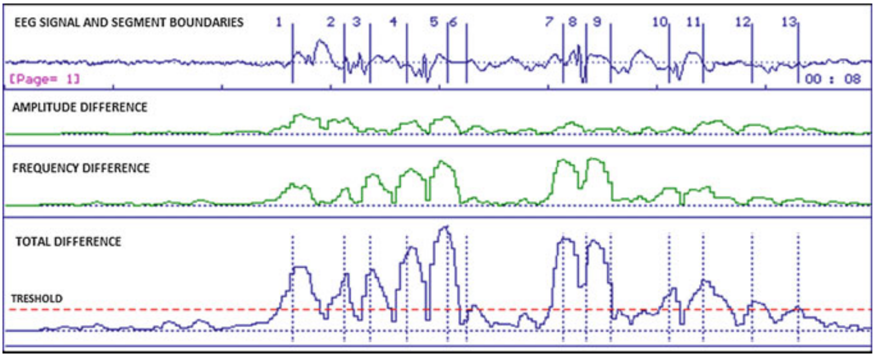


Fig. 1 A detailed example of adaptive segmentation of channel of EEG record. The *upper* part shows the signal (8 s) with segment boundaries. The *lower* part shows amplitude measure ADIF, frequency measure FDIF and total difference measure DIF. *Dashed line* represents the threshold for segment detection

were empirically set—k_A to 1 and k_F to 7. Next we used a threshold for the rejection of insignificant local maxima of the total difference measure. The threshold was computed adaptively from the incoming signal—BL is block of incoming data of length 8 s.

$$THR = \frac{1}{BL}\left(k_A|ADIF_1 - ADIF_2| + k_F|FDIF_1 - FDIF_2|\right) \quad (4)$$

We used another window for local maximum detection—GWL describes its length. The segment boundaries are represented by vertical lines and the numbers near segment boundaries denote the order of segments in this channel. The example of adaptive segmentation of one channel of EEG is shown in Fig. 1. Segment boundaries found by means of parameters: window length (WL) of 256 samples, window length for local maximum (GWL) of 32 samples, moving step was 1 sample and threshold (THR) for elimination of fluctuation in total difference measure was 0.68.

The adaptive segmentation is very useful for multichannel EEG records. Each channel is segmented by two connected windows sliding along each channel of EEG signal at a time independently. It means that the segment boundaries are positioned "asynchronously" in different places in different channels.

3.2 Feature Extraction

Next step of processing is computation of features. The performance of any method depends on a large extent on the extraction of the relevant features characterizing the EEG signals. We used the 24-dimensional feature vector, which was obtained

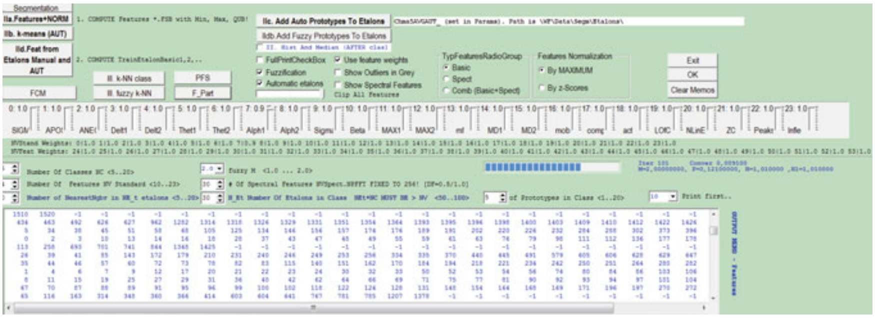


Fig. 2 An example of summary sheet of features for classification

from each segment of EEG signal. This vector formed the input into the classification step of the analysis.

This vector describes properties of the signal in frequency and time domain [10–12]. Frequency domain features: the power spectrum density (PSD) was estimated from the auto-correlation function of the signal and then FFT was taken with 2 s blocks (256 points), 50 % overlap and a Hanning window. Features in this domain are power in the frequency bands delta (0.5–3.5 Hz), theta (3.5–7.5 Hz), alpha (7.5–12.5 Hz) and beta1 (12.5–17.5 Hz), beta2 (17.5–25 Hz). For each band, absolute and relative band power were computed together with the total spectral power (TP) in the range 0.5–30 Hz. Peak frequency (PAF), bandwidth of dominant spectral peak (BW), weighted mean frequency (WMF) (Fig. 2).

Time domain features: variance of amplitude, difference between maximum positive and minimum negative segment sample, maximal value of the first derivative and of the second derivative of the signal, Hjorth parameters—activity (A), mobility (M) and complexity (C), zero crossings and others. Detailed description of the parameters can be found in [5, 8].

From above it is obvious that the proposed symptoms have different physical dimensions and therefore it is first necessary to standardize, thus it converted into a same scale, so they have unit variance and zero mean. We carried it out by dividing the maximum in all columns.

3.3 Fuzzy Clustering

It is not possible to use a classic set theory for an object which has class membership in several classes. Fuzzy c-means (FCM) is a method of clustering [10, 13–15] which allows one piece of data to belong to two or more clusters in the whole range of interval ‹0, 1›. The properties of objects of fuzzy sets can be practically interpreted by α-levels of fuzzy sets. They are sets of those objects that have fuzzy class membership u_i above some threshold α [10].

$$\alpha - level\ sets = \{x \in X: u_i(x) > \alpha\} \tag{5}$$

They can be used for eliminating the less typical class members with threshold above alpha.

Let $X = (x_{1,}\ x_2,\ \ldots\ x_N)$ denotes a set of N segments to be portioned into c clusters, where x_i represents multivariate data (features). The algorithm [10] is based on minimization of the following objective function:

$$J = \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij}^m \left\| x_j - v_i \right\|^2 \tag{6}$$

where u_{ij} represents the membership of segment x_j in the ith cluster, v_i is the cluster center, ‖ · ‖ is a norm metric (the Euclidean norm is used) and m is a constant. The parameter m controls the fuzziness of the resulting partition ($m = 2$ is used in this study). The cost function is minimized when pixels close to the centroid of their clusters high membership values are assigned and low membership values are assigned to pixels with data far from centroid. The membership function represents the probability that an EEG segment belongs to a specific cluster. In the FCM algorithm, the probability is dependent solely on the distance between the object and each individual cluster center in the feature domain. The membership function and cluster centers are updated by the following:

$$u_{ij} = \left[\sum_{k=1}^{c} \left(\frac{\left\| x_j - v_i \right\|}{\left\| x_j - v_k \right\|} \right)^{2/(m-1)} \right]^{-1} \tag{7}$$

and

$$v_i = \frac{\sum_{j=1}^{N} u_{ij}^m x_j}{\sum_{j=1}^{N} u_{ij}^m} \tag{8}$$

Starting with an initial guess for each cluster center, the FCM converges to a solution for v_i representing the local minimum or a saddle point of the cost function. Convergence can be detected by comparing the changes in the membership function or the cluster center at two successive steps.

The number of clusters is crucial and largely affects clustering. Optimal number of classes is estimated by partition coefficient (PC), see Table 1. The example in the table shows that the ideal number of classes for the case in Figs. 3 and 4 is ten. The partition coefficient, PC, is defined as [16]:

$$PC = \frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij}^2 \tag{9}$$

Table 1 Finding the ideal number of classes

Number of classes	Partition coefficient	Difference
2	0.7808	+0.7808
3	0.5679	−0.2130
4	0.5388	−0.0290
5	0.4619	−0.0769
6	0.3914	−0.0705
7	0.3428	−0.0486
8	0.3144	−0.0284
9	0.2918	−0.0226
10	0.2780	−0.0137
11	0.2562	−0.0218
12	0.2337	−0.0225

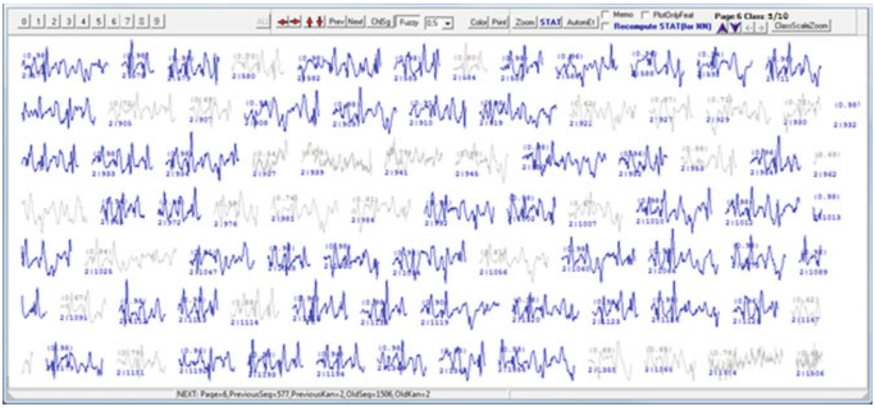


Fig. 3 Examples of class number five of fuzzy classification where graphoelements with low membership coefficient are eliminated. Fuzzy class membership number is written above each segment. The identification number of each segment is written below the segment, each segment is identified by the channel and position in time

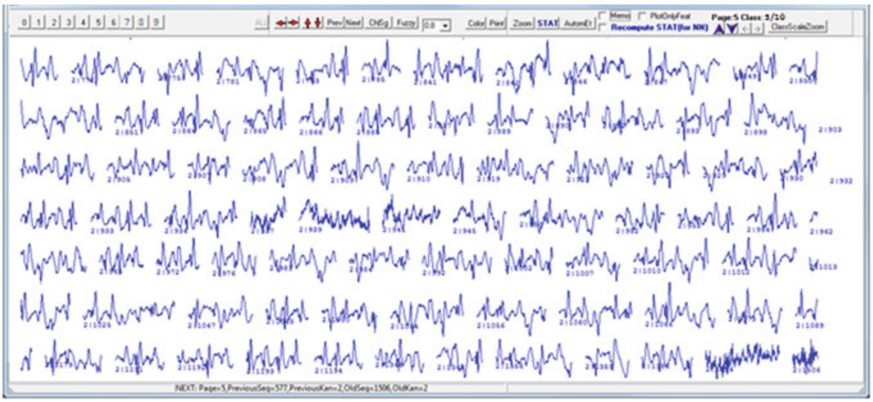


Fig. 4 Examples of class number five of fuzzy classification where graphoelements with higher membership

4 Results

The adaptive segmentation of multichannel EEG recording followed by the fuzzy c-means analysis creates more number of EEG classes with similar graphoelements. In Fig. 3 shows an EEG class number five. Threshold 0.5 was used; so that all segments were drawn including the nontypical (hybrid) segments which are eliminated and highlighted in grey color because of their membership lower than α-level. The fuzzy class membership number is written above each segment. The number channel and the order segment in this channel is written below each segment. Figure 4 shows same EEG class number five but their membership higher than α-level.

The classification that is present in this study allows the doctors to see the global class profile of long record for each patient. Each channel of EEG with adaptively segmented and fuzzy classified graphoelements can be reduced to a figure containing corresponding classes as shown in Fig. 4.

The temporal profile is very useful for displaying the distribution in time particular types of activity in the recording. The above steps of the analysis give us a segmented multichannel EEG signal and allow identification of EEG segment. That means we know the position of the segment boundary, the number of class to which the relevant segment belongs. We can distinguish the different classes by color—Fig. 5.

The temporal profile [17] is very useful for displaying the distribution in time particular types of activity in the recording. The above steps of the analysis give us a segmented multichannel EEG signal and allow identification of EEG segment. That means we know the position of the segment boundary, the number of class to which the relevant segment belongs. We can distinguish the different classes by

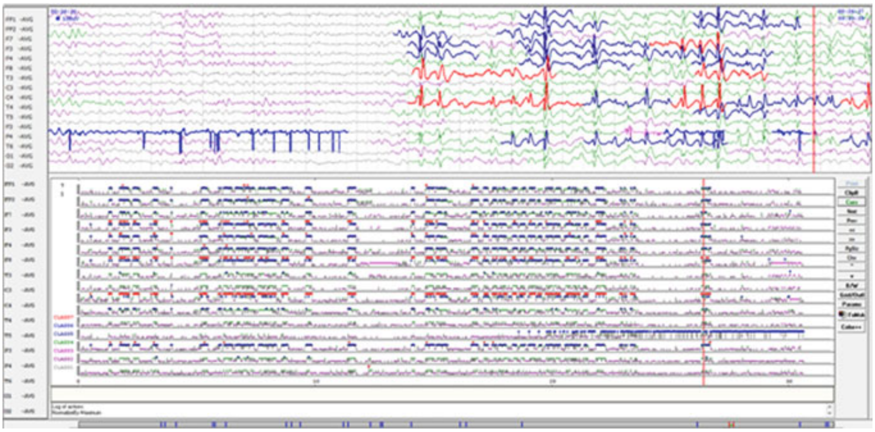


Fig. 5 Example of temporal profile of multichannel EEG after fuzzy classification algorithm; upper part demonstrates colored graphoelements and the lower part shows the global distribution of graphoelement classes. This approach reflects the macro-structure of the EEG signal

color—Fig. 5. The classification allows the doctors to see the global class profile of long record for each patient. This global representation of classes helps the doctors to clearly see the dominating classes in every part of the 24-h recording and simply slide between events.

5 Conclusion

Adaptive segmentation extracts the discriminative features much better than segmentation with fixed segment boundaries. Our approach of automated detection does not contain only few classes (normal and epileptic). Moreover, all graphoelements that are adaptively segmented are classified and colored according to their membership in the relevant class.

The approach of fuzzy c-mean algorithm for classification of all EEG graphoelements improves significantly the homogeneity of classes compared to the classical crisp classification algorithm.

The segments with low membership in the certain class are not left out, but just colored in grey. Due to the fact that they may contain some additional information for the doctor and the doctor can reclassify them into more relevant classes. That is why our approach is semiautomatic.

The global temporal profiles of long recordings can be plotted with clearly distinguishable macrostructures.

In future, cluster analysis will be used in the learning phase for etalon definition and the incoming object will be compared with the etalon (prototype). This way will be improved classification of different graphoelements.

Acknowledgments This work was supported by the Grant Agency of VSB-TU Ostrava No. SP2016/151 and by the Grant Agency of the Czech Technical University in Prague, No. SGS15/229/OHK4/3T/17.

References

1. Acharya, U.R., Sree, S.V., Swapna, G., Martis, R.J., Suri, J.S.: Automated eeg analysis of epilepsy: a review. Knowl.-Based Syst. **45**, 147–165 (2013)
2. Prasad, M., Lin, C.T., Yang, C.T., Saxena, A.: Vertical Collaborative Fuzzy C-Means for Multiple EEG Data Sets. Springer Lecture notes in Computer Science, vol. 8102, pp. 246–257 (2013)
3. Krajča, V., Petránek, S., Pietilä, T., Frey, H.: Wave-Finder: a new system for an automatic processing of long-term eeg recordings. In: Rother, M., Zwiener, U. (eds.): Quantitative EEG Analysis-Clinical Utility and New Methods. Universitätsverlag GnbH, Jena (1993), pp. 103–106
4. Schaabova, H., Krajca, V., Sedlmajerova, V., Bukhtaieva, O., Petranek, S.: Supervised learning used in automatic EEG graphoelements classification. In: E-Health and Bioengineering Conference (EHB). IEEE (2015), pp. 1–4

5. Krajca, V., Petranek, S., Patakova, I., Varri, A.: Automatic identification of significant graphoelements in multichannel EEG recordings by adaptive segmentation and fuzzy clustering. Int. J. Biomed. Comput. **28**(1), 71–89 (1991)
6. Barlow, J.S.: Methods of analysis of nonstationarity EEGs with emphasis on segmentation techniques a compare review. J. Clin. Neurophysiol. **2**(3), 267–304 (1985)
7. Barlow, J.S., Creutzfeldt, O.D., Michael, D., Houchin, J., Epelbaum, H.: Automatic adaptive segmentation of clinical eegs. Electroencephalogr. Clin. Neurophysiol. **51**(5), 512–525 (1981)
8. Krajca, V., Hozman, J., Mohylova, J., Petranek, S.: Pattern recognition of epileptic EEG graphoelements with adaptive segmentation, supervised and unsupervised learning algorithms. In: Biomedizinische Technik, Biomedical Engineering, BMT, Abstract of the Conference, 2012, vol. 57/08 (2012)
9. Lhotska, L., Krajca, V., Mohylova, J., Petranek, S., Gerla, V.: EEG processing. In: Data Mining and Medical Knowledge Management: Cases and Applications. IGI Publishing, Hershey (2009), pp. 161–180, ISBN 978-1-60566-218-3
10. Bodenstein, G., Praetorius, H.M.: Feature extraction from the electroencephalogram by adaptive segmentation. Proc. IEEE **65**(5), 642–652 (1977)
11. Havens, T.C., Bezdek, J.C., Leckie, C., Hall, L.O., Palaniswami, M.: Fuzzy c-means algorithms for very large data. IEEE Trans. Fuzzy Syst. **20**(6), 1130–1146 (2012)
12. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Springer Science & Business (1981)
13. Hsu, W.-Y., Lin, C.-Y., Kuo, W.-F., Liou, Michelle, Sun, Y.-N., Tsai, C.-H., Hsu, H.-J., Chen, P.-H., Chen, I.-R.: Unsupervised fuzzy c-means clustering for motor imagery eeg recognition. Int. J. Innov. Comput. Inf. Control **7**(8), 4965–4975 (2011)
14. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)
15. Chuang, K.S., Tzeng, H.L., Chen, S., Wu, J., Chen, T.J.: Fuzzy c-means clustering with spatial information for image segmentation. Comput. Med. Imaging Graph. **30**(1), 9–15 (2006)
16. Windham, M.: Cluster validity for fuzzy clustering algorithms. Fuzzy Sets Syst. **5**, 177–185 (1981)
17. Krajca, V., Petranek, S., Mohylova, J., Paul, K., Gerla, V., Lhotska, L.: Neonatal EEG sleep stages modeling by temporal profiles. In: Computer Aided Systems Theory-EUROCAST 2007. Springer (2007), pp. 197–201

Part III
Logical-Algebraic Methods and Reasoning Models for Intelligent Systems: Theory and Applications

Implementation of the Argumentation System Based on Defeasible Reasoning Theory

Andrew Derevyanko, Oleg Morosin and Vadim Vagin

Abstract This paper considers the software implementation of the argumentation system based on defeasible reasoning. Argumentation is used for finding and solving possible conflicts and inconsistence in data. The implemented software could be used in different intelligent systems, such as expert systems, decision support systems and machine learning. The main accent of this paper is made on the implementation and practical use of the argumentation system. The paper presents the architecture, main algorithms and examples of usage of the implemented system.

Keywords Argumentation · Defeasible reasoning · Justification degrees · Conflicts · Inconsistence

1 Introduction

The knowledge bases used in different intelligent systems are often obtained from unreliable sources and can contain inconsistent information. Classical methods of logical inference are not applicable to these knowledge bases. One way to detect and resolve internal conflicts is the use of the methods of argumentation.

By argumentation, we understand the process of analyzing the problem, in which we extend the set of assumptions (also called set of arguments), define the causal relationships between arguments, find conflicts between them and, if possible,

A. Derevyanko · O. Morosin (✉) · V. Vagin
National Research University "Moscow Power Engineering Institute",
Krasnokazarmennaya 14, 111250 Moscow, Russia
e-mail: omorsik@gmail.com

A. Derevyanko
e-mail: 777alterego777@gmail.com

V. Vagin
e-mail: vagin@appmat.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 450, DOI 10.1007/978-3-319-33609-1_14

determine the plausibility of the arguments. Typically, this process includes the detection of conflicts and the search for solutions. In contrast to classical logic, the argumentation presumes that there may be arguments as “for” and “against” certain assumptions.

We consider the architecture of the developed system of argumentation. We will present the main definitions of the argumentation system and will give the architecture of the implemented argumentation system. In conclusion, the paper provides example of the application of the system for solving the problem containing inconsistency.

2 Argumentation Theory

There exists many formalizations of the argumentation system: abstract argumentation systems [1], coherence systems [2], system based on defeasible logic [3, 4] and some others. In this paper we will consider the defeasible reasoning systems, as they are better suite for practical tasks. For the sake of completeness, in this section we will give the main definitions of the argumentation systems, based on defeasible reasoning and in Sects. 2 and 3 will consider the practical aspects of implementation of such system.

In defeasible reasoning systems, arguments are defined as chains of reasoning that lead from premises to a conclusion where each step in this chain can be rebutted. Arguments can be defeated because conflicts can be detected as new information arrives or even as new conclusions from the existing knowledge are obtained; i.e., as knew knowledge is obtained, earlier conclusions can turn out to be incorrect. An *argument* is a pair consisting of a set of premises and a conclusion. Such pairs are written as p/X, where p is the consequence and X is the set of premises. In all illustrations, we will present arguments as ovals. For arguments with the empty set of premises (such arguments are called facts), we will write only the conclusion. All arguments are considered as formulas of the first order logic.

The relationship between arguments are presented on an *inference graph*. Inference graph is a graph that shows the way of obtaining new arguments from already existing ones and shows the conflicts between arguments.

The concept of conflict is the foundation of the argumentation system. Consider two types of conflicts—*rebutting* and *undercutting* [3, 5]. Rebutting is a situation in which certain arguments rebut the conclusions of other arguments. Rebutting is a symmetric form of attack. *Undercutting* is an asymmetric form of attack in which an argument disclaims the connection between the premises and conclusion of another argument.

In the argumentation system based on defeasible reasoning, three types of inference rules are used—*deductive*, *defeasible*, and *undercutting rules*. Deductive rules are the inference rules claiming that if P is true then Q is true. Such rules are indefeasible. For such rules, we use the notation $P \Rightarrow Q$. In the inference graphs, they are presented by ordinary arrows (see the arguments P and Q in Fig. 1a). Defeasible

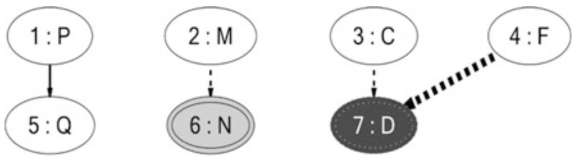


Fig. 1 Examples of inference rules

rules are the inference rules whose reliability is questionable. For example, such rules can be obtained by induction or abduction. The arguments obtained by such rules are called defeasible. Defeasible rules are denoted by $M| \Rightarrow M$. In inference graphs, they are depicted by dashed arrows, and defeasible arguments are depicted by double ovals (see the arguments *M* and *N* in Fig. 1b). Undercutting rules are the inference rules that formalize the conflict of the defeat type. These rules indicate that there are arguments that defeat the connection between two other arguments connected by a defeasible implication. For example, there is an argument *E* that undercuts the defeasible relationship $C \mid \Rightarrow D$ between the arguments *C* and *D*. Such undercutting rules are written in the form $E \Rightarrow (C \ @ \ D)$. In the inference graphs, the undercutting arguments and the arguments defeated by them are connected by a bold dashed arrow. The defeated arguments are shaded in dark grey (see argument D in Fig. 1c).

Detection of conflicts is the most difficult part in defeasible reasoning. Main difficulties arise from necessity of maintenance of the first order logic. Speaking of propositional logic it is clear that rebutting is a situation, when two contradicting arguments *A* and $\sim A$ exist. Speaking of the first order logic the main idea is the application of a unification mechanism for finding rebutting arguments. A substitution $U = \{t_1/x_1, t_2/x_2 \ldots t_n/x_n\}$, where t_i—term and x_i—variable, is called the *most general unifier* for expressions W_1 and W_2 if $W_1 * U = W_2 * U$, where $W_1 * U$ is the result of replacing variables $x_1, x_2, \ldots, x_n$ on terms $t_1, t_2, \ldots, t_n$ and it is considered that W_1 and W_2 do not have variables with the same names (if so it is necessary to rename them) [6]. For example, two expressions $P(x_1) \ \& \ G(x_2) \to H(x_1,x_2)$ and $P(f(a)) \ \& \ G(b) \to H(f(a),b)$ are unifiable with the unifier $U = \{f(a)/x_2, b/x_2\}$, and expressions $P(x) \to G(x)$ and $P(a) \ \& \ G(b)$ are not unifiable.

Two arguments $A_1 = p_1/X_1$ *and* $A_2 = p_2/X_2$ have a conflict of the type “rebutting” if exists such unifier U that:

1. $\sim p_1 * U = p_2 * U$, where $\sim p_1$ is the negation of the conclusion p_1.
2. $X_1 * U \in X_2 * U$ or $X_2 * U \in X_1 * U$.

An argument A_1 is the undercutting argument for arguments A_2 and A_3 if:

1. There is a defeasible link between A_2 and A_3.
2. There is the undercutting rule $E \Rightarrow (C@D)$ E(C@D) and there are unifiers U_1 and U_2 such that:

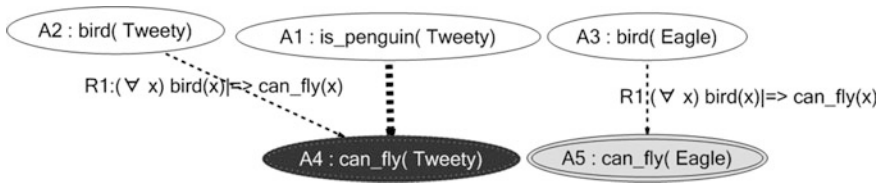


Fig. 2 Example of finding conflicts

a. $E * U_1 = A_1 * U_1$
b. $(C * U_1) * U_2 = (A_2) * U_2$
c. $(D * U_1) * U_2 = (A_3) * U_2$

Let's illustrate the application of unification for finding conflicts on a small example.

Suppose we have the following arguments and inference rules:

A_1: *is_ penguin*(*Tweety*), A_2: *bird*(*Tweety*), A_3: *bird*(*Eagle*);
R_1: (∀ *x*) *bird*(*x*) | =>*can_fly*(*x*), R$_2$: (∀ *x*)*is_penguin*(*x*) =>(*bird*(*x*)@ *can_fly* (*x*)).

The inference graph of this example is presented in Fig. 2.

In this example arguments A_1, A_2 and A_4 suits for the undercutting rule R_2, as exists:

1. the unificator U_1 = {Twetty/*x*}, such as (∀ *x*)*is_pengion*(*x*) · U_1 = A_1:is_pengion(Twetty) · U_1;
2. the unificator U_2 = { }, such as (*bird*(*x*) · U_1) · U_2 = (A_2: *bird*(*Twetty*)) · U_2 and (*can_fly*(*x*) · U_1) · U_2 = (A_4: *can_fly* (*Twetty*).

Therefore, argument A_1 has a conflict with argument A_4 and it should be considered as defeated.

On the contrary, for arguments A_1, A_3 and A_5 there exists the unificator U_1 = {*Twetty*/*x*}, such as (∀ *x*)*is_pengion*(*x*) · U_1 = A_1:*is_pengion*(*Twetty*) · U_1, but there are no such unificator U_2, that (*bird*(*x*) · U_1) · U_2 = (A_3: *bird*(*Eagle*)) · U_2 and, consequently, undercutting rule R_2 is not applicable for arguments A_1, A_3, A_5 and argument A_5: *can_fly*(*Eagle*) remains undefeated.

Now let us turn to the main problem—finding statuses of arguments. At each step of reasoning, finding the status of arguments (defeated it or not) plays a main role.

Argument is called the ***initial*** one, if the set of its ancestors is empty, i.e. it means it is set initially.

An ***argument basis*** is a set of arguments involved in inference of an argument.

Let us give the definition of a function, that assigns statuses defeated or not to the graph nodes [7].

The function σ assigns a provisional status to arguments, giving value *defeated* or *undefeated* to a subset of nodes(arguments) of the graph in such a way that:

- σ assigns the status *undefeated* to all initial nodes.
- σ assigns the status *undefeated* to the node *n* iff σ assigns the status *undefeated* to all nodes of the node *n* basis and σ assigns the status *defeated* to all nodes that attacks the node *n*.
- σ assigns the status *defeated* to the node *n* iff either some nodes of the basis of *n* has status *defeated*, or the node *n* is attacked by some node with status *undefeated*.

The status assigned by σ is the final status of the argument n, if σ assigns a provisional status to n and σ is not involved in the defining of statuses of the other arguments related with n.

So, to sum up there are three available status assignments—*undefeated*, *defeated* and *provisionally defeated*. The first status *undefeated* means that there are no reasons for not believing this argument. The second one *defeated* means that there are valuable reasons for not believing it. The third status *provisionally defeated* means that at this stage of reasoning there are no significant data for believing or denying an argument. In particular that means that there are two (or more) arguments attacking each other (a collective defeat in Pollack's terminology), and there is no information to solve this conflict.

2.1 Justification Degrees in Defeasible Reasoning

Justification degrees allows to give quantitative estimation to arguments plausibility. For justification degrees we use the numerical scale [0, 1], where 0 corresponds to a defeated argument and 1 corresponds to the most justified argument.

We need to define a function *Jus*(*A*) for calculating any argument in the inference graph. We assume that the justification degree is already defined for the initial arguments. The value of this function is affected by two factors—the inference tree of the argument (i.e., the justification degrees of the arguments used in the inference of the argument in question) and conflicts with other arguments.

For convenience, consider these factors separately: $Jus_{anc}(A)$ is the inherited justification degree, and $Jus_{con}(A)$ shows how much the conflict undercuts the argument justification. The inherited justification degree is the measure of justification of the arguments involved in the inference of *A*. It is clear that if unreliable arguments are used to prove an argument, then the reliability of the last argument is low. It is proposed to use the weakest link principle [7, 8], which states that the inherited justification degree of an argument *A* cannot be greater than the justification degree of the weakest argument involved in the inference of *A*.

$$Jus_{anc}(A) = \min(\{Jus(A_1), Jus(A_2) \ldots Jus(A_\mathrm{n})\}), \quad (1)$$

where $A_1, A_2 \ldots A_n$—arguments that were used in the inference of argument *A*.

Note that formula (1) implies that, if an argument has only one ancestor, then its inherited justification degree is equal to the justification degree of this ancestor.

In the calculation of $Jus_{anc,}$ we seek the weakest arguments; however, in order to find out how much a conflict reduces the justification degree, we use the strongest arguments. Let's consider A_{confl} as a set of arguments that has conflicts with the argument A and $Aconfl_i$ is a member of this set.

$$Jus_{con}(A) = \begin{cases} Max(\{Jus_{anc}(Aconfl_1), Jus_{anc}(Aconfl_2), \ldots, Jus_{anc}(Aconfl_n)\}), |Aconfl| > 0 \\ 0, otherwise \end{cases} \tag{2}$$

In formula (1) Jus_{anc} is used to avoid cases of mutual conflicts between two arguments.

Thus, the justification degree of an argument A could be calculated as follows:

$$Jus(A) = \begin{cases} Jus_{anc}(A) - Jus_{con}(A), Jus_{anc}(A) > Jus_{con}(A) \\ 0, \quad otherwise \end{cases} \tag{3}$$

It is necessary to note, that this is not the only way of calculating justification degree, e.g. in some works Bayesian approach is applied [9].

2.2 An Algorithm for Calculating the Degrees of Justification

Input data: *Inference graph G.*
Output data: Modified inference graph G' with calculated justification degrees of the arguments.

Step 1. Add to the queue *Qlist* all arguments in graph *G,* starting with initial arguments.
Step 2. While *Qlist* is not empty extract the next element q from it and execute *Justification*(q).
Step 3. Change the justification degrees of arguments in *G* according to found degrees in step 2.

Procedure ***Justification***(***argument q***, ***[argument ex]***);

Input: argument q, optional parameter—argument *ex*. (If *ex* is given, then it's justification degree is ignored when calculating the justification degree of q).
Output: justification degree of argument q.

Step 1. If q is initial argument, then set $Jus_{anc}(q)$ equal to the initial justification degree of an argument. Otherwise, find all arguments Anc_i, that are direct ancestors of argument q. Set $Jus_{anc}(q) = min\{Jus(Anc_1), Jus(Anc_2), \ldots, Jus(Anc_k)\}$, where k is the number of direct ancestors of argument q.

Step 2. If argument q doesn't have conflicts with other arguments, then set $Jus_{con}(q) = 0$. Otherwise, construct set *Acon*, consisting of arguments conflicting with q. If parameter *ex* is not null, than exclude it from the set *Acon*. If the justification degree of all arguments in *Acon* is already calculated, then set $Jus_{con}(q) = max\{Jus(Acon_1), Jus(Acon_2), \ldots, Jus(Acon_m), Jus_{anc}(ex)\}$, where m is the number of arguments, conflicting with argument q, $Acon_i$—elements of *Acon*. Go to step 4. If for some arguments in *Acon* justification degree is not calculated, then go to step 3.

Step 3. For each $Acon_i$ with undefined justification degree execute the following:

Step 3.1. If $Acon_i$ undercuts q, then execute $Justification(Acon_i)$, otherwise if $Acon_i$ has a symmetrical conflict with q, then execute $Justification(Acon_i, q)$, i.e. calculate justification degree of $Acon_i$ not taking into account the argument q. Return to step 2.

Step 4. Set $Jus(q) = Jus_{anc}(q) - Jus_{con}(q)$, if $Jus_{anc}(q) > Jus_{con}(q)$, otherwise set $Jus(q) = 0$.

Step 5. Change the justification degrees of arguments in inference graph and stop algorithm.

The computational complexity of the proposed algorithm for a fixed set of arguments $\mathbb{A}$ consisting of n arguments and containing k conflicts is $\underline{O}(kn)$. Since the number of conflicts is small compared with the number of arguments, the complexity of the algorithm is equal to the $\underline{O}(n)$.

3 Implementation

The described above argumentation system based on defeasible reasoning was successfully implemented as a standalone application in Visual Studio 2012, C#. The architecture of the implemented software is presented in the Fig. 3.

Let us give a brief description of the main modules.

The syntactical analyzer produces the initial transformation of input raw text data format into an internal object representation.

The format of the inputted data can be formulated in the following Backus–Naur form:

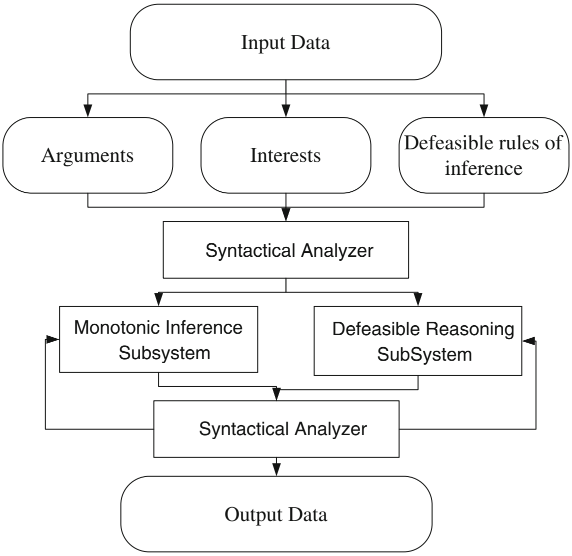


Fig. 3 Structure of argumentation system

<Argument>::=<BoolExpression>|<Quantors_list>(<BoolExpression >)

<Quantors_list>::=<Quantor> <Variable>

<Quantors_list>::=<Quantor><Variable><Quantors_list>

<Quantor>::=∀ | ∃

<BoolExpression>::=<BoolExpression>|¬<BoolExpression>|(<BoolExpression><LogOp><BoolExpression >)

<BoolExpression>::=<Name>|<Name>(<Variable_list>)

<Variable_list>::=<Variable>|<Variable>, <Variable_list>

<Variable>::=<Name>

<LogOp>::=’&’ | ‘|’ | ‘↔’ | ‘→’

< Name>:: =<Letter>|<Letter><LetterOrDigitS>

<LetterOrDigitS>::=<LetterOrDigit>|<LetterOrDigit>< LetterOrDigitS>

<Letter>::=a|b|c|…z|A|B|C|…Z|

<LetterOrDigit>::a|b|c|…z|A|B|C|…Z|1|2|3..0.

For instance, the argument can be written so: $\forall x\ \exists y\ (big\_town(x) \rightarrow (small\_town(y) \& near(x,y)))$.

(1) Grammar for deductive rules:

<Conclusive_reason>:: {<Quantors_list><ListOfBoolExpression >}⇒
<Quantors_list>< BoolExpression >

<ListOfBoolExpression >::
<BoolExpression >|<BoolExpression >,< ListOfBoolExpression>

Example of deductive rule: $\{A,B\} \Rightarrow C$.

(2) Grammar for defeasible rules:

<Quantors_list><PrimaFaccie_reason>::{<ListOfBoolExpression >} |⇒ <Quantors_list>< BoolExpression >

Example of defeasible rule $\{A\}$ |⇒ C.

(3) Grammar for undercutting rules:

<Quantors_list><Rebutting_reason>::{<ListOfBoolExpression >} ⇒ (< BoolExpression >@< BoolExpression >)

Example of undercutting rule $\{B\} \Rightarrow (A@C)$.

The Monotonic Reasoning Subsystem provides an ability of the system to operate with monotonic arguments and to infer new arguments from the existing ones [10]. The monotonic reasoning subsystem is based on the theory of natural deduction. The main modules of this subsystem is presented in Fig. 4.

The Defeasible Reasoning Subsystem is used for dealing with defeasible rules of inference, finding conflicts and calculating defeat statuses and justification degrees. The structure is presented in Fig. 5.

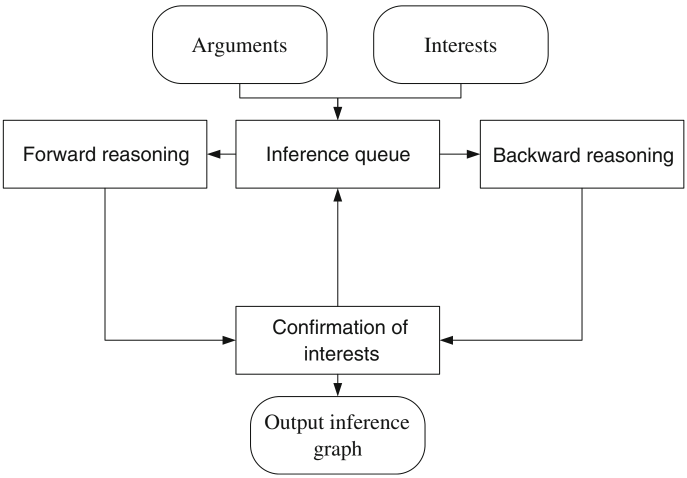


Fig. 4 Monotonic Reasoning subsystem

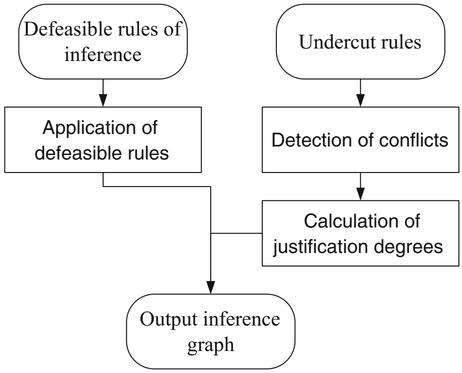


Fig. 5 Defeasible reasoning subsystem

4 Example

There is a company CMP that is engaged in IT business. According to statistics, IT market is on the rise and 75 % IT companies show a positive trend of development. CMP is the company that develops software for operating systems Linux. Only 10 % of users use Linux operating system and, consequently, the demand for products is not great. However, 80 % of the production company aimed at the international market of software for Web servers, which are mostly run on Linux. The businessman wants to invest a certain amount in the growing business and receive income. Should he invest in CMP company?

Let's rewrite this statements formally using the first order logic notation.

A1. IT(*CMP*);
A2. for_Linux(*CMP*);
A3. market(*CMP,Server_side*);
A4.∀*x* (*invest*(*x*)& *will_grow*(*x*) > *income*);
R1. ∀*x IT*(*x*)| => *will_grow*(*x*);
R2. ∀*x* (*will_grow*(*x*)) | => (*invest*(*x*));
R3. ∀*x* (*for_Linux*(*x*)) | => (*poor_demand*(*x*));
R4. ∀*x poor_demand*(*x*) | => (~*invest*(*x*));
R5. ∀*x* (*market*(*x,Server_side*)) =>(*for_Linux*(*x*) @ *poor_demand*(*x*)))

Inference graph for this example is shown in Fig. 6.

This example presents the problem that is not solvable in the classical logics. It has two conflicts of types "rebutting" and "undercut". Different rules were used for constructing arguments—rules of natural deduction, rules that helped us to cope with quantified expressions and defeasible rules of inference. These rules were built from textual task description and entered to the program complex in a formal way. Finally, we find that the interests of the study confirmed with a degree of 55 %. That is an investment in the company CMP generate revenue with a probability of 55 %.

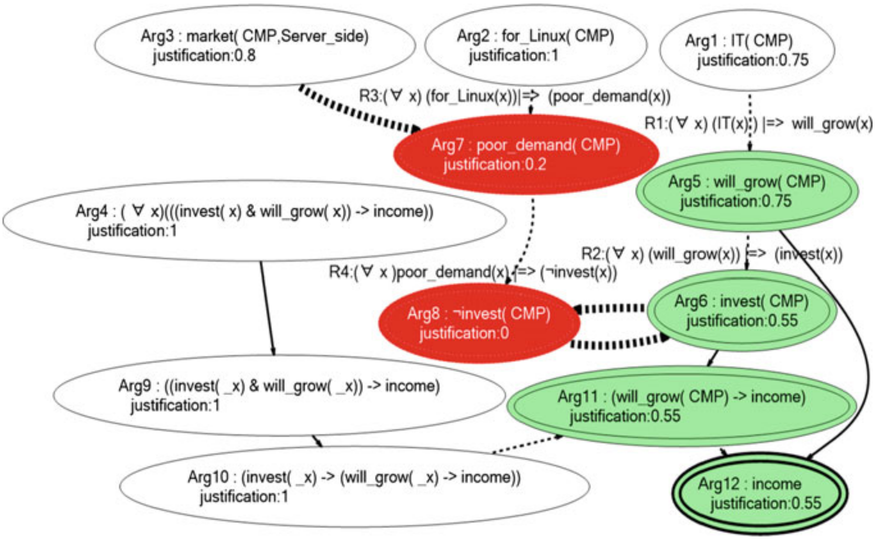


Fig. 6 Inference graph for example

5 Conclusions

In this paper we given an overview of main principles and ideas that were used for implementation of the argumentation system, based on defeasible reasoning theory. The developed system were tested on different benchmarks [10, 11], which showed that it is sound and correct. Speaking of practical aspect, the implied system could find application in different intelligent systems; in particular, it was shown [12] that it could be very useful in machine learning and inductive concept formation. The developed software system could be considered as a set of independent components: monotonic and defeasible inference subsystem, justification degree calculation subsystem, parsing module, module of graphical representation of the argumentation problem. Such module system allows flexible use of the developed components as a part of complex intelligent systems, such as expert systems, decision support systems and others.

Acknowledgments This paper was prepared under the financial support RFBR Grant No. 16-37-00309.

References

1. Bondarenko, A., Dung, P.M., Kowalski, R., Toni, F.: An abstract, argumentation-theoretic approach to default reasoning. J. Artif. Intell. **93**, 63–101 (1997)
2. Besnard, P., Hunter, A.: Elements of Argumentation. MIT press, New York (2008)

3. Pollock, J.L.: How to Reason Defeasibly. J. Artif. Intell. **57**, 1–42 (1992)
4. Nute, D.: Defeasible reasoning and decision support systems. J. Decision support systems **4**, 97–110 (1988)
5. Pollock, J.L.: Defeasible reasoning. J. Cogn. Sci. **11**(4), 481–518 (1987)
6. Vagin, V., Golovina, E., Zagoryanskaya, A., Fomina, M.: Exact and Plausible Inference in Intelligent Systems, Vagin, V., Pospelov, D. (eds.) FizMatLit, Moscow (2008) (in Russian)
7. Pollock, J.L: Oscar—A general purpose defeasible reasoner. J. Nonclassical Logics **6**(1), 89–113 (1996)
8. Betz, G.: On degrees of justification. J. Erkenntnis **2**, 237–272 (2012)
9. Haenni, R., Kohlas, J., Lehmann, N.: Probabilistic Argumentation Systems. Springer, Dordrecht (2000), pp. 221–288
10. Pollock, J.L.: Natural deduction. http://johnpollock.us/ftp/OSCAR-web-page/PAPERS/Natural-Deduction.pdf. Accessed 04 Oct 2015
11. Vreeswijk, G.A.W.: Interpolation of benchmark problems in defeasible reasoning. In: Second World Conference on the Fundamentals of Artificial Intelligence, pp. 453–468, Paris (1995)
12. Vagin, V., Fomina, M., Morosin, O.: Argumentation in inductive concept formation. In: 9th International Conference on Application of Information and Communication Technologies. IEEE (2015), pp. 133–137

Paraconsistency of Argumentation Semantics for Stepping Theories of Active Logic

Igor Fominykh and Michael Vinkov

Abstract Active logic is a conceptual system, the principles of which are satisfied by the formalism of reasoning, enabling them to correlate the results with the timing and which are tolerant to inconsistencies. However, tolerance to the contrary not yet received a theoretical substantiation in the form Paraconsistent semantics. The work proposes a variant of argumentation semantics for formalism of stepping theories built on the principles of Active Logic and logic programming. It is shown that the proposed argumentation semantics are Paraconsistent in the meaning of existence of contradictions in the stepping theories does not lead to their destruction, as is the case in standard logic systems.

Keywords Active logic · Time considerations · Stepping theories · Paraconsistent semantics

1 Introduction

Active Logic [1, 2] is a conceptual system that combines a number of formalism of so-called reasoning in time, which are considered not as a static sequence of statements, but following in time process. The principles embodied in systems of Active Logic relevant in solving the problems of managing complex objects in hard

This work was supported by the RFBR.

I. Fominykh (✉)
Moscow Power Engineering Institute, Moscow, Russia
e-mail: igborfomin@mail.ru; gborfomin@mail.ru

M. Vinkov
Bauman Moscow State Technical University, Moscow, Russia
e-mail: vinkovmm@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 450, DOI 10.1007/978-3-319-33609-1_15

real-time mode. These problems are characterized by situations where exceeding the allowable time allotted to their decision will lead to disastrous consequences instead of gradual degradation of performance. In addition, the information received in the course of solving such problems is often inconsistent and requires special handling techniques. Reasoning satisfying the concept of Active Logic have two important properties: (1) the final and intermediate results can be correlated to time and the reasoning agent is able to be aware of this (this property was called temporal sensitivity) and (2) these arguments are tolerant to the contradictions that arise in the course of their implementation (the arguments that have this property are called *paraconsistent*). But if the provision of the temporal sensitivity in systems of Active Logic was not associated with serious theoretical problems, the attempt to create a declarative semantics for such paraconsistent logical systems have not been successful, as was shown in [2].

In this report, issues relating to the paraconsistency of systems of Active Logic will be considered in relation to one of the variants of embodiment of this concept, the formalism of so-called stepping theories introduced in [3]. A specific feature of this formalism is the fact that it combines the principles of Active Logic and logic programming. It is known that the concept of logic programming is based on the semantics of its related non-monotonic formalisms in terms of sets of literals, rather than the more complex structures in the style of Kripke, as is the case in other non-monotonic systems. This fact has a beneficial effect on the computational complexity of logic programs. By combining the principles underlying these concepts nonmonotonic formalism was received corresponding to concept of Active Logic and, as it seems, is easier to implement and has the best features of the computational complexity compared to other Active Logic systems known at the present time.

In [4, 5] argumentation semantics has been defined for the formalism of stepping theory. The theory of argumentation [6] has been very fruitful in the presentation of non-monotonic reasoning [7]. It was stated that some well-known non-monotonic formalisms correspond to different individual cases of argumentation structure. In terms of the theory of argumentation for them it was managed to get a very elegant definition of the relation of logical consequence. This report shows that argumentation semantics proposed for the formalism of stepping theories is paraconsistent.

2 Logical Paraconsistency

Paraconsistent logic—a section of modern non-classical logic, in which the logical principle that allows to derive the arbitrary sentence from the fact of logical contradiction, does not hold. In classical logic a theory called *contradictory* when both a sentence and its negation can be proved. If an arbitrary proposal can be proved in this theory, the theory called *trivial*. In conventional systems, the logics of the concept of triviality and inconsistency do not differ, i.e. there is a contradiction in theory leads to its triviality. Paraconsistent logic treats inconsistency differently than classical logic. The possibility to derive from the contradictions of any proposals is ruled out in it, thereby ceases to be a contradiction of the theory of the threat of destruction. However, this does not resolve the fundamental need to get rid of contradictions in the course of further development of the theory. This implies another definition of paraconsistent logic, somewhat less general than the previous one: the logic called paraconsistent if it can be the basis of contradictory, but not trivial theories.

The strict definition of paraconsistent logic is associated with characteristic of relation of logical consequence. It can be called explosive, if it satisfies the condition that for any formulas A and B, A and Non-A will be an arbitrary to formula B (symbolically: $\{A, \neg A\} \vdash B$). Classical logic, intuitionistic logic, multi-valued logic and most other standard logic are explosive. Logic is called paraconsistent if and only if its ratio of logical inference is not explosive.

The need for the development of contradictory, but non-trivial theory was the impetus for the emergence of paraconsistent logic (the theory is called non-trivial if and only if it is not trivial). Standard systems of logic do not separate concepts of logic contradictions of the concept of triviality, i.e. a contradiction in theory leads to its triviality. Hence there is another definition of paraconsistent logic, somewhat less general than the previous one: the logic called paraconsistent if it can be the basis of contradictory, but non-trivial theories. That is a definition that was first time in the literature given by the Polish logician Jaśkowski [8] and independently by the Brazilian logician Da Costa [9]. Thus, paraconsistent logic allows you to "localize" the action of the contradictions in the sense that the existence of contradictions in the theory does not lead to the destruction of the last, in a sense, is the realization of the thesis of the non-universality of the law of contradiction.

3 Explosive Semantics for Theories of Stepping Active Logic

One of the Active Logic systems, as mentioned above, is the formalism of so-called stepping theories, which is a set of rules given to them with a binary relation of preference. In this section, for the purposes of this work will be introduced semantics, which is explosive or, in short, E-Semantics (E-explosive) in the meaning defined in the previous section. Note that the semantics of logic programming formalisms, where used classical negation (not just the denial by default), for example, sets the semantics of the responses in the formalism of extended logic programs [10], also is explosive.

Sets of rules of stepping theories are divided into two subsets—rigorous and credible rules. In the future, without sacrificing a community, we will be limited only to cases where the ratio of preference is empty, and there are only plausible theory of rules in form of

$$N:\ a_1,\ a_2,\ \ldots,\ a_m \Rightarrow b, \qquad (1)$$

where N—a character string that indicates the name of the rule, b—propositional literal, a_1, …., a_m—propositional literals or literal of language of first-order logic type later (j), where j—integer (in the antecedent rule literal latter species can be no more than one). Furthermore, for any literal, for example, q will denote its complement of contraries as a couple—q.

Rules express the principle of negative introspection in the following interpretation: if the formula a_1, a_2, …, a_m implements and at this stage the output is not known whether the formula (a literal)—b, one may assume that the formula b implements. For simplicity, the antecedents of the rules are considered as a set of literals. If this set is empty, such rules are considered to be equivalent to the rules with the same consequents, but in the antecedent of which there is a single literal later (0). Thus, this system can be considered as a variant of Active Logic entirely based on the rules and meet the basic principle of logic programming (model formulas are sets of literals).

Stepping theory will be called the theory of pair T = (R, >), where R—the final rule in the form of (1), > is an acyclic relation preference defined on the set R, and later in this report, as has been said, for the sake of simplicity, but not at the expense of community, is considered empty. For any stepping theory T = (R, >) will be denoted by R [q] set of rules, with the consequent q. Many literal constituting the antecedent rule r, denoted by A (r). Let Lit_T be the set of all literals found in the rules of stepping theory of T. *The set of beliefs* of stepping theory T = (R, >) is the set in the form of {now (t)} ∪ L_T^t, where t—positive integer or 0, interpreted as a certain moment of time, $L_T^t \subseteq Lit_T$. Consider the operator $\vartheta^E{}_T$ corresponding to E-semantics and converting sets of beliefs into sets of beliefs, so that if B—set of beliefs, such that now (t) ∈ B, then (1) now (t + 1) ∈ $\vartheta^E{}_T$ (B) and (2) sufficient conditions of belonging for the literal q to the set of beliefs $\vartheta^E{}_T(B)$ are shown below:

1. q ∈ B and q ≠ now (t), where t—any positive integer or 0.
2. There exists a rule r ∈ R [q], that for any literal p ∈ A (r) is done strictly one of two conditions: (1) p ∈ B or (2) if p = later (t) and now (t′) ∈ B, then $t \le t'$.
3. If there is a p ∈ B and—p ∈ B.

The latter condition determines that the entered E-semantics is explosive, i.e., if p ∈ B and—p ∈ B, then $q \in \vartheta^E{}_T(B)$, where q—any of the plurality of literal Lit_T, which determines the language current stepping theory. The formulas of this language are literals of this set. Note that the trivial stepping theories in terms of E-semantics are only those in which there is a conclusion of a pair of contraries literal, and, not necessarily both literals at the same time. Such stepping theories will be called *contradictory* and *ultimately trivial*, referring to the fact that the process of reasoning, formalized by stepping theories, as opposed to the standard of excessive logic, develops in time.

Let B be a set of beliefs of theory T, such that literal now (t) ∈ B. Let's call B as *quasi-fixed point* of $\vartheta^{E}{}_{T}$ operator if and only if (later by signed as ⇔) B \{now (t)} $= \vartheta^{E}{}_{T}$ (B) \{now (u)} for any u > t. *History* in stepping theory T is a finite sequence B of the set of beliefs. Wherein ***B*** (i) denotes the i-th element in the history, ***B*** (0) = {now (0)} and for any t $B(t+1) = \vartheta^{E}{}_{T}(B(t))$. The final element of the history is the set of beliefs, denoted B_{fin}, (*final*). It is the lowest quasi-fixed point of $\vartheta^{E}{}_{T}$ operator such that if now (n) ∈ B_{fin}, then for any j, if later (j) ∈ A (r), where r —a rule of a plurality of R, then n > j. *Inference step* of stepping theory T = (R, >) is every pair of the form (***B*** (i), ***B*** (i + 1)), wherein number equal to (i + 1) will be called *inference step number*. Every literal q, such that q ∈ B_{fin} (q ∈ ***B*** (t)) will be called the *consequence* (t-*consequence*) of the stepping theory T.

Stepping theories T and T′ are called *equivalent* ⇔ any result (t-result) of theory T is a consequence (t-consequence) of theory T'. The following two examples illustrate, respectively, a trivial, but not contradictory stepping theory and contradictory stepping theory that is ultimately trivial.

Example 1 Let set R_1 of stepping theory T_1 = (R_1, Ø) consists of the following elements:

N1: ⇒ p
N2: ⇒ q
N3: q ⇒ r

History of this theory is:

$\boldsymbol{B}_1$ = < $\boldsymbol{B}_1$ (0) = {now(0)}, $\boldsymbol{B}_1$ (1) = {now(1), p, q,}, $\boldsymbol{B}_1$ (2) = B_{fin} = {now (2), p, q, r}>.

In this case, the theory is not contradictory, although $\boldsymbol{B}_1$ (2) = Lit_{T1}

Example 2 Now consider a stepping theory T_2 = (R_2, Ø), wherein R_2 consists of the following elements:

N1: ⇒ p
N2: ⇒ q
N3: q ⇒ r
N4: later (3), p ⇒ ¬ r
N5: s ⇒ o

History of this stepping theory is:

$\boldsymbol{B}_2$ = <$\boldsymbol{B}_2$ (0) = {now(0)}, $\boldsymbol{B}_2$ (1) = {now (1), p, q}, $\boldsymbol{B}_2$ (2) = {now (2), p, q, r}, $\boldsymbol{B}_2$ (3) = {now (3), p, q, r} $\boldsymbol{B}_2$ (4) = {now (4), p, q, r, ¬ r}, $\boldsymbol{B}_2$ (5) = B_{fin} = {now (5), p, q, r, ¬ r, s, o}, B_{fin}\ {now (5)} = Lit_{T2}. Inference step 3 did not lead to a change in the set of beliefs $\boldsymbol{B}_2$ (2), except now (.)—in this step none of the rules of the R_2 "fired". This stepping theory is contradictory and ultimately trivial. Note that literals s and o got into the set of beliefs $\boldsymbol{B}_2$ (4) = B_{fin}, despite the fact that the rule N5 did not "fire." This was due to explosiveness of E-semantics.

4 Paraconsistent Argumentation Semantics for Stepping Theories of Active Logic

Argumentation system is usually have some *logic language* and *definition of the argument, the conflict between the arguments* and *the status of the argument* as basic elements. The last three items are often used to define the relationship of logical consequence. Below will be presented argumentation system consisting of said elements, built using the specific of stepping theories of Active Logic, and will be shown that the semantics of this system is paraconsistent, in the sense that the stepping theory, which are contradictory from the standpoint of E-semantics, are not trivial from the point of view of argumentation semantics.

Definition 1 Let T = (R, >) be a stepping theory. As argument for the T will be called 1) every literal h (of first-order logic) as later (t), where $t > 0$, for which there exists a rule $r \in R$, such that $h \in A(r)$, or 2) the sequence of the rules Arg = $<r_1,\ldots, r_n>$, where $r_1,\ldots, r_n \in R$, such that for every $1 \geq i \geq n$, if $p \in A(r_i)$, where p—a propositional literal, then there is exactly one $j < i$ that $r_j \in R[p]$.

For this stepping theory T = (R, >) the *set of all its arguments* will be denoted $Args_T$. If the argument is a literal first-order logic type later (t), then we say that such argument is a *limiting argument* (the effect of other arguments below in time). Otherwise, the argument is called a *supporting argument*. Propositional literal q is a *conclusion* of supporting argument Arg = $<r_1, \ldots, r_n> \Leftrightarrow r_n \in R[q]$.

Supporting subargument of argument Arg = $<r_1, \ldots, r_n>$ is any subsequence $<r_1,\ldots,r_n>$ that satisfies the Definition 1. Limiting argument is subargument of argument Arg = $<r_1,\ldots,r_n>$ if the corresponding literal is a part of the antecedent of any of the rules $r_1,\ldots,r_n$.

Every supporting subargument of argument Arg = $<r_1, \ldots, r_n>$ is called its *maximum subargument* if the literal that is the conclusion of this subargument is a part of antecedent of rule r_n. It is clear that an argument Arg = $<r_1, \ldots, r_n>$ may have more than one maximum subargument, it depends on the number of literals in the antecedent of rule r_n. Limiting argument is a maximum subargument of argument Arg = $<r_1,\ldots,r_n>$ if the corresponding literal is a part of the antecedent of rule r_n.

Example 2 (continued) Arguments of stepping theory $T_2 = (R_2, \emptyset)$ are Arg_1 = <N1, N2, N3, N4> with all their subarguments. The argument Arg_2 = <N1, N2, N3> are the maximum (supporting) subargument of argument Arg_1. Limiting argument of the theory T_2 (and the corresponding subargument of arguments Arg_1

and Arg_2) are Arg_3 = later (3). Rule N5 is not an argument of stepping theory T2 because it does not satisfy the Definition 1.

Turning to the definition of the status of the argument and the conflict between the arguments of the stepping theory, it should be noted that in contrast to other systems of reasoning, where we study the relationship between various arguments is "static", there, informally, it is necessary to consider the development of these relationships over time. At some stage of conclusion particular supporting argument may not yet be built, i.e., it does not have time to get into action, and after getting into action it can act until (conclusion step) out of operation. This means that the argument (in this step) rejected (=refuted) because of the aftermath of the conflict with the other arguments. Thus, the term "administering an argument in action" plays a key role in determining the status of arguments and conflict between them. For simplicity, output step number i will be referred to simply as a step i.

Definition 2 The introduction of the arguments of stepping theories made by the following rules.

1. Any type of limiting argument type later (i) being launched at step i + 1.
2. If the supporting argument has no subarguments and does not attack any other argument, it is put into action in step 1. Otherwise, it is put into action in step, from which it is not attacked by any other argument.
3. Each supporting argument is put into action in step i, if all its subarguments are put into action in the previous steps, and there is his maximum subargument being put into action in step (i − 1) and this supporting argument is not attacked by any other argument.

After putting the argument in the action on the next steps, unless otherwise stated, it is believed that he has the status of *active argument*. Otherwise, the argument has the status of *inactive argument*.

Definition 3 The supporting argument Arg_1 attacks supported argument Arg_2 in step i, ⇔ (1) propositional literals that are their conclusions form a pair of contraries, (2) all maximal subarguments of both the arguments have the status of active.

Note that in this case, the supporting arguments Arg_1 and Arg_2 attack each other. It should be kept in mind that both arguments have the status of inactive, unlike all their subarguments. If at least one of these subarguments lose the status of active, mutual attacks stop.

Definition 4 The supporting argument *loses the status of active* and *becomes inactive* in the step i ⇔ in this step (1) it is attacked by any other argument or (2) one of its subarguments loses the status of active.

Thus, supporting arguments in the course of their "life cycle" may change their status few times. Consider the operator ϑ^{A}_{T} that is in charge of the current argumentation semantics and otherwise similar to the above operator ϑ^{E}_{T}. The following Definition gives a theoretical-argumentation characterization of stepping theories of Active Logic.

Definition 5 Let B be the set of beliefs, such that now (t) ∈ B. Then (1) now (t + 1) ∈ ϑ^{A}_{T} (B) and (2) the necessary and sufficient condition for a literal q to be the part of the set of beliefs ϑ^{A}_{T} (B) is that there is at least one valid supporting argument, the conclusion of which is q and that was put into action no later than at step t, or there is at least one supporting argument with the conclusion q, derived from the operation in step t + 1.

Example 3 Let the set R_3 of stepping theory T_3 = (R3, ∅) consists of the following elements:

N1: ⇒ p	N5: s ⇒ ¬ q
N2: p ⇒ r	N6: later (3) ⇒ ¬ s
N3: r ⇒ q	N7: q ⇒ o
N4: ⇒ s	

Supporting arguments of stepping theory T_3 are Arg_1 = <N1, N2, N3, N7>, Arg_2 = <N4, N5>, Arg_3 = N6, and all of their subarguments. Limiting argument of this theory is Arg_4 = later (4). The history of the stepping theory is as follows: $\boldsymbol{B}_3$ = < $\boldsymbol{B}_3$(0) = {now (0)}, $\boldsymbol{B}_3$ (1) = {now (1), p, s}, $\boldsymbol{B}_3$ (2) = {now (2), p, s, r, ¬ q}, $\boldsymbol{B}_3$ (3) = {now (3), p, s, r}, $\boldsymbol{B}_3$ (4) = {now (4), p, r, q}, $\boldsymbol{B}_3$ (5) = {now (5), p, r, q, o}>.

Supporting argument Arg_2 and maximum subargument of supporting argument Arg_1 attacked each other in step 3, resulting in a literal ¬ q, who was in the set of beliefs $\boldsymbol{B}_2$ (2), did not get into the set of beliefs $\boldsymbol{B}_2$ (3), because supporting argument Arg_2, the conclusion of which was a literal ¬ q. In step 4 came into effect the limiting argument Arg_4, which is the subargument of argument Arg_3. The latter immediately attacked the subargument of argument Arg_2, made it inactive, and tore the attack of argument Arg_2 on the maximum subargument of argument Arg_1. The latter came into action in step 4.

Theorem 1 Let B = $\boldsymbol{B}$(i)—the set of beliefs of stepping theory T = (R, >). Then ϑ^{A}_{T}(B) ⊆ ϑ^{E}_{T} (B), moreover, if the appropriate stepping theory T is not inconsistent, the equality ϑ^{A}_{T}(B) = ϑ^{E}_{T} (B) be.

Proof If the stepping theory T is not contradictory, then by induction on number of step conclusion i it can be proved for any literal q, that q ∈ B in accordance with the E-semantics ⇔ according to the argumentation semantics there exists the supporting argument with the conclusion q, acting on the step i and came into action no later than the number of step conclusion (i − 1). If the stepping theory T is inconsistent, then by induction on number of step conclusion i it can be shown that if q ∈ B according to argumentation semantics, then q ∈ B in accordance with the E-semantics. □

Theorem 2 Let stepping theory T = (R, >) is inconsistent in accordance with the of E-semantics, $\boldsymbol{B}$ = <$\boldsymbol{B}$ (0), … , $\boldsymbol{B}$(n)>—its history, such that for all i, $0 \leq i \leq n$, $\boldsymbol{B}(i + 1) = \vartheta^{A}{}_{T} (\boldsymbol{B} (i))$. Then for all i, $0 \leq i \leq n$, there is at least one literal h ∈ Lit_T, such that $h \notin \boldsymbol{B}(i)$.

Proof Since T is inconsistent, in R there exists at least two rules are such that literals that are their conclusions, constitute a pair of contraries, and this pair of literals on one of the steps of the conclusion corresponding to the E-semantics is the elements of the corresponding set of beliefs. This means that according to the semantics of argumentation, arguments, conclusions of which are the literals in this step of conclusion attack each other, and none of the literals can not appear in the corresponding set of beliefs. These literals may appear in sets of beliefs apart when all the arguments, the conclusion of which is one of those literal, are inactive, but not both literals at the same time. □

Theorem 2, in fact, says that stepping theories corresponding argumentation semantics, although they may be contradictory, but not trivial in the sense defined in Sect. 1. Thus, argumentation semantics is Paraconsistent.

5 Conclusions

The article shows that the argumentation semantics presented for stepping theories of Active Logic in [5] is paraconsistent. Stepping theories may contain a contradiction (in terms of argumentation semantics contradictory is any Stepping theory in which there are arguments, attacking each other in certain moments of time), but it does not lead to their destruction, as in the case of E-semantics. Note that in terms of argumentation semantics and E-semantics corresponding sets of conflicting stepping theories stepper with each other.

References

1. Elgot-Drapkin, J., Perlis, D.: Reasoning situated in time I: basic concepts. J. Exp. Theor. Artif. Intell. **2**(1), 75–98 (1990)
2. Hovold, J.: On a semantics for active logic, MA Thesis, Department of Computer Science, Lund University (2011)
3. Vinkov, M.M.: Active logic in the context of well-founded semantics of logical programs with priorities. In: *Sb. nauch. tr. IX nats. konf. s mezhdunar. uchastiem "KII—2004"* [Proceedings of the 9th National Conference with International Participation "KII—2004"], vol. 1, pp. 86–94. Fizmatlit Publisher, Moscow (2004) (in Russian)
4. Vinkov, M.M.: Argumentation semantics for active logic step theories. In: *Sb. nauch. tr. X nats. konf. s mezhdunar. uchastiem "KII—2006"* [Proceedings of the 10th National Conference with International Participation "KII—2006"], vol. 1, pp. 64–72. Obninsk (2006) (in Russian)
5. Vinkov, M.M., Fominykh, I.B.: Argumentation semantics for active logic step theories with granulation time. In: *Iskusstvenny intellekt i prinyatie resheniy* [Artificial intelligence and decision-making], no. 3, pp. 3–9. URSS Publishers, Moscow (2015) (in Russian)
6. Dung, P.M.: Negation as hypothesis: an abduction foundation for logic programming. In: Proceedings of the 8th International on Logic Programming. MIT Press, Paris (1991)
7. Vagin, V.N., Zagoryanskaya, A.A.: Argumentatsiya v pravdopodobnom vyvode. In: *Sb. nauch. tr. VII nats. konf. s mezhdunar. uchastiem "KII—2000"* [Proceedings of the 9th National Conference with International Participation "KII—2000"], vol. 1, pp. 28–34. Fizmatlit Publisher, Moscow (2000) (in Russian)
8. Jaskowski, S.: Prepositional calculus for contradictory deductive systems. Stud. Logica **24** (1969)
9. da Costa N.C.A.: On the theory of inconsistent formal system. Notre Dame J. Formal Logic **15** (1974)
10. Gelfond, M., Lifsitz, V.: Classical negation in logic programs and disjunctive databases. New Gener. Comput. **9**(3,4), 365–385 (1991)

Query Answering Over Ontologies in the Extended Allen's Interval Logic

Gerald S. Plesniewicz

Abstract We introduce the extended Allen's interval logic whose sentences are Boolean combinations of sentences of Allen's interval logic with metrical information, and define, for this logic, the deduction method based on analytical tableaux. We applied the method to query answering over interval ontologies specified in the extended Allen's logic.

Keywords Knowledge representation · Ontologies · Temporal logics · Allen's interval logic · Query answering over ontologies

1 Introduction. Main Definitions

In 1983, J.A. Allen has published the seminal paper "Maintaining knowledge about temporal intervals", where he has proposed a simple temporal logic formalism [1]. Allen studied qualitative constraints with temporal intervals linked by means of elementary relations (like "before", "after", "during", "overlaps" and so on). His work was followed by a study of metric constraints for temporal intervals. Allen's interval logic and its extensions were applied to various problems of intelligent information systems designing (knowledge representation, common sense reasoning, natural language understanding, actions planning, ontology modeling et al.).

An ontology for a dynamic problem domain contains temporal dependencies between concepts [3, 5, 8]. For example, suppose that a given ontology contains the concept *Agent* with the attribute *Action*. Let a be one of the values of the attribute. Then the expression $Agent.Action = a$ denotes the event "the agent carries out the action a". We assume that the event takes place within a certain temporal interval $A = [A^-, A^+]$, where A^- and A^+ are time points which mark the beginning and end

G.S. Plesniewicz (✉)
National Research University MPEI, Moscow, Russia
e-mail: salve777@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 450, DOI 10.1007/978-3-319-33609-1_16

of the interval. We also assume that A^{-}, A^{+} are integers and $A^{-} < A^{+}$ (so, we consider only non-degenerated intervals).

Thus, with every ontology for dynamic problem domain it is associated the *interval ontology* whose concepts are temporal intervals and temporal relations.

In Allen's interval logic (let us denote it by **LA**), there are 7 basic relations between intervals: ***e*** (equals), ***b*** (before), ***m*** (meets), ***o*** (overlaps), ***f*** (finishes), ***s*** (starts), ***d*** (during) (see Table 1.) The inverted relations are marked by asterisks: ***b**** (after), ***m**** (met-by), ***o**** (overlapped-by), ***f**** (finished-by), ***s**** (started-by), ***d**** (contains). (So, $A\,\alpha^{*}B \Leftrightarrow B\,\alpha\,A$, where α is basic relation and *A*, *B* are temporal intervals.)

A *sentence* (*formula*) of **LA** is an expression of the form *A* ω *B* where ω is any subset of the set $\Omega = \{\boldsymbol{e}, \boldsymbol{b}, \boldsymbol{m}, \boldsymbol{o}, \boldsymbol{f}, \boldsymbol{s}, \boldsymbol{d}, \boldsymbol{b}^{*}, \boldsymbol{m}^{*}, \boldsymbol{o}^{*}, \boldsymbol{f}^{*}, \boldsymbol{s}^{*}, \boldsymbol{d}^{*}\}$ and *A, B* are interval names. If ω = {α}, then we have a *primitive* sentence, and instead of $A\{\alpha\}B$ we write *A* α *B*. If $\omega = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ then we write $A\ \alpha_1\alpha_2\ldots\alpha_k\ B$ instead of $A\{\alpha_1, \alpha_2, \ldots, \alpha_k\}B$. The formula *A* ω *B* is interpreted as disjunction of the primitive formulas: $A\,\omega\,B \Leftrightarrow \mathrm{V}\{A\,\alpha\,B | \alpha \in \omega\}$.

As every logic, **LA** induces the relation "|=" of *logical consequence*. Let ***O*** be an interval ontology written in **LA** and σ be a **LA** sentence. Then ***O*** | = σ if and only if there is no interpretation such that all sentences of ***O*** are true but σ is false.

Boolean extension **LA+** of Allen's interval logic **LA** consists of formulas which are Boolean combinations of **LA** formulas, i.e. **LA+** has the following syntax: (a) propositional variables are **LA+** sentences; (b) **LA** sentences belong to **LA+**; (c) $\sim \varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$ are **LA+** sentences if φ and ψ are **LA+** sentences. We also introduce $\varphi \rightarrow \psi$ as an abbreviation of $\sim \varphi \vee \psi$. Semantics of **LA+** formulas is defined as usual.

Take, for example, the primitive **LA** sentence *A* ***o*** *B*. The sentence is characterized by the inequalities $A^{-} < B^{-}$, $B^{-} < A^{+}$, $A^{+} < B^{+}$ or by the inequalities $B^{-} - A^{-} > 0$, $A^{+} - B^{-} > 0$, $B^{+} - A^{+} > 0$. Suppose, $B^{-} - A^{-} = 3$, $A^{+} - B^{-} \geq 1$. Then we have metric information which constrains interpretation of the sentence *A* ***o*** *B*, and we write the sentence $A\ \boldsymbol{o}(B^{-} - A^{-} = 3; A^{+} - B^{-} \geq 1)\ B$. The sentences of this type make up the language **μLA** of the Boolean extension interval Allen's logic with metrical information. The language **μLA** has the following syntax:

Table 1 Basic relations of Allen's interval logic

Interval relation	Illustration	Inequalities and equalities for intervals ends
A ***b*** *B*	|===*A*===||===*B* ===|	$A^{+} < B^{-}$
A ***m*** *B*	|===*A*===||=====*B*=====|	$A^{+} = B^{-}$
A ***o*** *B*	|===*A*===||=====*B*=====|	$A^{-} < B^{-}$, $B^{-} < A^{+}$, $A^{+} < B^{+}$
A ***d*** *B*	|===*A*===||=====*B*=====|	$B^{-} < A^{-}$, $A^{+} < B^{+}$
A ***s*** *B*	|===*A*===||=====*B*=====|	$A^{-} = B^{-}$, $A^{+} < B^{+}$
A ***f*** *B*	|===*A*===||=====*B*=====|	$B^{-} < A^{-}$, $A^{+} = B^{+}$
A ***e*** *B*	|=====*A*=====||=====*B*=====|	$A^{-} = B^{-}$, $A^{+} = B^{+}$

(a) *primitive constraints* in the ontology $\boldsymbol{O}$ with intervals $A_1, A_2, \ldots, A_n$ have the forms $X - Y = r, X - Y > r, X - Y \geq r, X - Y < r, X - Y \leq r, X - Y < s$, where $X, Y \in \{A_1^-, A_1^+, A_2^-, A_2^+, \ldots, A_n^-, A_n^+\}$ and r is an integer;

(b) an arbitrary *constraint* λ is a conjunction of primitive constraints, i.e. λ has the form $\theta_1; \theta_2; \ldots; \theta_m$ where θ_i are primitive constraints, and semicolon denotes conjunction;

(c) a **μLA** *sentence* has the form $A\ \delta_1\delta_2\ldots\delta_m\ B$, where every $\delta_i \in \Omega$ or $\delta_i = \alpha(\lambda)$, $\alpha \in \Omega$ and λ is a constraint such that only members of $\{A^-, A^+, B^-, B^+\}$ can occur in λ.

We extend the language **μLA** to the language of the *extended Allen's interval logic* **μLA+** in the same way as we have extended **LA** to **LA+**. The language **μLA+** has the following syntax: (a) propositional variables are **μLA+** sentences; (b) constraints are **μLA+** sentences; (c) **μLA** sentences are **μLA+** sentences; (d) $\sim \varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$ are **μLA+** sentences if φ and ψ are **μLA+** sentences.

Example 1 Suppose, there is an agent which can carry out the actions a, b and c. Each action requires some time; therefore, temporal intervals A, B, C are associated with the actions a, b, c. Moreover, suppose that there are conditions p and q such that:

(1) If p is true then there is no time point at which both actions a and b are carried out;

(2) If q is true then the action b is carried out only when the action c is carried out. Consider the question:

(3) What Allen's relations are impossible between the A and C if both conditions p and q take place?

It is clear that the assertions (1) and (2) can be represented in **LA+** by the formulas $p \rightarrow A\ \boldsymbol{bb^*}B$ and $q \rightarrow B\ \boldsymbol{edfs}\ C$. Also, we may write the question (3) as the query ? $x - p \wedge q \rightarrow \sim A\ x\ C$ to the ontology $\boldsymbol{O} = \{p \rightarrow A\ \boldsymbol{bb^*}B,\ q \rightarrow B\ \boldsymbol{edfs}\ C\}$. The answer to the query consists of those relations $x \in \Omega$ that the logical consequence $\boldsymbol{O} \mid = p \wedge q \rightarrow \sim C\ x\ A$ holds. Later, in Example 3, we will find that the answer consists of the relations $\boldsymbol{d^*}$, $\boldsymbol{e}$, $\boldsymbol{f^*}$ and $\boldsymbol{s}$.

Consider the following two assertions about the actions a, b, c, and the query:

(4) If p is true then the action a is carried out before b with time distance ≥ 2;

(5) If p is true then the action b is carried out before c with time distance ≥ 3;

(6) Suppose if $p \wedge q$ holds. Find the greatest x such that distance between intervals A and C is not less than x.

These assertions are written in **μLA+** as $p \rightarrow A\ \boldsymbol{b}(B^- - A^+ \geq 2)\ B$, $q \rightarrow B\ \boldsymbol{b}(C^- - A^+ \geq 3)\ C$.

The query can be written as ? max $x - p \wedge q \rightarrow A\ \boldsymbol{b}(C^- - A^+ \geq x)\ C$. (*End of Example* 1.)

In this paper, we give the deduction method based on analytical tableaux for the extended Allen's interval logics **LA+** and **μLA+**. We show how to apply this method for query answering over ontologies written in **LA+** and **μLA+**.

2 Inference Rules and Deduction in the Logics LA+ and μLA+

In Tables 2, 3 and 4, there are the inference rules for the logics **LA+** and **μLA+**. Consider an example of deduction by means of these rules.

Example 2 Take the ontology $\boldsymbol{O} = \{p \to A\ \boldsymbol{b}(B^- - A^+ \geq 2)\ B,\ q \to B\ \boldsymbol{b}(C^- - A^+ \geq 3)\ C\}$ and the formula $-\,p \wedge q \to A\ \boldsymbol{b}(C^- - A^+ \geq 3)\ C$. To prove that $\boldsymbol{O} \mid = -\,p \wedge q \to A\ \boldsymbol{b}(C^- - A^+ \geq 3)\ C$, we construct the deduction tree by applying the inference rules to the following set of formulas with signs "+" and "–": $\{+\,p \to A\ \boldsymbol{b}(B^- - A^+ \geq 2)\ B,\ +\,q \to B\ \boldsymbol{b}(C^- - A^+ \geq 3)\ C\}$ (see Fig. 1).

Constructing the deduction tree starts with the initial branch containing the formulas $+p \to A\ \boldsymbol{b}(B^- - A^+ \geq 2)\ B$, $+q \to B\ \boldsymbol{b}(C^- - B^+ \geq 3)\ C$, $-\,p \wedge q \to A\ \boldsymbol{b}(C^- - A^+ \geq 3)\ C$. At the first step we have applied the rule from Table 2 in the second row and fourth column (denote it T2.24) to the formula $-\,p \wedge q \to A\ \boldsymbol{b}(C^- - A^+ \geq 3)\ C$ and have put the label "[7]" on the right of the formula. As a result of the application of the rule T2.24, two formulas $+\,p \wedge q$ and $-\,A\ \boldsymbol{b}(C^- - A^+ \geq 3)\ C$ have been added in sequence to the initial branch, and the label "1:" has been put on the left of these formulas. At the step 7, the rule T2.23 has been applied to $+q \to B\ \boldsymbol{b}(C^- - B^+ \geq 3)\ C$. As a result, the "fork" with formulas $-\,q$ and $+B\ \boldsymbol{b}(C^- - B^+ \geq 3)\ C$ has been added to the current branch. The tree has four branches. The first (left) branch is *closed*, i.e. it is inconsistent since it contains the contrary pair $+p$ and $-\,p$. The second branch also closed since it contains the contrary pair $+q$ and $-\,q$. Let us write out of the third and fourth branches all inequalities (without signs), and add the valid inequalities $A^+ - A^- \geq 1$, $B^+ - B^- \geq 1$, $C^+ - C^- \geq 1$:

$$S_1 = \{B^- - A^+ \geq 2,\ C^- - B^+ \geq 3,\ A^+ - C^- \geq 0,\ A^+ - A^- \geq 1,\ B^+ - B^- \geq 1,\ C^+ - C^- \geq 1\},$$
$$S_2 = \{B^- - A^+ \geq 2,\ C^- - B^+ \geq 3,\ A^+ - C^- \geq -3,\ A^+ - A^- \geq 1,\ B^+ - B^- \geq 1,\ C^+ - C^- \geq 1\}.$$

The sets S_1 and S_2 are inconsistent. Indeed, S_1 includes the inequalities $B^- - A^+ \geq 2$, $B^+ - B^- \geq 1$, $C^+ - B^+ \geq 3$, $A^+ - C^- \geq 1$. Adding up these inequalities ($B^- -$

Table 2 Inference rules for propositional connectives

$\dfrac{+\sim\varphi}{-\varphi}$	$\dfrac{-\sim\varphi}{+\varphi}$	$\dfrac{+\varphi\wedge\psi}{\begin{matrix}+\varphi\\+\psi\end{matrix}}$	$\dfrac{-\varphi\wedge\psi}{-\varphi \mid -\psi}$
$\dfrac{+\varphi\vee\psi}{3+\varphi \mid +\psi}$	$\dfrac{-\varphi\vee\psi}{\begin{matrix}-\varphi\\-\psi\end{matrix}}$	$\dfrac{+\varphi\to\psi}{-\varphi \mid +\psi}$	$\dfrac{-\varphi\to\psi}{\begin{matrix}+\varphi\\-\psi\end{matrix}}$

Table 3 Inference rules for the Allen's connectives

$\dfrac{+A\,b\,B}{B^- - A^+ \geq 1}$	$\dfrac{-A\,b\,B}{A^+ - B^- \geq 0}$
$\dfrac{+A\,m\,B}{\begin{array}{c}B^- - A^+ \geq 0\\ A^+ - B^- \geq 0\end{array}}$	$\dfrac{-A\,m\,B}{B^- - A^+ \geq 1 \mid A^- - B^+ \geq 1}$
$\dfrac{+A\,o\,B}{\begin{array}{c}B^- - A^- \geq 1\\ A^+ - B^- \geq 1\\ B^+ - A^+ \geq 1\end{array}}$	$\dfrac{-A\,o\,B}{A^- - B^- \geq 0 \mid B^- - A^+ \geq 0 \mid A^+ - B^+ \geq 0}$
$\dfrac{+A\,f\,B}{\begin{array}{c}A^- - B^- \geq 1\\ B^+ - A^+ \geq 0\\ A^+ - B^+ \geq 0\end{array}}$	$\dfrac{-A\,f\,B}{B^- - A^- \geq 0 \mid B^+ - A^+ \geq 1 \mid A^+ - B^+ \geq 1}$
$\dfrac{+A\,s\,B}{\begin{array}{c}B^- - A^- \geq 0\\ A^- - B^- \geq 0\\ B^+ - A^+ \geq 1\end{array}}$	$\dfrac{-A\,s\,B}{B^- - A^- \geq 1 \mid A^- - B^- \geq 1 \mid A^+ - B^+ \geq 0}$
$\dfrac{+A\,d\,B}{\begin{array}{c}A^- - B^- \geq 1\\ B^+ - A^+ \geq 1\end{array}}$	$\dfrac{-A\,d\,B}{B^- - A^- \geq 0 \mid A^+ - B^+ \geq 0}$
$\dfrac{+A\,e\,B}{\begin{array}{c}B^- - A^- \geq 0\\ A^- - B^- \geq 0\\ B^+ - A^+ \geq 0\\ A^+ - B^+ \geq 0\end{array}}$	$\dfrac{-A\,e\,B}{B^- - A^- \geq 1 \mid A^- - B^- \geq 1 \mid B^+ - A^+ \geq 1 \mid A^+ - B^+ \geq 1}$
$\dfrac{+A\,\alpha(\lambda)\,B}{\begin{array}{c}+A\,\alpha\,B\\ +\lambda\end{array}}$	$\dfrac{-A\,\alpha(\lambda)\,B}{-A\,\alpha\,B \mid -\lambda}$
$\dfrac{-A\,\beta\theta\,B}{+A\,\beta\,B \mid +A\,\theta\,B}$	$\dfrac{+A\,e\,B}{\begin{array}{c}-A\,\beta\,B\\ -A\,\theta\,B\end{array}}$
λ is a constraint, $\alpha \in \Omega$, $\beta \in \Omega$ or $\beta = \alpha(\lambda)$, θ is a sequence β.	

$A^+) + (B^+ - B^-) + (C^+ - B^+) + (A^+ - C^-)$, we obtain $0 \geq 2 + 1 + 3 + 0 = 6$ (contradiction). Similarly, S_2 is inconsistent, since it includes the inequalities $B^- - A^+ \geq 2$, $B^+ - B^- \geq 1$, $C^+ - B^+ \geq 3$, $A^+ - C^- \geq -3$, and adding up them we obtain $0 \geq 2 + 1 + 3 + (-3) = 3$ (contradiction). Hence, the deduction tree is closed, and this means inconsistency of the ontology ***O***. (*End of Example* 2.)

Let S be any set of inequalities of the forms $X - Y \geq r$, where X, Y are integer variables and r is an integer. We associate with S the following labeled directed graph Γ(S). Its vertices are integer variables and its labeled arcs are triples (X, Y, r) with the condition that the inequality $X - Y \geq r$ enters the set S.

Figure 2 shows the graphs $\Gamma(S_1)$ and $\Gamma(S_2)$ for the above sets S_1 and S_2. (Here we did not draw the labels "1".) We see that the graphs contain the cycle A^+, B^-, B^+, C^+, D^-, A^+. Length of this cycle in $\Gamma(S_1)$ is 6, and in $\Gamma(S_2)$ is 3.

Table 4 Inference rules for constraints

$\frac{+\lvert A\rvert = r}{A^+ - A^- \geq r \quad A^- - A^+ \geq -r}$	$\frac{-\lvert A\rvert = r}{A^- - A^+ \geq 1-r\,\vert \quad A^+ - A^- \geq r+1}$	$\frac{+X-Y=r}{X-Y \geq r \quad Y-X \geq -r}$	$\frac{-X-Y=r}{X-Y \geq r+1\,\vert \quad Y-X \geq r+1}$
$\frac{+\lvert A\rvert \geq r}{A^+ - A^- \geq r}$	$\frac{+\lvert A\rvert > r}{A^+ - A^- \geq r+1}$	$\frac{+\lvert A\rvert \leq r}{A^- - A^+ \geq -r}$	$\frac{+\lvert A\rvert < r}{A^- - A^+ \geq 1-r}$
$\frac{-\lvert A\rvert \geq r}{A^- - A^+ \geq r+1}$	$\frac{-\lvert A\rvert > r}{A^- - A^+ \geq -r}$	$\frac{-\lvert A\rvert \leq r}{A^+ - A^- \geq r+1}$	$\frac{-\lvert A\rvert < r}{A^- - A^+ \geq r}$
$\frac{+X-Y \geq r}{X-Y \geq r}$	$\frac{+X-Y > r}{X-Y \geq r+1}$	$\frac{+X-Y \leq r}{Y-X \geq -r}$	$\frac{+X-Y \leq r}{Y-X \geq -r}$
$\frac{-X-Y \geq r}{Y-X \geq 1-r}$	$\frac{-X-Y \geq r}{Y-X \geq -r}$	$\frac{-X-Y \leq r}{X-Y \geq r+1}$	$\frac{-X-Y < r}{X-Y \geq r}$
$\frac{+X-Y=r}{X-Y \geq r \quad Y-X \geq -r\,\vert}$	$\frac{-X-Y=r}{X-Y \geq 1+r\,\vert \quad Y-X \geq 1-r}$	$\frac{+\theta;\lambda}{+\theta \quad +\lambda}$	$\frac{-\theta;\lambda}{-\theta\,\vert -\lambda}$
$\frac{+X-Y=r}{X-Y \geq r \quad Y-X \geq -r\,\vert}$	$\frac{-X-Y=r}{X-Y \geq 1+r\,\vert \quad Y-X \geq 1-r}$	$\frac{+\theta;\lambda}{+\theta \quad +\lambda}$	$\frac{-\theta;\lambda}{-\theta\,\vert -\lambda}$

θ is a primitive constraint, λ is an arbitrary constraint

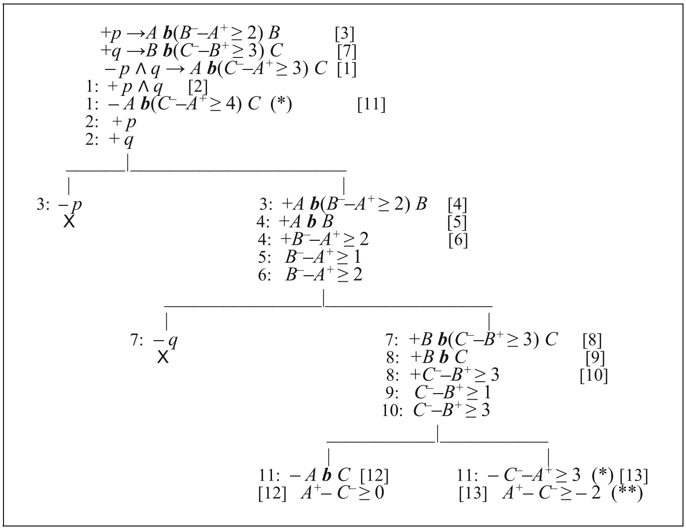


Fig. 1 Deduction tree for Example 2

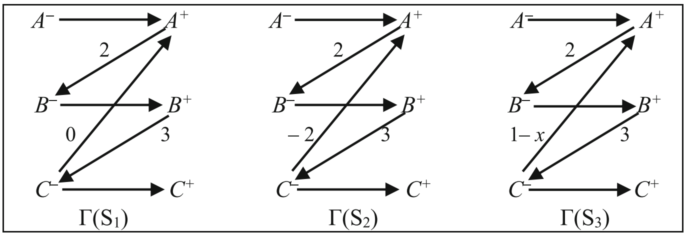


Fig. 2 Graphs for systems of inequalities

In general, for arbitrary set S of inequalities the following assertion is true: S *is inconsistent if and only if* Γ(S) *has a positive cycle*. There is a computationally effective algorithm for detecting positive cycles in labeled graphs. This algorithm can be used to prove the closing of deduction trees.

3 Query Answering Over Ontologies in LA+ and μLA+

Let ***O*** be an ontology in **LA+**. A *query* to ***O*** has the form

$?(x_1, x_2,\ldots, x_n) - \kappa[x_1, x_2,\ldots, x_n]$,

where $\kappa[x_1, x_2,\ldots, x_n]$ is a **LA+** formula that contains primitive sentences of the form $A\ x_i\ B$. The answer to this query is the tuple $(\alpha_1, \alpha_2,\ldots, \alpha_n)$ (where $\alpha_n \in \Omega$) such that the formula, which is the result of replacing x_i with α_i, is logical consequence of the ontology: ***O*** $\models \kappa[\alpha_1, \alpha_2,\ldots, \alpha_n]$.

Example 3 Take the ontology ***O*** = {*p* → *A* ***bb****B*, *q* → *B* ***edfs*** *C*} and the query ?*x* – *p* ∧ *q* → ~ *A x C*. In Fig. 3, the deduction tree for the set {+ *p* → *A* ***bb****B*, +*q* → *B* ***edfs*** *C*} of formulas (with signs "+" and "–") is shown. Here at step 6 we have applied the second inference rule from Table 3 to formulas *A* ***bb**** *B* and *B* ***edfs*** *C*. As result, we have get +*A* ***bb*dmm*o*s*** *B* since (see Table 5)

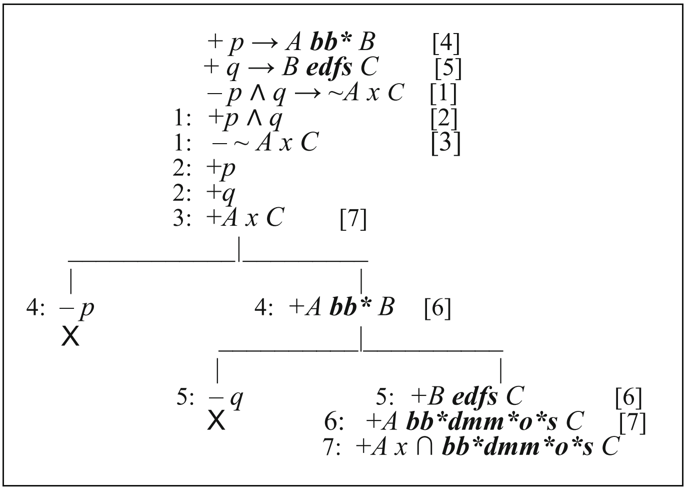


Fig. 3 Deduction tree for Example 3

Table 5 Inference rules for logic **LA**

$\frac{+A\,\alpha\,B}{+A\,\alpha^* C}$	$\frac{+A\,\alpha\,B \quad +B\,\beta\,C}{+A\,\alpha\circ\beta\,C}$	$\frac{+A\,\alpha\,B \quad +B\,\beta\,C}{+A\,\alpha\cap\beta\,C}$	$\frac{-A\,\omega\,B}{A(\Omega\backslash\alpha)C}$

Table 6 Fragment of Allen's multiplication table

	d	***f***	***s***
b	***bdmos***	***bdmos***	***b***
b*	***b*dfmo****	***b****	***b*dfmo****

$$\begin{aligned} bb^*\circ edfs &= b\circ e \cup b\circ d \cup b\circ f \cup b\circ s \cup b^*\circ e \cup b^*\circ d \cup b^*\circ f \cup b^*\circ s \\ &= b \cup bdmos \cup bdmos \cup b \cup b^* \cup b^*dfm^*o^* \cup b^* \cup b^*dfm^*o^* \\ &= bb^*dfmm^*oo^*s. \end{aligned}$$

The third branch will be closed if and if the $x \cap$ ***bb*dmm*o*s*** $= \varnothing$. Hence, the logical consequence $\boldsymbol{O} \models p \wedge q \rightarrow \sim A\ x\ C$ takes place if and only if $x \subseteq \Omega \setminus$ ***bb*dmm*o*s*** $=$ ***d*ef*s***. Thus, the answer to query ? $x - p \wedge q \rightarrow \sim A\ x\ C$ consists of the relations ***d****, ***e***, ***f**** and ***s***. (*End of Example* 3.)

For logic **μLA+** we consider queries of the form

? $\max\{x_1, x_2,\ldots, x_m\}, \min\{y_1, y_2,\ldots, y_n\} - \kappa[x_1, x_2,\ldots, x_m, y_1, y_2,\ldots, y_n\}$,

where $\kappa[x_1, x_2,\ldots, x_m, y_1, y_2,\ldots, y_n\}$is **μLA+** formula that contains inequalities with integer variables x_i and y_j. The answer to that query is the maximal values of x_i and minimal values of y_i such that $\kappa[x_1, x_2,\ldots, x_m, y_1, y_2,\ldots, y_n\}$ is logical consequence of the given ontology (Table 6).

Consider, by example, how to find answers to these kind of queries.

Example 4 Let us take the ontology $\boldsymbol{O} = \{p \rightarrow A\ \boldsymbol{b}(B^- - A^+ \geq 2)\ B,\ q \rightarrow B\ \boldsymbol{b}(C^- - A^+ \geq 3)\ C\}$ and the query ? max $x - p \wedge q \rightarrow A\ \boldsymbol{b}(C^- - A^+ \geq x)\ C$. To prove that $\boldsymbol{O} \models p \wedge q \rightarrow A\ \boldsymbol{b}(C^- - A^+ \geq x)\ C$, we must construct the deduction tree by applying the inference rules to the following set of formulas with the signs "+" and "–": $\{+\ p \rightarrow A\ \boldsymbol{b}(B^- - A^+ \geq 2)\ B,\ +\ q \rightarrow B\ \boldsymbol{b}(C^- - A^+ \geq x)\ C\}$. This tree is obtained from the tree shown in Fig. 1 by replacing the formulas $-\ A\ \boldsymbol{b}(C^- - A^+ \geq 4)\ C$ (*) and $-\ C^- - A^+ \geq 4$ (*) with $-\ A\ \boldsymbol{b}(C^- - A^+ \geq x)\ C$ and $-\ C^- - A^+ \geq x)$ (correspondingly), and replacing the formula $A^+ - C^- \geq -3$ (**) with $A^{+n} - C^- \geq 1 - x$. Figure 2 shows the graph $\Gamma(S_3)$ which is constructed for fourth branches of the deduction three. The graph $\Gamma(S_3)$ contains the cycle A^+, B^-, B^+, C^-, A^+ of the length $2 + 1 + 3 + 1 - x$, and that cycle will be positive if and only if $x \leq 6$. Therefore, $x = 6$ is the answer to that query. (*End of Example* 4.)

4 Conclusion

We have defined the extended Allen's interval logic **μLA+**. The sentences of this logic are Boolean combinations of the sentences of Allen's interval logic with incorporated inequalities and equalities for ends of temporal intervals. For this logic, we have defined the deduction method based on analytical tableaux [2]. This method was used for query answering over ontologies written in **μLA+**. Possible practical applications of the results of the paper are in domain of workflows technology [7, 4, 6, 8].

Acknowledgment This work was supported by Russian Foundation for Basic Research (project 14-07-0387) and Ministry of Education and Science of Kazakhstan (project0115 RK 00532)

References

1. Allen, J.A.: Maintaining knowledge about temporal intervals. Commun. ACM **20**(11), 832–843 (1983)
2. D'Aggostino, M., Gabbay D., Hahnle R., Possega J. (eds.): Handbook of Tableaux Methods, p. 672. Kluwer Academic Publishers, Dortrecht (1999)
3. Fisher, M., Gabbay, D., Vila, L. (eds.): Handbook of Temporal Reasoning in Artificial Intelligence, vol. 1. Elsevier Science Inc. (2005)
4. Gagne, S., Trudel, D.: A Temporal Semantics for Workflow Control Patterns. In: 2008 32nd Annual IEEE International Computer Software and Application Conference (COMPSAC'08), pp. 999–1004. IEEE Computer Society, Washington, USA (2008)
5. O'Connor, M., Das, A.: A method for Representing and Querying Temporal Information in OWL. In: 3th International Joint Conference BIOSTEC 2010, pp. 97–110, Valencia, Spain, no. 127 in Communications in Computer and Information Science, Springer (2011)
6. Okeo, G., Chen, L., Wang, Y., Sterritt, R.: A Hybrid and Temporal Approach for Composite Activity Modeling. In: 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communication (TrustCom 2012), pp. 1763–1770, Liverpool, UK, June 25–27, 2012, IEEE Computer Society (2012)
7. van der Aalst, W., van Hee, K.: Workflow Management Models, Methods, and Systems, p. 53. The MIT Press, Cambridge, Massachusetts London (2002)
8. Wang, Y., Fan, Y.: Using Temporal Logics for Modeling and Analysis of Workflows. In: IEEE International Conference on E-Commerce Technology for Dynamic E-Business, pp. 163–174. IEEE Computer Society, Beijing (2004)

Logical-Algebraic Methods in Constructing Cognitive Sensors for Railway Infrastructure Intelligent Monitoring System

Maria N. Svyatkina, Valery B. Tarassov and Alexander I. Dolgiy

Abstract A complex problem of monitoring railway infrastructure objects is faced. The need in constructing intelligent monitoring system is shown. Ambient intelligence architecture for intelligent monitoring system is introduced, its components are discussed. An important component of ambient intelligence architecture is a hybrid system of knowledge acquisition/discovery that contains sensor data mining (interpretation), expert knowledge acquisition and ontological engineering facilities. Four principles of cognitive measurements are formulated in the context of intelligent monitoring. Cognitive measurement is seen here as a two-leveled hierarchical granulation process realized by using cognitive sensors. Here cognitive sensor is interpreted as not only measurement-information device for studying monitoring object and enabling knowledge discovery from measurement, but also as an "artificial understanding system". The operation of cognitive sensor is based on logical-algebraic pragmatics of measurements. In this context some concepts and definitions of understanding are discussed. A dual descriptive-prescriptive nature of understanding is shown. A generalized logical-algebraic approach to constructing cognitive sensors is proposed. It is based on the concept of logical world and visual representation of various algebraic structures specifying logical worlds by Hasse diagrams. Colored Hasse diagrams are introduced in the paper enabling a better understanding of logical world pragmatics for measurement results. Three basic

This work was supported by RFBR, grants №14-07-00846, №14-07-00653.

M.N. Svyatkina (✉) · V.B. Tarassov
Bauman Moscow State Technical University, Moscow, Russia
e-mail: maria.svyatkina@gmail.com

V.B. Tarassov
e-mail: vbulbov@yahoo.com

A.I. Dolgiy
Rostov State Transport University, Rostov on Don, Russia
e-mail: adolgy@list.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 450, DOI 10.1007/978-3-319-33609-1_17

logical worlds for cognitive sensors are constructed: Vasiliev's, Kleene's and Dunn-Belnap's worlds. These worlds underlie both the cognitive sensors of the same names and cognitive networks based on bilattices and multilattices. As a result, an algorithm of logical-algebraic synthesis for cognitive sensors is developed; the appropriate flowchart is constructed.

Keywords Artificial intelligence • Intelligent monitoring system • Ambient intelligence • Knowledge acquisition/discovery • Cognitive measurement • Information granulation • Cognitive sensor • Pragmatics • Multi-valued logics • Logical world • Bilattice

1 Introduction

The paper is devoted to developing logical-algebraic approach for constructing new generation sensors called cognitive sensors and cognitive sensor networks. Primarily, railway infrastructure monitoring problem is considered, a novel ambient intelligence architecture for intelligent monitoring system is suggested. The concept of measurement as a cognitive process is clarified, its links with evaluation and reasoning processes are shown. Basic principles of cognitive measurements are discussed, where the principle of pragmatic information granulation is of special concern. A natural instrument for performing cognitive measurement are cognitive sensors; they are viewed as both measurement and understanding devices.

We develop a new logical-algebraic approach to constructing cognitive sensors on the basis of logical worlds and visual representation of appropriate algebraic structures (in particular, semi-lattices, lattices, product lattices, bilattices) by Hasse diagrams. By logical world we mean a set of logical (truth or modal) values V together with a subset of designated values $D \subset V$, equipped with different order relations on V.

A natural interpretation of logical values for cognitive sensors is based on "traffic lights" pragmatics and its extensions. A new interpretation of truth values (measured truth, measured falsity, measured contradiction, uncertainty) is proposed, three typical examples of cognitive sensors using three-valued and four-valued logical pragmatics (Vasiliev's sensor, Kleene's sensor, Belnap's sensor) are discussed.

Moreover, the generalization of the proposed approach for developing cognitive sensor networks is suggested. Here the first step is the interpretation of two-sensor data with taking a bilattice-based pragmatics. Finally, an algorithm for logical-algebraic synthesis of cognitive sensors with an appropriate flowchart is presented.

2 Monitoring Railway Infrastructure

The development of high-speed railway transportation systems requires new strategies and tools of monitoring complex artificial structures. Typical examples of such structures are bridges, viaducts, tunnels, dams, etc. In most cases, these structures are rather old and need careful inspection and maintenance.

Here monitoring in a narrow sense means observation, checking and analysis of the railway infrastructure objects to prevent or avoid accidental situations. Nowadays, the problem of monitoring becomes more and more wider and includes such tasks as: (1) quantitative measurement of basic parameters of inspected infrastructure objects and their environment; (2) qualitative interpretation; (3) diagnostics of their current states; (4) prognosis of further evolution; (5) decision-making concerning control actions. Here various measurements are of primary concern. For instance, to reveal a current state of bridge in case of tornado we need to measure both meteorological parameters (primarily, wind direction and speed) and bridge reaction parameters (displacement, load and oscillation, etc.) [1].

Actually, three basic approaches to monitoring railway infrastructure are often used: (a) systematic visual inspection; (b) space satellite-based monitoring; (c) application of SCADA-systems, computer vision and sensor networks [1, 2]. In this paper we develop the concept of Intelligent Monitoring System on the basis of Ambient Intelligence and Smart Environments. This approach can improve the safety of infrastructure objects and provide some urgent assistance in potentially harmful situations.

3 Ambient Intelligence Architecture for Intelligent Monitoring Systems

In 1998, Philips coined the term “Ambient Intelligence” (AmI) in order to illustrate a vision of the future where various information technologies seamlessly interact and adapt to human needs while being none obtrusive. Ambient Intelligence systems aim at augmenting real world environments to create Smart Spaces where users are provided with pervasive virtual services.

According to classical definition, an AmI system is a digital environment that proactively, but sensibly, supports people in their daily lives [3, 4]. Here Ambient Intelligence is often used in a wider sense as a generic concept, and the close term “Smart Environment” is reserved to describe the physical infrastructure (e.g. sensors, actuators and networks) that supports the AmI-system [4, 5].

It is worth noticing that AmI represents a first step towards synergistic computer science [6, 7], where heterogeneous sciences and technologies co-operate to enable a new seamless user-friendly artificial environment of smart networked devices. So AmI scientific problems are inter-disciplinary and are placed on the crossroad of

Computer Science, Cybernetics, Ergonomics, Artificial Intelligence, Cognitive Science, Behavior Science, Mechatronics and Robotics.

Primarily, a cybernetic problem of organizing the required behavior of artificial micro-environment is solved by using both negative feedback and positive feedback; here such problems as user recognition, needs awareness, behavior context understanding, as well as the comparison of real control results with expected ones, are of special concern [8]. Also the black-box concept is augmented by Weiser's ideas [9]: "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it". Thus, the miniaturization of embedded sensors and actuators brings about their dissolution in real environment (or disappearance), so the people perceive only the friendly user interface.

Besides, we deal with a new generation of "man-machine-environment" systems usually studied in ergonomics. Indeed, the AmI represents a new concept of man-machine interface, where people are surrounded by useful micro-devices embedded into physical world and tied to intelligent software. Thus, a hybrid intelligent physical-technical environment emerges on the basis of invisible collective (web) intelligence. The AmI technological devices ought to be embedded, ubiquitous, context aware, personalized, transparent, anticipatory, well-adapted to human senses.

Generally, technological resources of AmI are related to the following six areas: (1) sensors and pervasive measurement devices; (2) actuators, mechatronic devices and robotic systems; (3) ubiquitous networks; (4) ubiquitous computing; (5) user friendly (anthropocentric) interface; (6) Artificial Intelligence (AI) and multi-agent systems.

Below we represent the AmI-system in the following way: AmI = complex distributed sensing and measuring system + complex distributed actuator system + operation system + intelligent kernel (ontologies, multivalued and fuzzy logics, data mining and knowledge discovery, machine learning, knowledge bases, hybrid expert systems, dialogue control,…). Its architecture is given in Fig. 1.

From Fig. 1 it is clear that the perspectives of AmI systems for monitoring are tightly connected with the development of intelligent measurements [10, 11], cognitive sensors, intelligent actuators and mechatronic devices, intelligent dialogue interfaces. Below we shall focus on cognitive measurements by using cognitive sensors [12, 13].

Let us stress that in our approach Cognitive Sensors are a basic component of sensor data mining in hybrid knowledge acquisition/discovery system [12]. Here knowledge acquisition/discovery sub-system for implementing the AmI architecture of intelligent monitoring can be viewed as an integrated 3rd generation system that encompasses expert knowledge, ontological modeling and cognitive measurements.

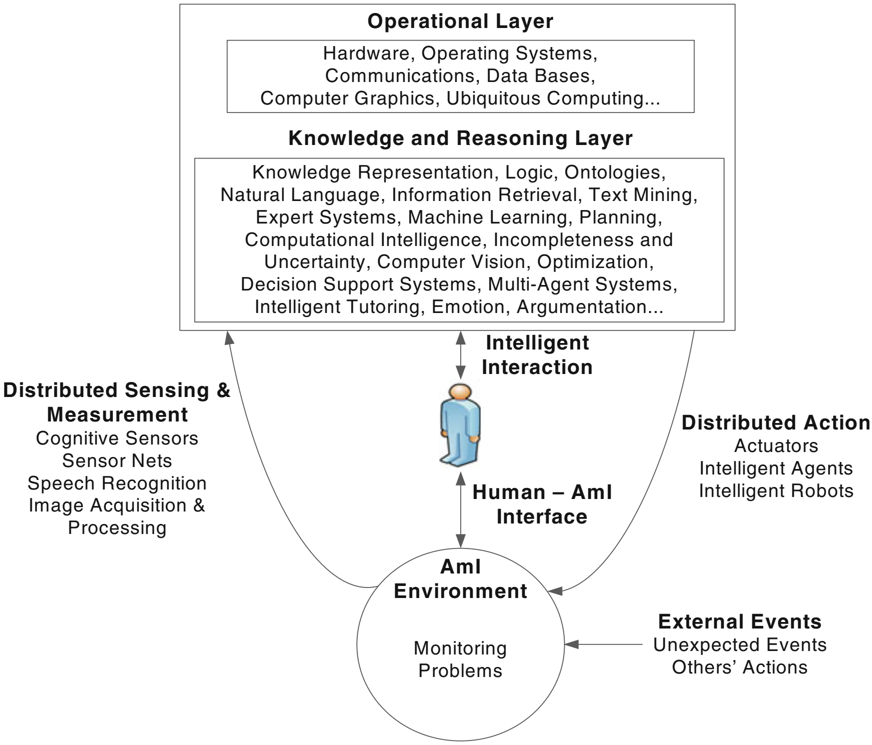


Fig. 1 Architecture of AmI system for monitoring (a revised version of Augusto's figure in [4])

4 Problem Formulation: How to Perform Cognitive Measurement by Cognitive Sensors?

4.1 Cognitive Measurements and Pragmatic Information Granulation

The consideration of measurement as a cognitive process has a long tradition in Russian metrology. One of the deep definitions for measurement was given by prominent Russian philosopher P. A. Florensky in "Technical Encyclopedia": "Measurement is a basic cognitive process in science and technology, in which an unknown quantity intended to be measured (measurand) is compared with other homogeneous and known quantity". This tradition was continued by Knorring [14], Rozenberg [15], Solopchenko, etc. In [14] measurement is conceived as a fundamental technique of cognition (gnoseotechnique) resulting in quantitative description of objects under investigation, and in [15] measurement is viewed as an experimental technique aimed at forming truth-valued judgments about investigated object. Therefore, it is tightly connected with such cognitive processes as computations and reasoning.

The term “Cognitive Measurement” itself has been introduced by Prokopchina [16]. She focuses on metrological knowledge elicitation by generating measurement data networks and transforming them into knowledge networks with employing Bayesian intelligent technologies.

We have proposed in [12, 13, 17] a more general approach based on cognitive prerequisites of measurement and four main principles: (1) principle of open measurement as a cognitive process; (2) principle of measurement-evaluation-reasoning unity; (3) principle of synthesizing truth theories in interpreting measurement results; (4) pragmatic granulation principle for measurement information.

According to the first principle, quite often measurement is not an isolated process; it is included into the sequence of decision support or monitoring processes. For example, in monitoring systems measurement is open in the sense that it is followed by interpretation, diagnosis, prognosis, decision-making. So an open measurement depends on measurement goals, measurement environment, measurement instruments, etc. Various parameters of infrastructure objects are measured with different instruments and, hence, different precision. In case of open measurement the main objective is not to obtain precise quantitative information for each object’s parameter, but to suggest practical recommendations concerning the state of infrastructure object.

The second principle claims the unity of measurement and evaluation (Fig. 2a) or, more generally, the unity of measurement, evaluation and reasoning (Fig. 2b).

In case of measurement-evaluation unity (Fig. 2a) cognitive measurement supposes the creation of a two-level hierarchy where ordinary measurements with using sensor networks provide fine-grained information on lower level and obtained measurement results are mapped into a pragmatic coarse-grained scale of values on higher level. These values usually have axiologic or deontic nature (measured values of object’s parameter are “in norm”, “almost in norm”, “out of norm” etc., for instance, “*a measured value of wind speed equal to 15 m/s is* “*almost in norm*””).

So a pragmatic evaluation scale contains a set of axiological values (“good”, “very good”, “fair”, “bad”, “rather bad” and so on) or deontic values (“obligatory”, “permissible”, “non-obligatory”, “forbidden”). The main objective of cognitive

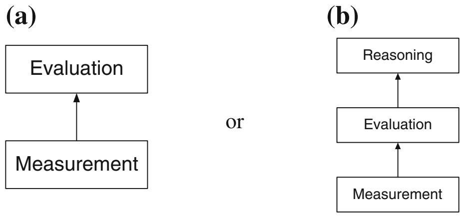


Fig. 2 Two ways of constructing cognitive sensors

measurement in monitoring process is not obtaining perfect quantitative information, but developing useful practical recommendations by further reasoning.

By measurement pragmatics we mean interpretation, understanding and utility of measurement results for solving real world problems. In other words, we associate with any measurement result m from the set of measurement results M a value $v \in V$ interpreted as truth value or, more generally, as a pragmatic value.

Let us give some examples of measurement-based evaluations and reasoning. For instance, the measurement result for wind speed obtained by anemometer is 28–29 m/s. Then cognitive anemometer interprets it as "*a storm*" and outputs the deontic sentence "any traffic on the bridge is forbidden".

Now let us take two monitoring parameters—the wind speed measured by anemometer and bridge load received by tensometer. In case of cognitive measurements we have as a result of interpreting obtained quantitative values some fuzzy linguistic information, for instance, "*The wind speed is rather high*" and "*the load value is close to maximum allowable*". Then we can construct a fuzzy production rule like "**If** *the wind speed is rather high* **AND** *the load value is close to maximum allowable* **Then** *impose restrictions on the mass of trains crossing the bridge about 3500 t*".

The principle of synthesizing truth theories in interpreting measurement results claims that both correspondence theory and coherence theory and pragmatic theory of truth are needed to interpret the measurement results and to understand their meaning with respect to some norm, for example "*the speed of lateral wind equal to 15 m/s is in norm*".

Now we smoothly move to the fourth (in some sense, the most important) principle—measurement information granulation principle. It is based on the concept of granule. L.A. Zadeh specifies *granule* is a collection of objects (values) which are drawn together by indistinguishability, similarity, proximity or functionality [18]. This concept supposes the determination of accuracy and uncertainty in measurement.

Granulation principle claims that the core of cognition is *data granulation*—generation of information granules of required size. Here the term granulation encompasses both composition process (generation of coarse-grained information units) and decomposition process (formation of fine-grained units). The precision of cognitive measurement depends on pragmatic factors. In particular, by selecting different levels of granulation one can obtain different levels of knowledge.

A classical measurement admits (at least as a special case) a singleton value of measurand. Contrarily, contemporary measurement science (see [19]) agrees tacitly that both measured value and its uncertainty are granules. In this context we share Kreinovich-Reznik's viewpoint that any measurement as a cognitive process represents in essence a granulation process [20]. It is the formation of homogeneous and heterogeneous, multi-dimensional granules through one-to-many links between primary measurements and their linguistic interpretations. Therefore, a basic granulation procedure is a pragmatic interpretation of earlier obtained sensor data (Fig. 3).

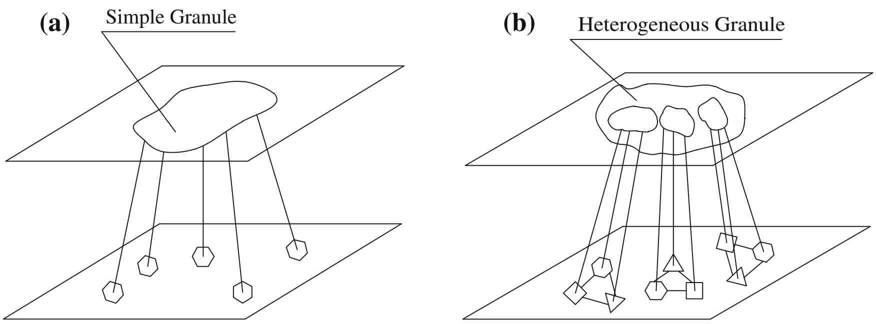


Fig. 3 Granulation in measurement: a two-leveled hierarchy. **a** Formation of simple granules, **b** formation of complex heterogeneous granules

4.2 Cognitive Sensors and Understanding Models

Cognitive measurement supposes the creation of cognitive sensors able both to measure some parameter of considered object and understand the obtained information. Hence cognitive sensor is not only measurement-informational, but also interpreting instrument equipped with logical-linguistic pragmatics, i.e. having the capacity of information granulation by representing measurement results in terms of reduced natural language. Since cognitive sensor is seen here as both an instrument of obtaining knowledge through measurements and an understanding artificial system, let us discuss below some basic concepts and definitions of understanding (see also [21]).

Up to recent time, understanding was reduced to understanding a text provided with some sense. Now it is evident that apart from texts various messages, events, behaviors, measurements, etc. can be also objects of understanding. Understanding may be viewed as operation, capacity, process, result, finally as a tool or technology for knowledge assimilation. It is usually deployed by taking such categories as truth, value, norm, sense, meaning, for example, "find truth", "conceive information value", "make clear the sense" and so on. Besides, it is often considered via dichotomies "cognition—understanding", "learning—understanding", "understanding—action", "understanding—interpretation", etc.

We will be focused on three basic definitions: (1) understanding is a universal cognitive operation based on correspondence between empirical facts and theoretical representation; (2) understanding is a universal cognitive operation enabling assimilation of a new subject, its inclusion into an existing system of ideas and representations; (3) understanding is universal cognitive operation that evaluates some object (for example, text or measurement result) from some pre-defined position on the basis of some standard or norm. Let us stress that these three definitions perfectly correspond to three basic truth theories in logic.

The understanding mechanism has a dual descriptive-prescriptive nature. A pre-understanding of measurement results is based on their logical-linguistic

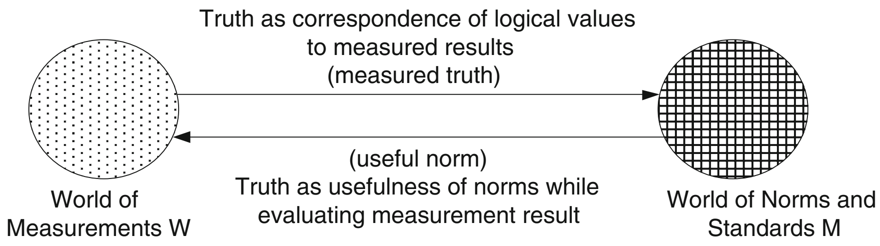


Fig. 4 Illustration of two dual truth concepts for cognitive sensors

description by logical (truth) values. Everything may be understood, for what we have some example, standard or norm, and misunderstanding is generally explained by the lack or implicitness of such an example or norm.

To make the principle of cognitive sensor operation clear let us consider two worlds—a Measurement World W and a Norm World M. It is shown in Fig. 4 that in cognitive sensor based on logical pragmatics two measurement concepts have to be used: (1) Truth as a correspondence of logical values to measured results (measured truth); (2) Truth as a useful norm (and falsity as a useful anti-norm) that allows us take into consideration designated or non-designated logical values.

4.3 *Specification of Logical World as the Core of Cognitive Sensor*

Our logical-algebraic approach to constructing cognitive sensors is based on the concept of Logical World and visual representation of various logical worlds by Hasse diagrams.

Following Karpenko [22] and Shramko [23], we first consider logical world basis as a non-empty set of logical (truth or modal) values. Three basic principles of giving a specific logical world are: (1) *distinguishability principle* (objects in logical world should be different); (2) *designation principle* (some of them can have a particular status, i.e. be designated); (3) *structure generation principle* (objects in logical world should form logical systems expressed by appropriate algebraic structures—chains, lattices, semi-lattices, bilattices, and so on). To differ from [23] we specify Logical World by a set of logical values together with the set of different order relations, for example, truth order, information order, consensus order, etc.

Let us define a Logical World by a triple

$$\boldsymbol{LW} = \langle \boldsymbol{V}, \boldsymbol{R}, \boldsymbol{D} \rangle, \tag{1}$$

where V is a universe of logical values v, $v \in V$, R is the set of order relations defined on V, for instance, truth orders $<_V$, $\leq_V$, information orders $<_I$, $\leq_I$, and D is the set of designated values, $D \subset V$.

Let us stress that a specific logical world is determined by the number of truth values in V, for example, V_3, V_4, $V_{\infty,}$ the number and the type z of designated truth values D, $z \in Z = \{+, -\}$, and the number of dimensions given by order relations from R.

A granular logical world is associated with generalized (granular) truth values. According to Dunn's approach [24], such a generalization supposes the transition from V to 2^V, hence we obtain

$$\boldsymbol{GLW} = \langle \boldsymbol{2}^V, \boldsymbol{R}, \boldsymbol{D} \rangle, \tag{2}$$

where $D \subset 2^{\,V}$. A natural extension of Dunn's approach consists in considering Zadeh's fuzzy logical worlds $[0,1]^V$ with granular and gradual logical values and fuzzy order relations or Goguen's L-fuzzy logical worlds.

Let us consider three basic logical worlds for cognitive sensors: Vasiliev's world, Kleene's world and Dunn-Belnap's world. To specify these worlds we use the following logical values: T—truth, F—falsity, N—uncertainty (it is the abbreviation from "none" = neither truth nor falsity) and B—contradiction (the abbreviation from "both" = "both truth nor falsity"). Here a Hasse diagram for three-valued Vasiliev's lattice Vas3 (paraconsistent Vasiliev's world) is shown in Fig. 5a, and another Hasse diagram for three-valued Kleene' lattice K3 (para-complete Kleene's world) is depicted in Fig. 6a. Two other Figs. 5b and 6b correspond to information orders.

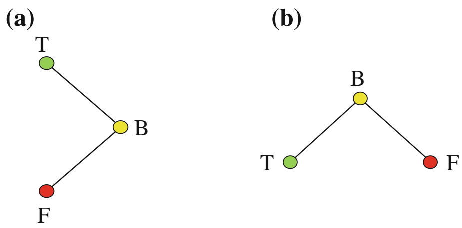


Fig. 5 Structures of paraconsistent logical world: **a** Vasiliev's logical lattice Vas3; **b** information upper semilattice obtained by 90° turning of Vas3

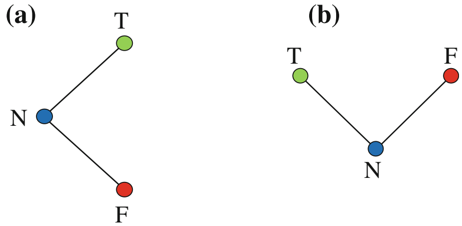


Fig. 6 Structures of paracomplete logical world: **a** Kleene's lattice K3; **b** information lower semilattice

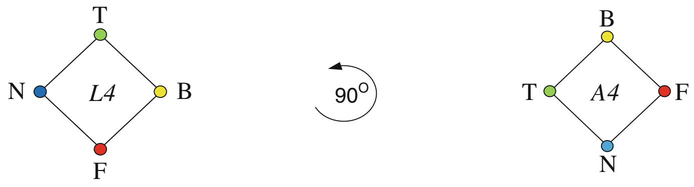


Fig. 7 Hasse diagrams for Belnap's logical lattice L4 and its dual A4

Furthermore, appropriate Hasse diagrams for Belnap's logical lattice L4 and its dual Scott's lattice A4 (information lattice) are shown in Fig. 7a, b.

4.4 Pragmatic Interpretation of Logical Values for Cognitive Sensors

Let is specify some simple cases of cognitive sensors by using a pragmatic specification of logical values for three above mentioned logical worlds.

At first, let us represent Vasiliev's sensor by the logical world

$$\boldsymbol{LW}_{\mathbf{Vas3}} = \langle\{\boldsymbol{T}, \boldsymbol{B}.\boldsymbol{F}\}, >_{\mathbf{V}}, \{\boldsymbol{F}\}\rangle, \tag{3}$$

where the choice of anti-designated value F is explained by the need in designating bad or even forbidden sensor values. So Vasiliev's sensor is a cognitive sensor equipped with a three-valued "traffic lights" pragmatics: (1) T—"measured truth" ("norm"—sensor data are located in a "green zone"); (2) F—"measured falsity" ("out of norm"—sensor data are located in a "red zone"); (3) B—"measured ambiguity" ("boundary situation"—sensor data are located in a "yellow zone"). It is a typical example of logical pragmatic granulation.

At second, instead of (3) we take into consideration the case of sensor's exhausted resources (it can occur in wireless sensor networks): (4) N—"total uncertainty" (the sensor is sleeping—blue-colored sensor value). The appropriate Kleene's sensor is given by

$$\boldsymbol{LW}_{\mathbf{K3}} = \langle\{\boldsymbol{T}, \boldsymbol{N}, \boldsymbol{F}\}, >_{\mathbf{V}}, \{\boldsymbol{N}, \boldsymbol{F}\}\rangle. \tag{4}$$

Finally, Belnap's sensor characterized by four-valued pragmatics (1)–(4) is specified as follows

$$\boldsymbol{LW}_{\mathbf{Bel4}} = \langle\{\boldsymbol{T}, \boldsymbol{B}, \boldsymbol{N}, \boldsymbol{F}\}, >_{\mathbf{V}}, \{\boldsymbol{N}, \boldsymbol{F}\}\rangle. \tag{5}$$

Here the use of pragmatic truth, falsity, ambiguity and uncertainty concepts for understanding the measurement results is an intrinsic capacity of cognitive sensor.

5 Interpretation of Multi-sensor Data: Bilattice Pragmatics

Now let us take the concept of cognitive sensor network including homogeneous sensors, for instance Vasiliev's sensor network or Belnap's sensor network. The pragmatics of Vasiliev's network is given by 3^n, and Belnap's network—by 4^n, where n is an integer, $n > 1$. So a simple networked structure generated from two Belnap's sensors forms a set of pragmatic values $4^2 = 16$, the network generated from three Belnap's sensors—$4^3 = 64$, and so on.

A general approach to constructing interpretation granules from multi-sensor data supposes the use of product lattices and multi-lattices. Following Shramko [23], we specify a complex logical world by n-dimensional logical multi-lattice or simply a logical n-lattice as a structure

$$\boldsymbol{ML} = \langle \boldsymbol{V},\ \leq_{\boldsymbol{1}}, \ldots, \leq_{n} \rangle, \tag{6}$$

where V is a non-empty set of logical values, for instance, $V = V_1 \times \cdots \times V_{\mathrm{n}}$ and $\leq_1, \ldots, \leq_n$ are partial order relations defined on V such that $(V, \leq_1), \ldots, (V, \leq_n)$—different lattices.

We shall take as a basic unit of cognitive sensor network a minimal micro-net that consists of two-sensors. Then a considered problem is reduced to constructing a set of pragmatic values by using bilattices [25, 26]. Informally, bilattice may be seen as a product lattice with two different order relations, where a link between these orders is given by a compound heterogeneous negation operation extending Belnap's negation.

A pragmatic bilattice of logical values is defined by a quadruple [24]

$$\boldsymbol{BL} = \langle \boldsymbol{V},\ \leq_{\mathrm{v}}, \leq_{\boldsymbol{i}}, \neg_{\boldsymbol{G}} \rangle, \tag{7}$$

where $V = V_1 \times V_2$ is a product set, $\leq_{\mathrm{v}}$, $\leq_{\mathrm{i}}$ are two order relations, for example, truth order $\leq_{\mathrm{v}}$ and information order $\leq_{\mathrm{i}}$, $\neg_{\mathrm{G}}$ is Ginsburg's negation (it may be also interpreted as semi-negation, semi-affirmation operation).

It is obvious that the bilattice (7) can be viewed as algebra with two different meet and join operations

$$\boldsymbol{BL} = \langle \boldsymbol{V}, \wedge, \vee, \otimes, \oplus, \neg_{\boldsymbol{G}} \rangle, \tag{8}$$

where (1) $\langle V, \wedge, \vee \rangle$ and $\langle V, \otimes, \oplus \rangle$ are complete lattices; (2) $\neg$ is a mapping $\neg$: $V \rightarrow V$, such that: (a) $\neg^2 = 1$; (b) $\neg$ is a lattice homomorphism of $\langle V, \wedge, \vee \rangle$ to $\langle V, \vee, \wedge \rangle$ and a lattice homomorphism of $\langle V, \otimes, \oplus \rangle$.

Let us interpret basic granular pragmatic truth values for two sensors in bilattices:

- T_1T_2—"measured concerted truth" (data from both sensors take norm values);
- F_1F_2—"measured concerted falsehood" (data from both sensors are out of norm that witnesses for object's failure state);
- T_1B_2–B_1T_2—"measured partial contradiction as the first-order fault" (the first sensor shows a norm value and the second sensor gives partial fault, and vice versa);
- T_1N_2–N_1T_2—"measured partial truth with uncertainty" (the first sensor indicates a norm value and the second sensor sleeps, and vice versa);
- T_1F_2–F_1T_2 "measured full contradiction" (the first sensor indicates a norm value, and the second sensor informs about failure state, and vice versa);
- B_1B_2—"measured concerted ambiguity" (the data from both sensors inform about first order fault);
- N_1N_2—"total uncertainty" (the resources of both sensors are exhausted or both sensors sleep);

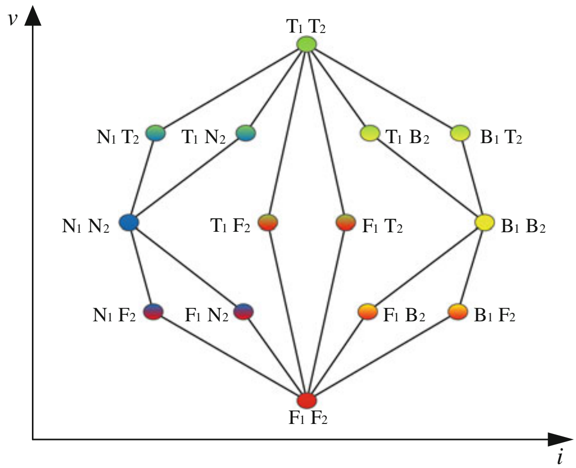


Fig. 8 Bilattice-based representation of logical pragmatics for two Belnap's sensors

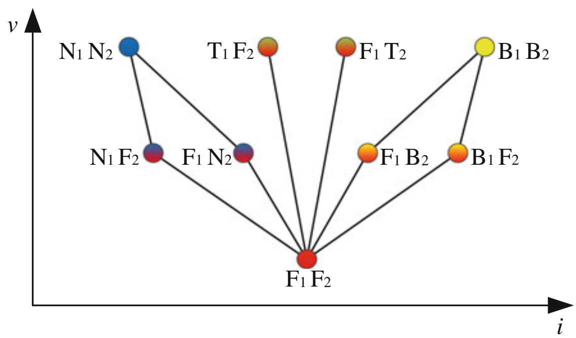


Fig. 9 Logical map of fault values for the structure of two Belnap's sensors

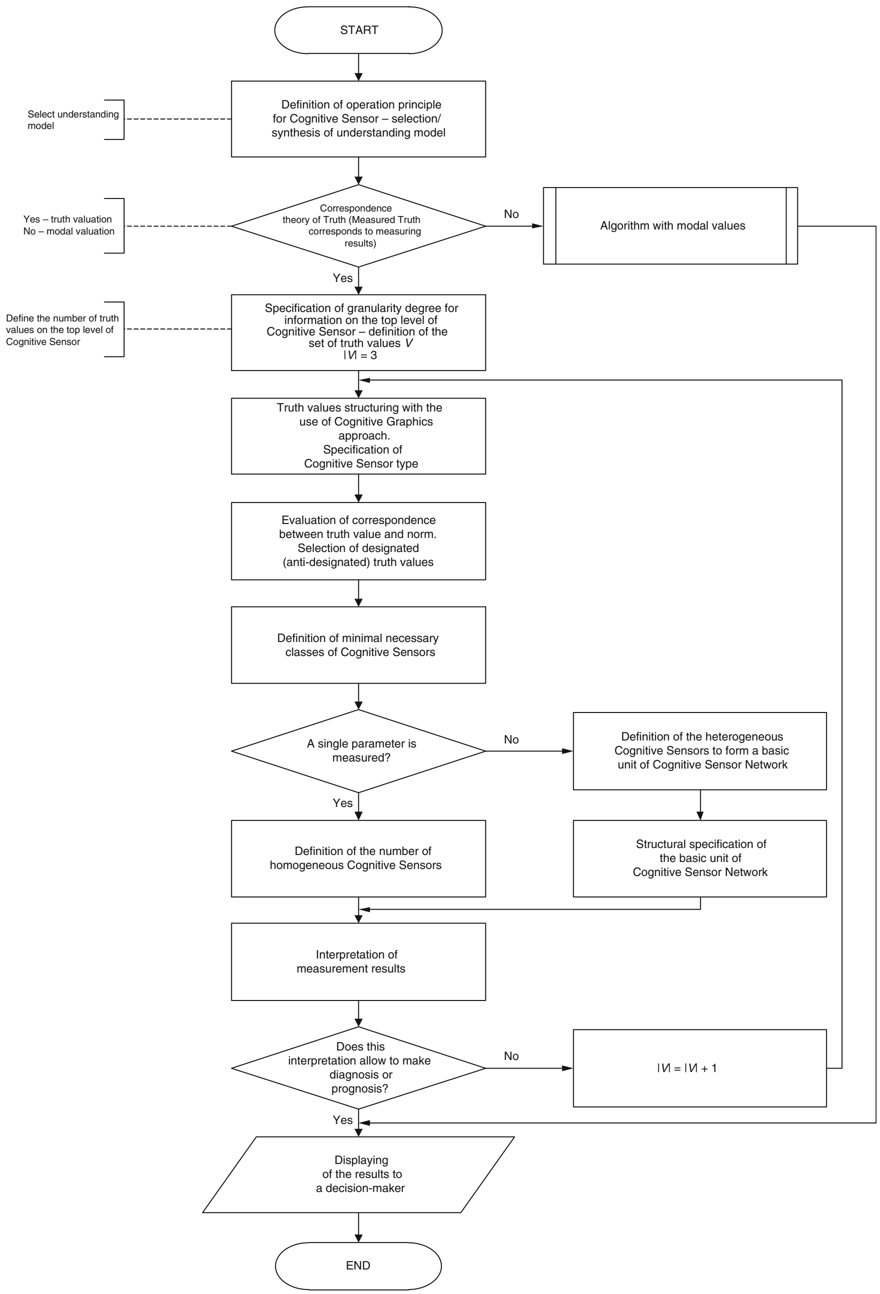


Fig. 10 An algorithm of constructing cognitive sensor for monitoring railway infrastructure objects

- F_1B_2–B_1F_2—"measured partial contradiction as the second-order fault" (the first sensor indicates an out of norm value and the second sensor indicates a fault), and vice versa;
- F_1N_2–N_1F_2—"measured partial falsehood with uncertainty" (the first sensor indicates a failure value and the second sensor sleeps), and vice versa.

The sensor data pragmatics for two Belnap's sensors is given by double Hasse diagram (Fig. 8), and logical map of fault values in shown in Fig. 9.

Since in Belnap's logic the values of B and N are considered independently, its conservative extension does not admit compound values B_1N_2 and N_1B_2. Thus, for two Belnap's sensors we obtain $|V| = 14$.

6 Algorithm of Logical-Algebraic Synthesis for Cognitive Sensors

As a crucial result of our logical-algebraic approach, an algorithm of constructing cognitive sensors has been proposed. Its flowchart is given in Fig. 10. The key steps of logical-algebraic synthesis are: (1) specification of operation principle for cognitive sensor related to a pre-selected understanding model; (2) definition of required measurement precision and granularity degree for pragmatic evaluations; (3) specification of initial cognitive sensor type with the possibility of its further change on the basis of algebraic model of logical values; (4) selection of necessary composition and number of cognitive sensors; (5) primary and final interpretation of measurement results for solving monitoring problems.

References

1. Shabelnikov, A.N., Shabelnikov, V.A., Kovalev, S.M.: Intelligent systems for distributed monitoring on the basis of wireless sensor networks with using systems of mobile agents. In: Proceedings of the 5th International Conference on Integrated Models and Soft Computing in Artificial Intelligence, vol. 1, pp. 538–543. Physmathlit, Moscow (2009)
2. Dolgiy, A.I., Kovalev, S.M., Samsonov, V., Khatlamadzhiyan, A.E.: Processing of fuzzy graphic images in intelligent computer vision systems on railway transport. In: Proceedings of the 9th International Conference on Application of Information and Communication Technologies (AICT-2015), pp. 118–121. SFU Publishers, Rostov-on-Don (2015)
3. Aarts, E., Harwig, R., Schuurmans, M.: Ambient intelligence. In: Denning, P.J. (ed.) The Invisible Future: The Seamless Integration of Technology into Everyday Life. McGraw-Hill Companies, New York (2001)
4. Nakashima, H., Adhajan, H., Augusto J.C. (eds.): Handbook of Ambient Intelligence and Smart Environments. Springer, Heidelberg (2010)
5. Chong, N.-Y., Mastrogiovanni, F. (eds.).: Handbook of research on Ambient Intelligence and Smart Environments: Trends and Perspectives. IGI Global, New York (2011)
6. Tarassov, V.B.: From Multi-Agent Systems to Intelligent Organizations. Editorial URSS, Moscow (2002). (in Russian)

7. Batyrshin, I.Z., Tarassov, V.B., Yazenin. A.V., et al.: Fuzzy Hybrid Systems. In: Yarushkina, N.G. (ed.). PhysMathLit, Moscow (2007) (in Russian)
8. Tarassov, V.B.: From hybrid systems to ambient intelligence. In: Proceedings of the 1st International Symposium on Hybrid and Synergistic Intelligent Systems: Theory and Practice, part 1, pp. 42–54. Kant Baltic Federal University Ed., Kaliningrad (2012) (in Russian)
9. Weiser, M.: Hot topics: ubiquitous computing. IEEE Comput. **26**(10), 71–72 (1993)
10. Finkelstein, L.: Intelligent and knowledge-based instrumentation. An examination of basic concepts. Measurement **14**, 23–30 (1994)
11. Reznik, L.: What is intelligent measurement? In: Zadeh, L., Kacprzyk, J. (eds.) Computing with Words in Information/Intelligent Systems, pp.78–97. Physica-Verlag, New York (1999)
12. Tarassov, V.B., Svyatkina, M.N.: Intelligent systems on the basis of cognitive measurements In: Proceedings of the 7th International Conference on Integrated Models and Soft Computing in Artificial Intelligence, vol. 2, pp. 611–623, Physmathlit, Moscow (2013) (in Russian)
13. Tarassov, V.B., Svyatkina, M.N.: Using Zadeh's granulation concept: granular logics and their application to sensor data analysis. In: Recent Developments and New Directions in Soft Computing, pp. 453–466. Springer, New York (2014)
14. Knorring,V.G.: Gnoseotechnique: a technique of cognition. Measurem. Inspect. Autom. №.1–2, pp. 3–12 (1992) (in Russian)
15. Rozenberg, V.Ya.: The development of conceptual-terminological apparatus of metrology on the basis of new information technology. Measurem. Tech. №11, pp. 20–22 (1990) (in Russian)
16. Prokopchina, S.V.: Cognitive measurements on the basis of Bayesian intelligent technologies. In: Proceedings of the XIIIth International Conference on Soft Computing and Measurement, pp. 28–34. LETI Publishers, St. Petersburg (2010) (in Russian)
17. Tarassov, V.B., Svyatkina, M.N.: Cognitive measurements for intelligent systems. In: Aliev, R.A. et al. (eds.) Proceedings of the 7th International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control (ICSCCW'2013), pp. 189–202. b-Quadrat Verlag, Kaufering (2013)
18. Zadeh, L.A.: Toward a Generalized Theory of Uncertainty (GTU): an outline. Inf. Sci. Inf. Comput. Sci. **172**(1–2), 1–40 (2005)
19. Potter, R.W.: The Art of Measurement. Prentice Hall, Englewood Cliffs (2000)
20. Reznik, L., Kreinovich, V. (eds.): Soft Computing in Measurement and Information Acquisition. Springer, New York (2003)
21. Tarassov, V.B.: Problem of understanding: present and future of artificial intelligence. In: Proceedings of the 5th International Conference on Open Semantic Technologies for Intelligent Systems (OSTIS-2015), pp. 25–42. BSUIR Editions, Minsk (2015)
22. Karpenko, A.S.: Development of Multi-Valued Logics. Editorial URSS, Moscow (2014)
23. Shramko, Y., Wansing, H.: Truth and Falsehood: An Inquiry into Generalized Logical Values. Springer, Berlin (2012)
24. Dunn, J.M.: Intuitive semantics for first-degree entailment and "coupled trees". Philos. Stud. **29**, 149–168 (1976)
25. Ginsberg, M.: Multi-valued logics: a uniform approach to reasoning in AI. Comput. Intell. **4**, 256–316 (1988)
26. Fitting, M.: Bilattices and the theory of truth. J. Philos. Logic **19**, 225–256 (1989)

Case-Based Reasoning Module for Intelligent Decision Support Systems

Alexander Eremeev, Pavel Varshavskiy and Roman Alekhin

Abstract In this paper the problem of the application of case-based reasoning in intelligent decision support systems (IDSSs) is considered. The hybrid method of case-based reasoning for the solution of problems of diagnostics and forecasting in (IDSS) is described. This paper demonstrates how the case-based reasoning module can be used in (IDSS), including IDSS of real time.

Keywords Analogous reasoning · Case-based reasoning · Intelligent decision support systems

1 Introduction

One approach to solving the problem of modeling commonsense reasoning in artificial intelligence (AI) systems and especially in intelligent decision support systems (IDSSs) is to use inductive reasoning, temporal reasoning, fuzzy logic as well as methods of reasoning based on analogies and precedents (cases) [1–5].

IDSSs are usually characterized by strict constraints on the duration of the search for the solution. One should note that, when involving models of case-based in

This work was supported by RFBR (projects № 14-01-00427-a, № 15-07-04574-a and project on the state task № 2.737.2014/K)

A. Eremeev (✉) · P. Varshavskiy · R. Alekhin
Applied Mathematics Department of the National Research University “MPEI”, Krasnokazarmennaya str., 14, Moscow 111250, Russia
e-mail: eremeev@appmat.ru

P. Varshavskiy
e-mail: VarshavskyPR@mpei.ru

R. Alekhin
e-mail: r.alekhin@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 450, DOI 10.1007/978-3-319-33609-1_18

IDSS, it is necessary to take into account a number of the following requirements to systems of this kind [1, 2]:

- the necessity of obtaining a solution under time constraints defined by real controlled process;
- the necessity of taking into account time in describing the problem situation and in the course of the search for a solution;
- the impossibility of obtaining all objective information related to a decision and, in accordance with this, the use of subjective expert information;
- multiple variants of a search, the necessity to apply methods of plausible (fuzzy) search for solutions with active participation of a decision making person (DMP);
- nondeterminism, the possibility of correction and introduction of additional information in the knowledge base of the system.

Case-based reasoning (CBR) can be used in various applications of AI and for solving various problems, e.g., for diagnostics and forecasting or for machine learning [1, 2, 4].

CBR may be applied in units of analysis of the problem situation, search for solutions, learning, adaptation and modification, modeling, and forecasting. The use of the respective methods in IDSS broadens the possibilities of IDSS and increases the efficiency of making decisions in various problem (abnormal) situations.

2 Case-Based Reasoning

2.1 CBR Cycle

CBR based on the accumulation of experience and the subsequent adaptation of previously successful solutions to similar new problems. This approach allows to simplify the decision-making process in condition of time limits and in the case of unexpected (abnormal) situations that may occur in the presence of various kinds of factors (incompleteness, inconsistency, uncertainty, etc.) In the input data and expert knowledge. The precedent is defined as a case that took place earlier and serving as an example or justification for future cases of this kind.

The processes involved in CBR can be represented by a CBR cycle (Fig. 1).

According to Aamodt and Plaza [6, 7], the CBR cycle comprising the four steps:

- RETRIEVE the most similar case(s) from the case library (CL);
- REUSE the retrieved case(s) to attempt to solve the current problem;
- REVISE the proposed solution in accordance with the current problem if necessary;
- RETAIN the new solution as a part of a new case.

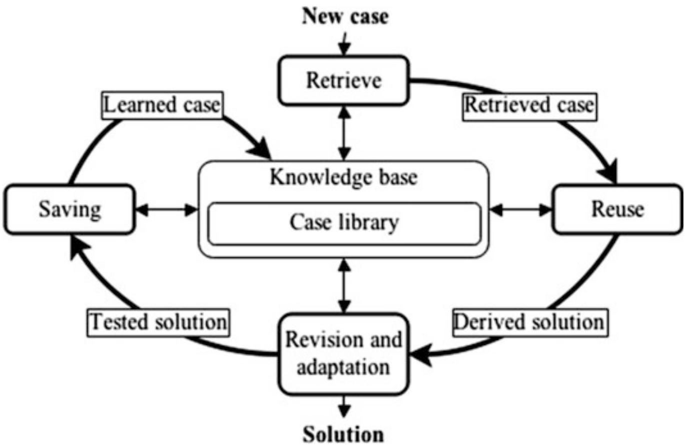


Fig. 1 CBR cycle

Use of the mechanism of cases for IDSS consists in output of the decision to the operator (DMP) for a current situation on the basis of the cases which contain in system. As a rule, last stage in a CBR cycle is excluded and realized by the expert (DMP) because the case library should contain only reliable information which is confirmed by the expert. Reconsidering and adaptation of the taken decision is required seldom because the same object (subsystem) is considered.

2.2 *Methods of Case Representation*

The successful implementation of CBR is necessary to ensure the correct case retrieval from CL. The choice of case retrieval method directly linked to the way of a case representation. There are different ways of the representation and storage of cases—from the simple (linear) to the complex hierarchical [3–5].

Generally, a case representation model comprises a description of the situation, the solution for this situation and the result of applying solution: *CASE = (Situation, Solution, Result)*, where *Situation* is the problem that describes the object state when the case occurred, *Solution* is the solution of the problem (e.g. diagnosis of problem situation and recommendations for DMP), *Result* is the outcome which describe the object state after the case occurred.

Different ways of case representation are concluded in different ways to describe these components. Cases can be represented in the form of database records, conceptual graphs, semantic nets, tree structures, predicates, frames and rules.

2.3 Advantages and Disadvantages of CBR Methods

The main advantages of CBR include the possibility to use the experience gained by the system for solution a new problem situation, without the intensive involvement of experts in a particular problem domain, and the exception of the repeated erroneous decision. In addition, CBR does not require an explicit problem domain model.

The disadvantages of CBR may include the following: the description of cases is usually limited to superficial knowledge of a problem domain; a large number of cases may lead to a decrease in system performance; complexities in definition of criteria for indexation and case comparison [7].

3 Method of Case Retrieval

3.1 Well-Known Methods for Case Retrieval

There are large number of case retrieval methods and their modifications, for example, the nearest neighbor (NN) method, induction method (based on decision trees), method on the basis of knowledge, method taking into account the applicability of cases, methods using neural network models and others. Well-known methods for case retrieval (NN, induction et al.) can be used alone or combined into hybrid retrieval strategies [5, 7].

The choice of case retrieval method is directly depends on the organization of CL and method of case representation. The NN algorithm and its modifications are used for the definition of a similarity for most common and simple parametrical case representation. For more complex case representations, like temporal or structural cases, the methods of case retrieval on the basis of structural analogy (e.g. structure-mapping theory (SMT) are used [7–11].

3.2 Structure-Mapping Theory

SMT allows formalize the set of implicit constraints, which are used by man who operates with concepts such as analogy or similarity. This theory uses the fact that an analogy is a mapping of knowledge of one domain (base) in another domain (target) based on the system of relations between objects of the base domain, as well as the target domain. The main principle of SMT is a principle of systematicity [8], that reflects the fact that humans (DMP) prefer to deal with a system of connected relations, not just with a set of facts or relations.

According to SMT, the analogous reasoning divides into three stages [9].

1. *Definition of potential analogues*. Given a current situation (target), define another situation (base) from the knowledge base that similar to it.
2. *Mapping and inference*. Build a mapping consisting of matchings between target and the base. Using this mapping we can obtain the candidate inferences, which represents the additional knowledge (facts) about the base that can be transferred to the target.
3. *Estimation*. Estimate the "quality" of obtained mapping. The structural criteria can be used, such as the degree of structural similarity.

Consider the structure-mapping engine (SME) which is based on SMT. The input data for the SME algorithm are structural representations of the base and target domains.

SME algorithm consists of the following steps.

1. *Construction of local mappings.* Determine the matchings (match hypotheses) between elements in the base and target domains.
2. *Construction of global mappings (Gmap)*. Form the mapping systems that use compatible pair of objects from local mappings.
3. *Construction of candidate inferences*. For each Gmap, construct a set of the facts (possibly empty) which occur in the base domain, but does not initially contains in the target domain.
4. *Estimation of Gmap*. Compute the global mappings structural estimation score taking into account the plausibility of local matchings.

Thus, as a result, the most systematic consistent mapping structure Gmap has following components [8, 9]:

- mapping is a set of matchings between base and target domains;
- candidate inferences is a set of new facts that presumably contained in the target domain;
- structure estimation is a numeric equivalent of the mapping quality based on the structural properties of Gmap.

The main advantages of SME, which are especially important for IDSSs, are the domain independence, polynomiality and the simplicity of importing the candidate inferences in the target domain.

3.3 Hybrid Method for Case Retrieval

This paper proposes a hybrid approach to finding solutions based on CBR. SME algorithm used to determine similarity at the first stage, and NN algorithm used at the second stage to compare values.

Case retrieval and determination of similarity of case and current situation is proposed to implement in two stages:

- determination of similarity of case and current situation on the basis of problem domain ontology and formation of pair matchings by an algorithm based on the SMT;
- determination of similarity of case and current situation by the nearest neighbor method taking into account the obtained pair matchings.

The descriptions of the current situation and the case are structurally compares at the first step. The purpose of this step—to identify possible pairwise matchings between case and current situation and estimate their similarity.

The pairwise matchings are formed by the following procedure *Pairs*:

Input: *Q*—current situation, *C*—case, *O*—problem domain ontology, CF_1 and CF_2—values to refine plausibility from SME algorithm.

Output: *F*—result pairwise matching set, *LS*—matching estimations set.

Local variables: *i, j*—loops parameters, *m*—result of equality check, *s*—result of similarity check.

Checks: *Eq(A,B)*—true if name of concept A is equal to name of concept B, false otherwise; *S(A,B)*—true if concept A is similar to concept B (we will determine the similarity between the concepts of ontology guided by the principle that the concepts A and B are similar, if the parent concept of A match the parent concept of B), false otherwise.

```
i ← 0
F ← {}
while (i < |Q|) {
 if (Q_i) {
  j ← 0
  while (j < |C|) {
   m ← Eq(Q_i, C_j)
   s ← S(Q_i, C_j)
   if (m OR s)
    F ← F ∪ (Pairs(Q_i, C_j, O) × <Q_i, C_j>)
   if (m)
    LS_j ← LS_j + CF_1
   if (s)
    LS_j ← LS_j + CF_2
   j ← j + 1
  }
 }
 i ← i + 1
}
return F
```

At the second stage to estimate similarity of the current situation and case the nearest neighbor method is used. For each matching pair in the selected metric the distance d_{CQ} between the current situation and case is defines. To estimate the similarity degree $Sim(C, Q)$ there are need to find the maximum distance d_{MAX} in the selected metric, using the bounds of range of parameters (x_i^{left} and x_i^{right}, $i = 1, \ldots, n$).

The result is a set of precedents, each of which corresponds with two estimates of similarity to the current situation, which can be expressed as a percentage:

- estimate on the basis of problem domain ontology: $Sstruct = \sum_{i=1}^{k} LS_i / SES_{MAX}$, k—number of correspondences, LS_i—plausibility estimation of i correspondence, SES_{MAX}—estimation for the case where a base is selected as the target;
- estimate by nearest neighbor method: $Sim(C, Q) = 1—d_{CQ}/d_{MAX}$, d_{CQ}—distance between the current situation and case, d_{MAX}—maximum distance in the selected metric.

Based on these two estimates, the DMP can choose the most appropriate precedent and get a solution to the current situation.

4 Implementation of CBR Module for IDSS

The proposed approach has been implemented in a CBR module prototype. Figure 2 shows the architecture of developed module.

Implemented CBR module prototype was examined by an example of solving a problem of complex object state diagnostics and detection of controlling impacts on the example of the automatic cooling system of pressurized water reactor of nuclear power plant [2, 9]. On the basis of technological rules and operational guidelines ten initial cases were established. We give a semantic interpretation of one of them (Fig. 3) and new problem situation on object (Fig. 4):

- it is recommended to inject TH11D01 with boric concentrate 40 g/kg caused by switching off ACS 1 (automatic cooling system) due to closing the gates TH11S24 and TH11S25;
- ACS is switched off due to the closed gates TH12S24 and TH12S25;
- the upper setting T517B01 (pressure in the container of ACS 1) is equal to 63;
- the lower setting T517B01 (pressure in the container of ACS 1) is equal to 56;
- the upper setting TH11T500 (temperature in the frame of ACS 1) is equal to 60;
- the lower setting TH11T500 (temperature in the frame of ACS 1) is equal to 20.

As a results of CBR module (Fig. 5)—DMP has the ability to choose the most appropriate case based on two similarity estimates: structural estimation based on the SMT; estimation based on the nearest neighbor method.

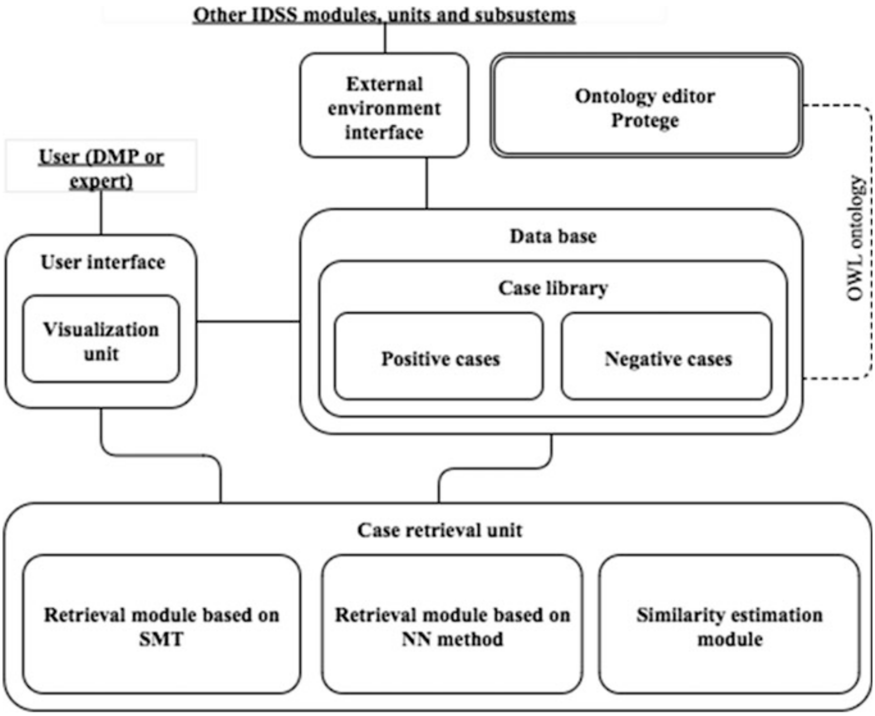


Fig. 2 Architecture of CBR module

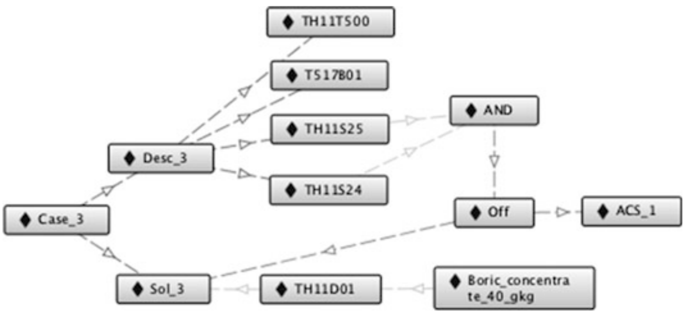


Fig. 3 Case representation

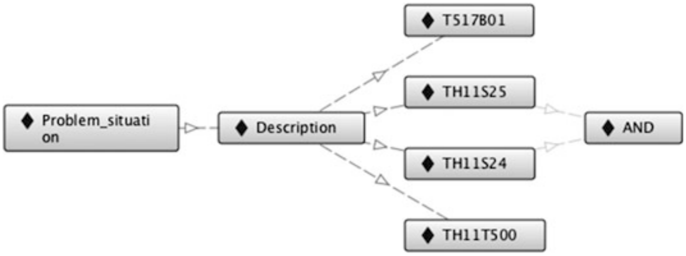
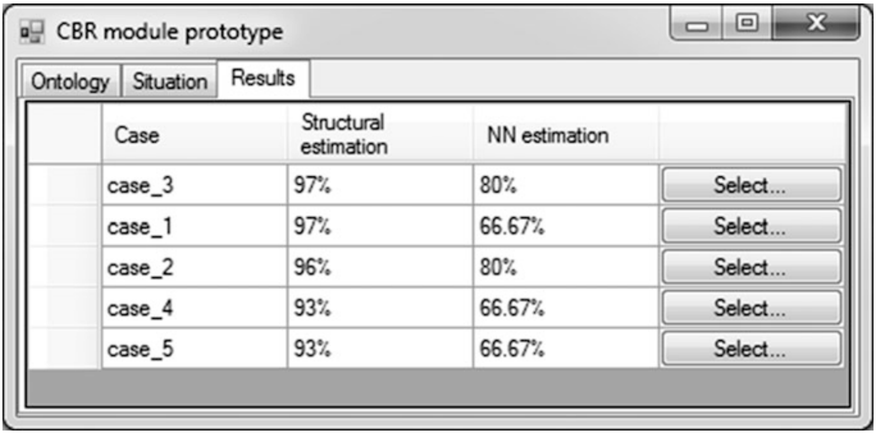


Fig. 4 Problem situation representation



Case	Structural estimation	NN estimation	
case_3	97%	80%	Select...
case_1	97%	66.67%	Select...
case_2	96%	80%	Select...
case_4	93%	66.67%	Select...
case_5	93%	66.67%	Select...

Fig. 5 Results of CBR module

5 Conclusions

Methods of the search for a solution on the basis of a structural analogy and cases were considered from the aspect of their applications in modern IDSS (including IDSS of real time), in particular, for a solution of problems of diagnostics and forecasting. The CBR cycle is considered and different methods of case representation and retrieval are investigated. This paper proposes a hybrid approach to finding solutions based on CBR. SME algorithm used to determine similarity at the first stage, and NN algorithm—at the second stage to compare values. The CBR module was considered from the aspect of its application in IDSS, in particular, for a solution of problems of complex object state diagnostics and detection of controlling impacts on the example from power engineering-operation control of nuclear power unit.

References

1. Eremeev, A.P., Vagin, V.N.: Methods and tools for modelling reasoning in diagnostic systems. In: ICEIS 2009—11th International Conference on Enterprise Information Systems, Proceedings, AIDSS, pp. 271–276 (2009)
2. Varshavskii, P.R., Eremeev, A.P.: Modeling of case-based reasoning in intelligent decision support systems. Sci. Tech. Inf. Process. **37**(5), 336–345 (2010)
3. Wilson, D.C., Leakes, D.B.: Maintaining case-based reasoners: dimensions and directions. Comput. Intell. **17**(2), 196–213 (2001)
4. Leake, D.B.: CBR in context: the present and future. Case-Based Reasoning: Experiences, Lessons and Future Directions, pp. 3–31. AAAI Press/The MIT Press (1996)
5. Eremeev, A.P., Kurilenko, I.E.: Temporal reasoning component for realtime intelligent decision-support systems. Sci. Tech. Inf. Process. **38**(5), 332–343 (2011)
6. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Communications, vol. 7(1), pp. 39–59. IOS Press (1994)
7. Eremeev, A., Kurilenko, I., Varshavskiy, P.: Application of Temporal Reasoning and Case-based Reasoning in Intelligent Decision Support Systems. International Book Series “Information science & Computing”, Number 10, Supplement to International Journal “Information technologies & Knowledge”. Volume 3/2009, pp. 9–16 (2009)
8. Falkenhainer, B., Forbus, K., Gentner, D.: The structure-mapping engine: algorithm and examples. Artif. Intell. **41**, 1–63 (1989)
9. Eremeev, A., Varshavsky, P.: Analogous reasoning and case-based reasoning for intelligent decision support systems. Int. J. Inf. Theor. Appl. (ITHEA) **13(**4), 316–324 (2006)
10. Eremeev, A., Varshavsky, P.: Methods and tools for reasoning by analogy in intelligent decision support systems. In: Proceedings of the International Conference on Dependability of Computer Systems. Szklarska Poreba, Poland, 14–16 June 2007, IEEE, pp. 161–168
11. Eremeev, A., Kurilenko, I., Varshavskiy, P.: Temporal case-based reasoning system for automatic parking complex. World academy of science, engineering and technology, international science index 101. Int. J. Comput. Electr. Autom. Control Inf. Eng. **9(**5), 1232–1238 (2015). http://waset.org/publications/10001487/temporal-case-based-reasoning-system-for-automatic-parking-complex

Automatic Document Classification Based on J.S. Mill's Ideas

Nicolay Lyfenko

Abstract The paper describes an approach to automatic classification problem using J.S. Mill's ideas of inductive logic. The technique uses the general principles (but not the technical details) of the JSM method of automatic hypothesis generation. The proposed method uses some *induction* procedure to form generic positive and generic negative objects (represented as vectors) and an *analogy* procedure to classify new documents. With an optimal selection of text preprocessing options, the suggested approach shows better precision than other implemented text classification methods.

Keywords Machine learning • Data mining • Natural language processing • JSM method

1 Introduction

Text classification problem is becoming even more important due to the exponential increase in the amount of data on the Internet and because of the development in technologies for processing vast data arrays, including documents in the natural language [1].

The most popular and effective text classification methods are: *k-nearest neighbor*, *support vector machines*, *naïve Bayes classifier* and *boosting-based systems*. However, the high efficiency of text classification also depends on the quality of document pre-processing and the selection of text representation models. The classical approaches of finding numerical representations of texts are: *tf-idf* in different variations, *the most frequent sequences* [2], *information gain*, *chi-square statistics*.

N. Lyfenko (✉)
Russian State University for the Humanities, Moscow, Russia
e-mail: lyfenkoNick@ya.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 450, DOI 10.1007/978-3-319-33609-1_19

In many systems of this kind texts are interpreted as feature vectors by well-accepted statistical methods [3]. These do not take account of any syntactic or semantic relationships among words in the text and make use of the local frequency of word occurrence in the text. However, the most semantically significant terms may have low frequency and in fact a little significance in the feature vector. In such a case, these terms may happen to become irrelevant in text classification process that makes the precision of categorization method less accurate.

This is the reason that this paper provides *document frequency* text representation model and some versions of text classification methods that are based on Mill's ideas. We are using not the JSM technique but only the ideas that are found at the heart of the classic JSM method [4]. Based on the fact that JSM method uses some of Mill's ideas literally and this logical-combinatorial method applies efficiently to small amounts of data but it does not work as well with increased data amount it is reasonable to use Mill's ideas and adapt his canons for stochastic relationships.

The current idea of JSM method was described by V.K. Finn in a lengthy article [4] and by Anshakov [5].

The proposed text classification method (Generic Object Based Inductive Method, GOBIM) uses two procedures: *induction* to form a generic positive and generic negative vectors and *analogy* to attach a new document to a predefined class. Text representation of a new vector and a generic one can be different.

More details about the described method and the text classification scheme one can find at works [6, 7]. The current paper is novel in the sense that: in the first place, here we give a new presentation of theory and in the second place, we run a new series of tests with a number of feature selection options on different subject domains.

2 Automatic Text Classification Method Based on J.S. Mill's Ideas

This section is written to show how to derive the appropriate informal rules of induction for stochastic data analysis from the informal rules of J. S. Mill's inductive logic [8] using the ideas from V. K. Finn's JSM method [9].

Usually, an object of JSM method is represented as a set of atoms. This set is convenient to interpret as a bit vector and the atoms proper are convenient to identify with attributes. The bit vector can be expanded by adding a field with the target property (which is also considered as a feature). In this instance we attain the object representation that is commonly used in different machine learning methods.

Thus, the data for JSM method can be represented as a bit matrix with strings that correspond to objects and columns that confirm to features (target properties included).

JSM method is used to work with three sorts of entities: *objects, object fragments, and target properties*. We define the *positive instance* and the *negative* *instance* (for the target property) as objects with and without target property. Object *fragments* are considered possible causes for both the presence and absence of the target property. The cause for its presence is called a *positive cause* or a plus cause. The cause for its absence is called a *negative cause*, a minus cause, or an anti-cause.

We use a "bag of words" model with extensions: terms are series of words with the length equal to 1, 2 or 3. They are nouns, adjectives, and verbs filtered by using *document frequency*.

2.1 Notation

Assume that d is a document from class of documents D, then the map $v^d : T \to \{0, 1\}$ is called a *vector representation* of document $d \in D$ if

$$v^d(t) = \begin{cases} 1, \ t \in d, \\ 0, \ t \notin d. \end{cases} \tag{1}$$

Suppose $D' \subseteq D$, S is an arbitrary non-empty set. Then we say that the map $v^{D'} : T \to S$ is a *generic vector* for D' if we can define an analytic representation for $v^{D'}$ by means of some variant of stochastic generalization method.

For *simple summation method*, $S = \mathbf{N}$ ($\mathbf{N}$ is a set of natural numbers),

$$v_S^{D'}(t) = \sum_{d \in D'} v^d(t) \tag{2}$$

For *document frequency method*, $S = \mathbf{R}$ ($\mathbf{R}$ is a set of real numbers).

$$v_F^{D'}(t) = \frac{\sum_{d \in D'} v^d(t)}{|D'|} \tag{3}$$

However, the JSM method always works not with one set of learning examples, but with two: the set of positive examples, here is marked as D^+; and the set of negative examples, here marked as D^-. The construction of the generic vector is done with the help of these two sets.

2.2 Induction

The training phase in machine learning corresponds to the *induction procedure* of JSM method. During this phase two generic vectors are formed. The main idea of induction phase for GOBIM is similar to the rules of induction of JSM Method.

The rules of induction of JSM method are: (Plus rule) *Assume that f is a common fragment of two or more positive instances for the target property and f is not a common fragment for two or more negative instances for this property. Then, f is a possible cause for the presence of the target property.* Minus rule is formulated symmetrically.

Then we will try to provide informal descriptions of rules that are suitable for retrieving regularities from stochastic data.

We note that JSM rules take into account two or more positive instances and the absence of negative instances. This cannot be assumed so definitely for stochastic data. The rules should be described informally, using indistinct restrictions, such as *many* or *few*. They are formalized subsequently by statistical criteria.

(Plus rule) *Assume that there are quite a large number of cases of the occurrence of fragment f in positive instances for the target property, and quite a few cases of the occurrence of f in negative instances. Then, f is a possible cause for the presence of the target property*. Minus rule is formulated symmetrically.

The rules of JSM induction, where statistical considerations could be used, have already been introduced in earlier works, for example, [10–13]. It should be noted that all the enumerated papers above, however, suggest the use of the standard JSM technique, which high computational complexity makes its application to vast amounts of data inconvenient. This paper suggest an entirely different and original approach, based on forming the generic objects: typical (generalized) positive instance and the typical (generalized) negative instance.

The copy vector is a generalized vector that represents an entire class of documents. The positive copy vector v^{D^+} implicitly contains all of the possible causes for the existence of the target property and the negative copy vector v^{D^-} implicitly contains all of the possible causes for the nonexistence of the target property. Copy vectors can be formed by different methods. The common idea is to emphasize the fairly frequent actual occurrence of terms in class documents. If a term frequently occurs in the positive documents and rarely occurs in the negative ones, it will have a high weight in v^{D^+}.

This approach agrees quite well with the information rule of induction that says that a fragment that is eligible to be the possible cause for a property's existence is found in a sufficiently large number of positive instances and a sufficiently small number of negative instances.

Assume that $v_M^{D^+ \Delta D^-}(t)\colon T \to \{0, 1\}$ is a mapping for calculating a *binary generic vector for a pair* of v^{D^+} and v^{D^-}, then

$$v_M^{D^+ \Delta D^-}(t) = \operatorname{sign}\left(v_M^{D^+} \Delta v_M^{D^-}\right) = \begin{cases} 1, & v_M^{D^+} \Delta v_M^{D^-} > 0 \\ 0, & v_M^{D^+} \Delta v_M^{D^-} \le 0 \end{cases} \quad (4)$$

To calculate generic positive and negative vectors the common equation is used. It depends on the operator (marked as Δ) and on the method of generalization (marked as M).

Where M can be *simple summation* method or *document frequency method*, operator Δ can be *difference* or *division*. Let introduce some possible variants for calculating $v_M^{D^+ \Delta D^-}(t)$ and $v_M^{D^- \Delta D^+}(t)$.

Division method for calculating binary generic vector of a pair

$$v_M^{D^+/D^-}(t) = \frac{v_M^{D^+}(t)}{v_M^{D^-}(t)} \tag{5}$$

$$v_M^{D^-/D^+}(t) = \frac{v_M^{D^-}(t)}{v_M^{D^+}(t)} \tag{6}$$

Proper subtraction method for calculating binary generic vector of a pair.

$$v_M^{D^+ \dot{-} D^-}(t) = v_M^{D^+}(t) \dot{-} v_M^{D^-}(t) \tag{7}$$

$$v_M^{D^- \dot{-} D^+}(t) = v_M^{D^-}(t) \dot{-} v_M^{D^+} \tag{8}$$

Here $\dot{-}$ is the *proper subtraction*, i.e.,

$$x \dot{-} y = \begin{cases} x - y, & x \geq y, \\ 0, & x < y. \end{cases} \tag{9}$$

2.3 Analogy

We will now describe the rules for generating predictions of the presence or absence of target properties in objects for which it is unknown. In the JSM method such rules are called rules of analogy. This name is given on the grounds that one can propose the conjecture that an object has a property if and only if this object resembles two or more other objects with this property and does not resemble objects without it. The similarity of objects is expressed exactly by their common fragment, which is defined as the cause of the presence or absence of the target property.

The informal analogical rule for our case can be formulated as follows: (Plus rule) An object possesses the target property if this object is essentially much more similar to positive instances (for this property) than negative. Minus rule is formulated symmetrically.

In our case the objects are documents; we will use vector modules and the cosine of the angle between document vector representation, expressed via the dot product to indicate the degree of similarity.

Let us use $\mathrm{sim}\left(\vec{a}, \vec{b}\right)$ to indicate the similarity degree between vectors $\vec{a}$ and $\vec{b}$ and calculate it using the following equation:

$$\mathrm{sim}\left(\vec{a}, \vec{b}\right) = \frac{\left(\vec{a}, \vec{b}\right)}{|\vec{a}| \cdot \left|\vec{b}\right|}. \tag{10}$$

In such a case the fact that the vector v that represents the document d^v is essentially much more similar to v^{D^+} than v^{D^-}, can be expressed using the following equation:

$$\mathrm{sim}\left(v^d, v^{D^+}\right) - \mathrm{sim}\left(v, v^{D^-}\right) \geq p^+, \tag{11}$$

where $p^+ \in (0, 1]$ is some threshold value.

The fact that the vector v that represents the document d^v is essentially much more similar to D^- than D^+, can be expressed using the following equation:

$$\mathrm{sim}\left(v^d, v^{D^-}\right) - \mathrm{sim}\left(v, v^{D^+}\right) \geq p^-, \tag{12}$$

where $p^- \in (0, 1]$ is some threshold value.

Thus, new document d, represented with vector v^d belongs to positive class D^+ if (11) holds and d belongs to negative class D^- if (12) holds. It is easy to see that (11) cannot hold simultaneously with (12). In addition, certain documents may appear to be unattributed to any of the classes. This is a typical case of JSM method. It means that it remains unknown whether the new object possesses the target property.

Other rules of identifying the relevance to classes D^+ and D^- are possible as well. Let us introduce some of them.

Simple intersection

$$d^v \in D^+ \ if \left|v^d \bigcap v^{D^+}\right| > \left|v^d \bigcap v^{D^-}\right| \tag{13}$$

$$d^v \in D^- \ if \left|v^d \bigcap v^{D^-}\right| > \left|v^d \bigcap v^{D^+}\right| \tag{14}$$

More flexible rules can be derived by including the weights of positive and negative predictions in the formulas:

Weighted intersection

$$d^v \in D^+ \ if \left|v^d \bigcap v^{D^+}\right| \cdot w^- > \left|v^d \bigcap v^{D^-}\right|, \tag{15}$$

$$d^v \in D^- \ if \left|v^d \bigcap v^{D^-}\right| \cdot w^- > \left|v^d \bigcap v^{D^+}\right|, \tag{16}$$

where w^- and w^- are the weight coefficients for positive and negative predictions.

To make more accurate predictions, we will take account not only of the occurrence of terms in the document but also of their nonoccurrence as well. We will expand the feature vector by adding to it "anti-terms" which weight is calculated using the formula:

$$v^d(\tilde{t}) = 1 - v^d(t) \quad (17)$$

3 Experiments

To attain relevant categorization results, the system requires fine adjustment, including the following set of Boolean options with marks:

1. the use of stop words in the feature vector (**SW**);
2. the filtering of terms (**F**);
3. the formation of terms consisting of noun groups only (**NP**);
4. the normalization of words (**N**);
5. the filtering of numbers (**D**).

Simple intersection (**u**), weighted intersection (**W**) cosine similarity (**C**) and extension (**E**) methods were implemented and tests were run.

The test involved two collections of 6012 documents in total: the first collection consisted of 3874 documents for the positive class and the second collection consisted of 2138 documents for the negative class. The problem set for the system was to perform a subject categorization of 2575 documents that had no class label yet. *Positive documents* include texts on political issues, *negative documents* encompass texts that are unrelated to politics, i.e., documents on sports, economy, accidents, culture, and healthcare. All these documents were taken from the automatically downloaded news feed of the web version of weekly newspaper *Argumenty i Fakty* for the period from February 10 to August 10, 2014. The class label for each document was known according to the data from the RSS feed. Then, in accordance with the above described rules, the obtained selection was divided in a training and test selection and the classifier was trained using the document *frequency method* and the *reduced difference* method for forming the binary generic vector of a pair (Fig. 1).

We should emphasize that for the methods *w* and *c* the reduction in size of the feature vector by *normalization*, consideration of *stop words*, *filtering*, and use of only *noun groups* leads to a far less accurate categorization (by 0.3 on average). In contrast, this index for the method of intersection and extension of feature vector ($u + e$) rises from 0.65 to 0.78, which allows us to characterize the applied approach as promising.

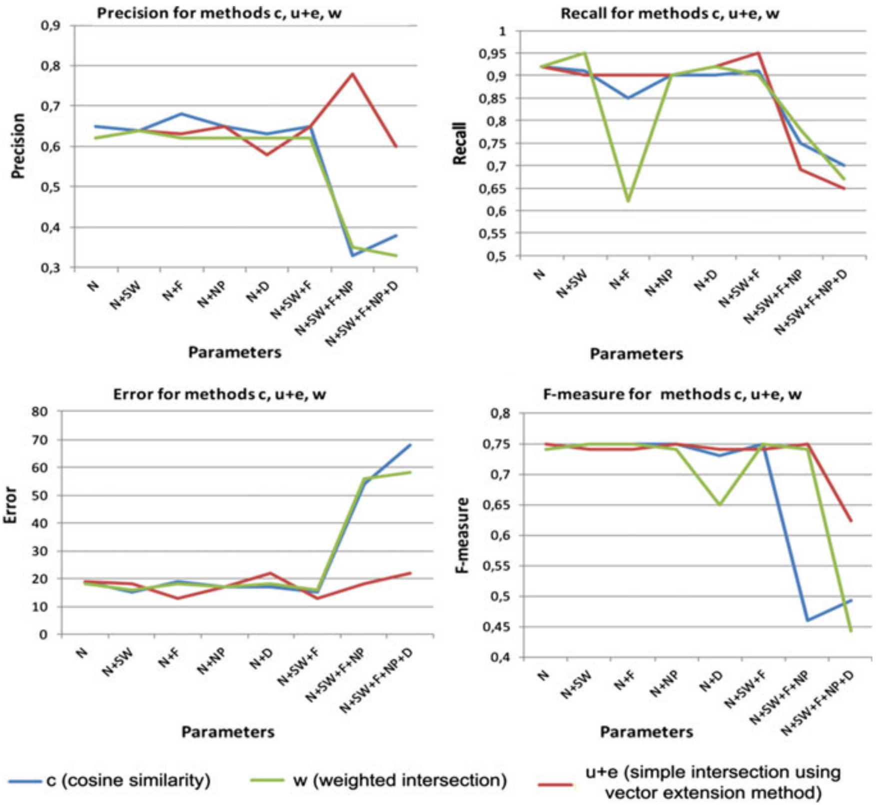


Fig. 1 Precision evaluation for methods c, w, u + e

4 Conclusions

In the article, we suggest an approach for text classification problem using J. S. Mill's ideas of inductive logic. Our tests have shown that the precision and recall of the general object inductive method in particular $u + e$ technique (*vector extension and using rules of intersection*) to categorization is commensurate with the results for *cosine similarity* and *weighted intersection* and even sometimes surpass this evaluation.

Some variants of classification rules, formulas for calculating generic object and generic binary vector were developed and implemented.

The follow-up activities include a series of tests on representative data from other subject domains, modification of algorithms for deciding on the class attribution of documents, and implementation of a JSM system algorithm that is more similar to the classical algorithm.

References

1. Sebastiani, F.: Text categorization. In: Zanasi, A. (ed.) Text Mining and Its Applications, pp. 109–129. Wit Press, Southampton (2005)
2. Yang, Y., Pederson, J.O.: A comparative study on feature selection in text categorization. In: Proceedings of ICML-97, pp. 412–420 (1997)
3. Bai, J., Nie, J.-Y.: Using language models text classifications. In: Processing of Asia Information Retrieval Symposium, Beijing (2004)
4. Finn, V.K.: Epistemological foundations of the JSM method for automatic hypothesis generation. Autom. Doc. Math. Lingust. **48**(2), 96–148 (2014)
5. Anshakov, O.M.: The JSM method: a set theoretical explanation. Autom. Doc. Math. Linguist. **46**(5), 202–220 (2012)
6. Lyfenko, N.D.: Automatic classification of documents in a natural language: a conceptual model. Autom. Doc. Math. Lingust. **48**(3), 158–166 (2014)
7. Lyfenko N.D.: An approach to text data categorization based on the ideas of J.S. Mill. Autom. Doc. Math. Lingust. **49**(6), 202–221 (2015)
8. Mill, J.S.: A System of Logic, Ratiocinative and Inductive. Harper & Brothers, NY (1882)
9. Finn, V.K.: Databases with incomplete information and a new method for automatic generation of hypotheses. In: Dialogovye i faktograficheskie sistemy informatsionnogo obespecheniya (Dialogue and Factual Information Support System), Moscow (1981)
10. Grigor'ev, P.A.: A method for automatic generation of hypotheses that is similar to JSM method: the use of statistical considerations. Nauchn.-Tekhn. Inform. Ser. 2. Protsessy Sist. 5–6, 52–55 (1996)
11. Grigor'ev, P.A., Sword-systems or JSM systems for chains using statistical considerations, Nauchn.-Tekhn. Inform., Ser. 2. Protsessy Sist. 5–6, 45–51(1996)
12. Anshakov, O.M.: Generalized quantifiers are defined using templates. Part I, Nauchn.-Tekhn. Inform. Ser. 2. Protsessy Sist. 11, 5–17 (2000)
13. Anshakov, O.M.: Generalized quantifiers are defined using templates. Part II, Nauchn.-Tekhn. Inform. Ser. 2.Protsessy Sist. 5, 35–48 (2001)

The Analysis of Diagrammatic Models of Workflows in Design of the Complex Automated Systems

Alexander Afanasyev, Nikolay Voit and Rinat Gaynullin

Abstract The workflows of industrial and design enterprises are represented by diagrams in notation languages: Business Process Model Language (BPML), Unified Model Language (UML), and Architecture of Integrated Information Systems (ARIS), Methodology Integration Definition (IDEF0 and IDEF3). To analyze, control and correct made by a designer during development errors in diagrams, the authors have developed a new temporal grammar called TRVM-grammar. It allows the authors to increase quality control and compatibility of the flows of the project works by 20 % in comparison with the competing use of preserving, positional and relational grammars.

Keywords Workflows · Grammar · Checking

1 Introduction

Over the past 20 years in the field of management of automated systems development important problems of creation, organization, analysis and control of workflows and their implementation of real systems development in practice have been solved. In [1] the author gives a detailed review of methods and tools of workflow organization. It includes the following phases: modeling the workflow "as it is"; their optimization through the creation of models "to be"; changing the

The reported study was partially supported by RFBR, research project №15-07-08268a.

A. Afanasyev (✉) · N. Voit · R. Gaynullin
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 450, DOI 10.1007/978-3-319-33609-1_20

organizational structure; "scrolling" the workflow; improving the process. In the research work [2] the authors also discuss the use of the following systems on the phases of workflow development: ERP-systems and workflow management systems (WfMS). Life cycle of complex automated systems is presented as a sequence of phases: 'Design and Analysis', 'Configuration', 'Decision' and 'Assessments'.

The dynamic development of complex automated systems is connected with the adaptation of workflows to changes in system requirements. "We consider enterprise/business agility as a property of an enterprise to function in the highly dynamic world" [3] and it has two features: (1) to adapt to changes in the environment; (2) to discover new opportunities constantly appearing in the dynamic world to launch completely new products (services). The agility implementation requires a new approach that allows designers to discover changes and opportunities of complex automated systems development, and to influence on them accordingly. The need to develop this approach appeared when the degree of changes in requirements of development increased. For example, in [4] the author says: "Industry and technology move too fast, requirements change at rates that swamp traditional methods."

Whitestein Technologies, Magenta Technologies, SkodaAuto, Volkswagen, Saarstahl AG note that the first generation of static management systems of product lifecycle and project workflows [5] can no longer meet the requirements of many companies, approach and automated tools of the first generation of project workflow standardization have already exhausted its resources, and as a result, there are poorly formalized (poor-quality) processes, stimulating the growth of expenses for their development and improvement.

The authors of this paper used the definition given in [6], for a dynamic stream of project work—stream of project work, adjusted to the environmental changes.

ProBis [7] is considered to be a traditional workflow management system. Dynamic project workflow management systems according to the works [8–10] include YAWL (Yet Another Workflow Language) and iPB.

All such systems use graphic primitives and relations in the form of diagrams to represent workflows in a graphical form. There are two tasks of structure analysis (syntax) and meaning (semantics) of actual diagrams. In [11] the author uses colored Petri nets for dynamic semantic analysis of workflows and the pi-Calculus approach formalizing workflows in algebraic expressions of first-order logic is used in [12]. There is also the modern theory of graphical languages for representing diagrams. This theory contains syntax models of the spatial and logical types, including attributes of graphical objects (e.g., rectangle or circle) and types of relations. The spatial model has relative or absolute coordinates of graphical objects. The use of spatial models is difficult to control, to analyze the structure or topology (syntax) and attributes of diagrams. As a rule, the logical model used by to describe the syntax diagrams, the graph grammar.

To represent the diagrammatic models, the modern theory of graphical and visual languages considers two basic syntax models called spatial and logical models. They are based on the attributes of graphic objects and types of their links.

The spatial model contains relative or absolute coordinates of graphic objects. The use of such model to control and analyze structural, i.e. topological (syntax), features of diagrams is very difficult. Therefore, to describe the syntax of diagrams the logical model is used. This model is processed by the graphical grammars.

John L. Pfaltz and Azriel Rosenfeld proposed web-grammars [13] of two-dimensional generating type. Zhang developed the positional graphical grammar relating to the context-free grammar, later it was developed by Costagliola [14]. Wittenberg and Weitzman [15] developed a relational graphical grammar. Zhang and Orgun [16, 17] described preserving graphical grammar in their works.

Most tools, supporting computer-aided design (IBM Rational Unified Process, ARIS), 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.

Actual proposals of this paper are: a metamodel depicts the different artifacts, methods and activities, and indicates how they are connected with the processes, that helps to solve a common problem in a software development project using Scrum [6] and user stories is that the big picture or strategic fit is often lost [18]; applying of a TRVM-grammar to an extended set of possible errors.

2 Metamodel

When developing the complex automated systems based on Scrum approaches [6], LSPS [5], or other approaches the strategic conformity (completeness, adequacy and consistency) of the project is lost due to the large amount of information. To make the project meeting the requirements of customers and being successful the developer is proposed a metamodel which gives a general idea about the design organization with description of the actors (entities) and their roles, functions and relations with each other.

Development of project process always has stakeholders, and, one or several teams of developers (designers) working together. The number of teams depends on the complexity of the project (size of project). Moreover, each team has a project manager, an executive manager and 5–6 members of a team.

The executive manager is responsible for project compliance with the requirements of the customer. He tells the requirements in workflows (backlog), as well as sets the sequence of process execution and assesses the progress of implementation. He is also responsible for the payback of the project. He should always be available to the design teams when solving the problems, answering the questions during the Sprint (Sprint).

The executive manager (responsible officer) is responsible for project compliance with the requirements of the customer. He tells the requirements in workflows (backlog), as well as sets the sequence of process execution and assesses the

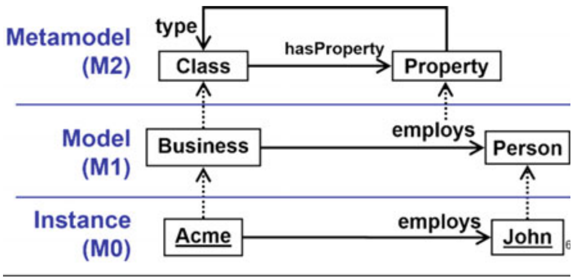


Fig. 1 A structure scheme of a metamodel [18]

progress of implementation. He is also responsible for the payback of the project. He should always be available to the teams when solving the problems, answering the questions during the Sprint (Sprint).

The project manager is responsible for adhering to agile methodology, approaches, the principles of development of complex computer-aided software (systems), and also makes clear the understanding of ambiguous statements. Team members are performers of tasks.

According to the Ref. [18] a metamodel can generate models, and parts of workflows in the development of complex automated systems, have an abstract meaning and be uniform.

The example is Fig. 1. It shows three levels of the workflow: the metamodel—Metamodel (M2), the model of workflow Model (M1) and the instance of the workflow—Instance (M0), and abstractions: Class (Class), Property (Property), Business (Business), Person (Person), Acme (AFM), John (Name) and relations between abstractions: type (type), hasProperty (field), employs (employees). The solid line defines the relationship of "contains" and dotted line means "inherits", the arrows indicate the order of the elements. This diagram provides abstract entities as the basis for consistent interpretation of specialized entities used different methods of modeling of workflow. The advantage of using abstractions (e.g., Fig. 1, Class, Property, type, hasProperty) is the polymorphism of methods and properties of the metamodel, the ability to implement access via a unified application programming interface (API).

We create our metamodel that is included in our grammar to increase a quality of checking of diagrams.

3 Grammars

In the modern theory of graphic and visual languages for representation of diagrammatical models including streams of works, two main syntactic models which basis attributes of graphic objects and types of their binding are considered: spatial and logical. The spatial model contains relative or absolute coordinates of graphic objects. Use of such model for control and the analysis structural, i.e. topological (syntactic), features of diagrams extremely difficult. Therefore, for the description of syntax of graphic languages of diagrammatical models the logical model is used. However, the existing logical models of grammars [19–23] have the following shortcomings:

1. Position grammars, developing on the basis of plex-structures, do not assume use of areas of connection and cannot be applied to graphic languages which objects have dynamically changeable quantity of inputs and outputs.
2. Authors of relational grammars make a reservation about imperfection of the mechanism of neutralization in respect of incompleteness of the formed list of mistakes.
3. There is no semantic control of text attributes of diagrammatical models of graphic languages.
4. Common faults of the above described grammars are: increase in number of production at creation of grammar for unstructured graphic languages (at invariable quantity of primitives of graphic language for the description of all options of not structure there is a significant increase in quantity of production), complexity of creation of grammar, big time expenditure of the analysis (the analyzers constructed on the basis of the considered grammars possess polynomial or exponential time of the analysis of charts of graphic languages).

In the field of a diagram and processing of visual languages are practically not conducted by the Russian scientists of research. Work of the prof. Zhogolev E. A. is known [19], which has survey character. In its importance of the direction connected with computer graphics is emphasized. However, the perspective connected with the analysis and control of diagrammatical models is not considered. In Stepanov P. A. work [20] the similar questions are also not raised.

4 A Temporal Grammar

When collective designing diagrammatical models have a complex hierarchical structure (complex diagrams), and they increase the number of terms in many times [21–23]. Classical graphical RV-grammar becomes more complicated, the process of its development becomes more sophisticated. The control of interconnected nodes and diagrams of complex models is not provided. To exclude these shortcomings, the multi-level grammar is proposed. Consider system of the temporal grammar given as a sequence of four elements:

$$TRVM = (\text{temporal}, \langle n, \Sigma^n, RV^n, r_0 \rangle, \textit{metamodel}) \tag{1}$$

where $\text{temporal} = \{\text{UML}, \text{BPMN}, \text{ARIS}, \ldots\}$ is a type of the temporal grammar;
n is a grammar index;
Σ^n is an alphabet of the n-th grammar;

RV^n is a set of the n-th grammar rules;
r_0 is an axiom of upper-level grammar;
metaModel is the metamodel.

As one of the states RV^i grammar contains RV^j grammar. RV^j grammar can also be compound:

$$G = (V, \Sigma, \tilde{\Sigma}, R, r_0) \tag{2}$$

where RV^n—the L(R) language grammar is an ordered set of five nonempty sets, where $V = \{v_l, l = \overline{1, L}\}$ is an alphabet of operations over the internal memory; $\Sigma = \{a_t, t = \overline{1, T}\}$ is a terminal alphabet of the graphical language (the set of the graphical language primitives); $\tilde{\Sigma} = \{\tilde{a}_t, t = \overline{1, \tilde{T}}\}$ is a quasi-terminal alphabet extending the terminal alphabet. The alphabet includes:

- quasi-terms of graphic objects that won't continue analysis,
- quasi-terms of graphic objects with more than one input,
- quasi-terms of links—marks with specific semantic differences defined for them;
- quasi-term for completing the analysis;
- $R = \{r_i, i = \overline{1, I}\}$ is the scheme of the grammar G (a set of names of complexes of rules, where each complex r_i consists of a subset P_{ij} of rules $r_i = \{P_{ij}, j = \overline{1, J}\}$);
- $r_0 \in R$ is an axiom of RV-grammar (the name of the initial complex of rules), $r_k \in R$ is a final complex of rules.

The rules of $P_{ij} \in r_i$ are given as:

$$\tilde{a}_t \xrightarrow{W\gamma(\gamma_1, \ldots, \gamma_n)} r_m \tag{3}$$

where $W_\gamma(\gamma_1, \ldots, \gamma_n)$—n-ary relation defining the form of operation on the internal memory depending on $\gamma \in \{0, 1, 2, 3\}$; $r_m \in R$ is the name of the receiver-rules.

The internal memory consists of stacks for processing the graphic objects that have more than one output, and elastic tapes for processing the graphic objects that have more than one input. The RVM-system receives the symbols of terminal alphabet from input tape and transmits them to the appropriate level. The elements that transmit the grammar to another level are called "subterms". Then the RV^n-grammar description is given as:

$$G = (V, \Sigma, \tilde{\Sigma}, \bar{\Sigma}, R, r_0) \tag{4}$$

where $\bar{\Sigma}$ is a set of subterms, i.e. grammar elements that transmit the automat to the next lower level. The rules containing subterms are the following:

$$\bar{a}_t \xrightarrow[r_0^{n+1}]{W_\gamma(\gamma_1, \ldots, \gamma_n)} r_m^n \tag{5}$$

where r_0^{n+1} is a receiver-rules the initial complex of next level grammar, r_m^n is a receiver-rules to which the transition is done when r_k^{n+1}.

As a simple example, we depict a use-case diagram in Fig. 2 that consists of an actor (Operator), an entity (DataBase) and operations (Update Record, Delete Record, Add Record, Select Record, Run Trigger, Call Function) to edit records in a table of a database. So, we create a grammar to describe use-case diagrams and present the description in Table 1. This mathematical description (see Table 1) helps to check and fix possible errors in our example.

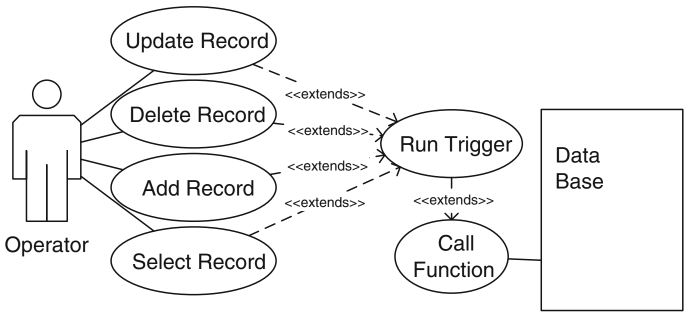


Fig. 2 A use-case diagram to describe a process to edit record in a database

Table 1 The tabular form of a grammar of use-case diagrams

№	Complex	Quaziterm	The complex-the successor	RV-relation
1	r_0	R_I	r_3	$W_1(i^{t(1)})/W_3(i^{t(1)} \equiv \varnothing \cup i^{t(2)} \equiv \varnothing)$
2		R_E	r_4	$W_1(i^{t(1)})/W_3(i^{t(1)} \equiv \varnothing \cup i^{t(2)} \equiv \varnothing)$
3		R_G	r_5	$W_1(i^{t(1)})/W_3(i^{t(1)} \equiv \varnothing \cup i^{t(2)} \equiv \varnothing)$
4		$\underline{R_I}$	r_6	$\varnothing$
5		$\underline{R_E}$	r_7	$\varnothing$
6		$\underline{R_G}$	r_8	$\varnothing$
7	r_1	R_A	r_2	$\varnothing$
8		R_G	r_5	$\varnothing$
9	r_2	C	r_0	$\varnothing$
10	r_3	C	r_0	$\varnothing$
11	r_4	C	r_0	$\varnothing$
12	r_5	C	r_0	$\varnothing$
13		A	r_1	$\varnothing$
14	r_6	C	r_0	$W_1(i^{t(2)})/W_3(i^{t(1)} \equiv \varnothing \cup i^{t(2)} \equiv \varnothing)$
15	r_7	C	r_0	$W_1(i^{t(2)})/W_3(i^{t(1)} \equiv \varnothing \cup i^{t(2)} \equiv \varnothing)$
16	r_8	C	r_0	$W_1(i^{t(2)})/W_3(i^{t(1)} \equiv \varnothing \cup i^{t(2)} \equiv \varnothing)$
17		A	r_1	$\varnothing$

5 Checking Diagrams

When designed complex computer based systems, we have distributed diagrams (complex diagrams) that must be check for errors with the help an analyzer. In our system, the semantic information of a diagram stored in the database as a graph of semantic concepts (a textual information), loaded to the graphical objects and links of diagrams. A new diagram is analyzed for concepts. To control the complex diagram, you need to construct a graph of semantic links between items of a diagram using lexical and syntax templates [22]. The proposed TRVM-grammar allows designers to check the following semantic errors: Large synonyms, Object's Antonyms, Conversion of links. One class of errors, called Incompatible objects, is not controlled with the proposed method. The developed TRVM-grammar of analysis and control can check and fix errors of a separate context as well as semantic errors that are not defined in most modern tools. Checking detects the following types of errors that occur in the UML diagrams (Table 2). The Table 2 uses the following abbreviations: UD is the Use of Diagram; AD is an Activity Diagram; SD is a Sequence Diagram, CD is a Class Diagram, DD is a Deployment Diagrams. The research of modern systems for creating the UML diagrams represents that the systems allow to find the first 16 types of mentioned errors.

Table 2 The error types of UML diagrams

№	Type of error	UD	AD	SD	CD	DD
1	Lack of link	+	+	+	+	+
2	Error of transfer of control					
3	Error of multiplicity of inputs		+			+
4	Error of multiplicity of outputs		+			
5	Invalid link	+	+	+	+	+
6	Error of link	+	+			+
7	Error of access level				+	
8	Error of message transfer		+	+		
9	Error of delegation of management				+	
10	Quantitative error of diagram elements		+			+
11	Excluding links of the wrong type				+	
12	Call to the lifeline			+		
13	Unbounded links	+	+	+	+	+
14	Violation of multiplicity of dependencies	+	+			
15	Mutually exclusive relations	+				
16	Multiple links	+				
17	Infinite cycle	+				
18	Circular links	+				
19	Synchronous call until get answer			+		
20	Error of a separate context		+			

The author's approach of TRVM-grammar allows to find all types of errors, including 4 additional: Multiple links, Circular links, Synchronous call until get answer, and Error of a separate context. The error of a separate context is an error that occurs in the paired elements as conditional, splitting and merging of conditional branches. This error is characterized with an unlimited number of units and links. For example, in the activity diagram, conditional links suppose the execution only one of the possible link, but transfer controls that all input links must be completed. This situation never allows to complete the active diagram successfully. To analysis and control the errors the program agent has developed: syntax oriented analyzer of UML diagrams for MS Visio, and network system of diagrams analysis and control, offering a full set of a functional to analyze and control syntax and semantic errors. The temporal grammar allows to perform follows: analyses any diagrams (BPML, UML, IDEF0, IDEF3, Petri net) using RVM-grammar; be integrated as plug-ins into MS Visio; 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.

6 Conclusions

In the article the industrial branch in which there is an interest in applied control devices and diagnostics of flows of project works in the form of diagrams is considered. Authors formulated 20 types of the mistakes arising in BPML, UML, IDEF0, IDEF3, Petri's network. Modern means can check only 16 types of mistakes, author's development of TRVM grammar fixes and corrects all 20 types of mistakes, has linear dependence of expenses of time for the analysis and correction in comparison with other types of the grammars reckoning exponential and polynomial dependences of expenses of time for the analysis, and also parallel work in all listed types of diagrams. The metamodel also gives the big picture or strategic fit about the design organization.

References

1. Becker, J., Kugeler, M., Rosemann, M., (eds.): Process Management: A Guide for the Design of Business Processes, 2nd edn. Springer (2011)
2. Weske, M.: Business Process Management: Concepts, Languages, Architectures. 2nd edn. Springer (2012)
3. Sherehiy, B., Karwowski, W., Layer, J.K.: A review of enterprise agility: concepts, frameworks, and attributes. Int. J. Ind. Ergon. **37**, 445–460 (2007)
4. Highsmith, J., Orr, K., Cockburn, A.: E-Business Application Delivery, pp. 4–17 (2000). [Online] Available at: http://www.cutter.com/freestuff/ead0002.pdf
5. A global Swiss company offering advanced intelligent application software for multiple business sectors. [Online] Available at: http://whitestein.com/. Accessed 05 Feb 2016

6. Bider, I., Jalali, A.: Agile business process development: why, how and when applying Nonaka's theory of knowledge transformation to business process development. In: Information Systems and e-Business Management. Springer, Berlin (2014). [Online] Available at: https://www.researchgate.net/publication/266078141. Accessed 05 Feb 2016
7. Andersson, T., Andersson-Ceder, A., Bider, I.: State flow as a way of analyzing business processes-case studies. Logist. Inf. Manag. **15**(1), 34–45 (2002)
8. YAWL Foundation: YAWL (2004). [Online] Available at: http://www.yawlfoundation.org/. Accessed 05 Feb 2016
9. Bider, I.: Analysis of agile software development from the knowledge transformation perspective. In: Johansson, B. (ed.) To appear in 13th International Conference on Perspectives in Business Informatics Research (BIR 2014). Springer, LNBIP, Lund, Sweden (2014)
10. IbisSoft: iPB Reference Manual (2009). [Online] Available at: http://docs.ibissoft.se/node/3. Accessed 05 Feb 2016
11. Jalali, A., Wohed, P., Ouyang, C., Johannesson, P.: Dynamic weaving in aspect oriented business process management. In: 21st International Conference on COOPERATIVE INFORMATION SYSTEMS (CoopIS 2013), pp. 2–20. Springer (2013)
12. Business Process Technology Group. [Online] Available at: http://bpt.hpi.uni-potsdam.de. Accessed 05 Feb 2016
13. Fu, K.: Structural Methods of Pattern Recognition, p. 319. Mir, Moscow (1977)
14. Costagliola, G., Lucia, A.D., Orece, S., Tortora, G.: A parsing methodology for the implementation of visual systems. [Online] Available at: http://www.dmi.unisa.it/people/costagliola/www/home/papers/method.ps.gz. Accessed 05 Feb 2016
15. Wittenburg, K., Weitzman. L.: Relational grammars: theory and practice in a visual language interface for process modeling (1996). [Online] Available at: http://citeseer.ist.psu.edu/wittenburg96relational.html. Accessed 05 Feb 2016
16. 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)
17. Zhang, K.B., Zhang, K., Orgun, M.A.: Using Graph Grammar to Implement Global Layout for a Visual Programming Language Generation System (2002)
18. Thiemich, C., Puhlmann, F.: An Agile BPM Project Methodology (2013). [Online] Available at: http://frapu.de/pdf/bpm2013.pdf. Accessed 05 Feb 2016
19. Zhogolev, E.A.: Graphical errors and graphic grammars. Progr. Comput. Softw. № 3, pp. 30–42 (2001)
20. Stepanov, P.A., Okhtilev, M.Yu.: Computational models of a visual language. In: Proceedings of Higher Schools № 11, pp. 28–32 (2006)
21. Sharov, O.G., Afanas'ev, A.N.: Syntax-directed implementation of visual languages based on automaton graphical grammars. Progr. Comput. Softw. № 6, pp. 56–66 (2005)
22. Sharov, O.G., Afanas'ev, A.N.: Neutralization of syntax errors in the graphic languages. Progr. Comput. Softw. № 1, pp. 61–66 (2008)
23. Sharov, O.G., Afanas'ev, A.N.: Methods and tools for translation of graphical diagrams. Prog. Comput. Softw. № 3, pp. 65–76 (2011)

Part IV
Ontological Engineering and Semantic Technologies

Integration and Processing of Problem-Oriented Knowledge Based on Evolutionary Procedures

Victoria Bova, Dmitry Zaporozhets and Vladimir Kureichik

Abstract Nowadays integration and processing of domain-specific knowledge is one of the most important tasks of providing access to heterogeneous information from different subject areas and support for interoperability of intelligent information systems aimed at the sharing of data and knowledge on structural and semantic level. The article discusses a modified approach to the problems of integration and knowledge processing, involves a comparison in their automatic ontology using semantic component metrics. The authors propose an evolutionary approach to the problem of integrating multiple ontologies for interoperability and representation of information and knowledge in intelligent information systems. The objectives of integration and processing of knowledge belong to the class of NP-hard optimization problems and can be implemented using genetic algorithms search for optimal solutions. The authors propose a genetic algorithm that is based on the use of analogues to the evolutionary processes of reproduction, crossover, mutation and natural selection. The researchers have conducted a series of experiments to analyze the developed approach. The findings confirmed the theoretical significance and the prospect of this approach, as well as possible to establish the optimal parameters of the algorithm.

Keywords Intelligent information systems · Integration of data and problem-oriented knowledge · Ontology · Semantic model · Genetic algorithms · Genetic operators

V. Bova · D. Zaporozhets · V. Kureichik (✉)
Southern Federal University, Rostov-on-Don, Russia
e-mail: vkur@sfedu.ru

V. Bova
e-mail: vvbova@yandex.ru

D. Zaporozhets
e-mail: elpilasgsm@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 450, DOI 10.1007/978-3-319-33609-1_21

1 Introduction

The development of new approaches and methods of performance, integration and treatment of problem-oriented knowledge is the main direction of development of modern intelligent information systems (IIS) [1–3]. The structure of the IMS includes diverse knowledge base with their own local information models with different standards of description, as well data and knowledge presentation. When you merge them into a global model, it generates many conflicts: the use of different terminology with referring to similar concepts in the domain of IMS; heterogeneity at the level of model specifications and conceptual semantics; identification and conversion of non-uniform data structures and knowledge [4–6].

All this makes the problem of integration in problems of processing domain-specific knowledge rather complex and multi-level. To resolve this problem we should take into account both the structural and syntactical differences in data models and knowledge that generate a schematic heterogeneity, and semantic properties of data objects to ensure semantic interoperability of data and resolution of semantic conflicts. For this reason, the integration of ontologies for the establishment of subsequent interaction information IIS models is an important task.

The paper proposes an evolutionary approach to the problem of integrating multiple ontologies for compatibility and representation data and knowledge in IIS. Such an approach would identify the priority projects semantic data and knowledge to represent them in the model of integration as well as to eliminate duplication and contradictions of entities and relationships at the level of the domain and data objects from the area of integration.

The objectives of integration and knowledge processing belongs to the class of NP-hard optimization problems and can be implemented with genetic algorithms for finding the optimal solutions.

2 Integration Problems of Knowledge in IIS Ontologies

The integration problem of data and knowledge is characterized by a wide variety of productions tasks, approaches and methods used to solve them [1, 3–5]. In general, the problem of integration is the logical association of data belonging to different sources, which provides a unified view of these data and operation.

Paying attention to the heterogeneity of the data, we should clarify this concept. It is not heterogeneous in terms of physical storage (i.e., regardless of their location and method of storage), but in terms of a model of representation—ontological specifications.

The heterogeneity of ontological specifications appears on the level of conceptual modeling and semantics. Differences that create heterogeneity at the level of factor models [1, 5–8]: in the syntax of the language defining the ontological model; in the expressive capacity models; the semantics of the primitives used in the models. Heterogeneity creates differences on the semantic level [5]: in the names of concepts and relationships; in approaches to the definition of concepts; in the partition on the domain concepts; in covering the subject area; in the points of view on the subject area. Accordingly, we have a problem of coordination for ontological specifications.

Today there are three main components of the problem of integration of data and knowledge: development of integrated circuits, providing a unified view of the data of different sources based on a unified ontological model; development of mappings between ontological models; development of methods of manipulation, the essence of which is disclosed below.

To solve the problem of semantic heterogeneity of data and knowledge, as well an access to heterogeneous information from different subject areas we propose a modified approach to the construction of the resulting ontologies for multiple source-level matching of concepts, relationships, and attributes. Integration model of heterogeneous IIS data and knowledge reduces to the construction of maps and establish relationships in a unified ontological model based on ontology matching multiple levels of IP model and conceptual semantics. The complexity of the model is selected based on the need for its expressive possibilities and a given set of semantic dependencies.

3 Modified Approach to the Integration Problem

One of the objectives of overcoming the lexical and semantic heterogeneity of IIS data and knowledge is to develop new approaches for sharing and support of ontologies developed independently of each other [2, 6]. We propose the following modified approach for solving the integration problem of IIS data and knowledge, as well providing an access to heterogeneous information from various subject areas. Especially suitable to the construction of the resulting ontology to multiple source is to provide consistency in the structural and semantic levels. The proposed approach to building a model of integration of multiple ontologies does not require replacement of individual single ontology, which is the result of combining. The integration process is understood as a process of establishing a non-uniform mapping of ontologies level of compliance with the possibility to extend the set of operations (methods of manipulation) over them in a semantically significant level. Such an approach would identify the priority projects semantic data and knowledge to represent them in the model of integration as well as to eliminate duplication and contradictions of entities and relationships at the level of the domain and data objects from the area of integration.

We need to solve the following problems for the integration of ontologies for matching application contexts [5–8]:

- binding specifications of information sources with the specifications of the domain to reflect its implicit semantics;
- reduction of different ontologies formalisms to one for comparison of ideas about the subject area;

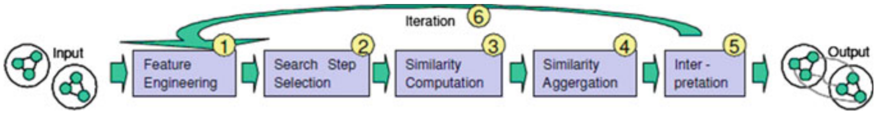


Fig. 1 Process of ontology mapping

- mapping of ontological contexts in one formalism to align them;
- semantic linking of object diagrams elements of information sources and the task based on ontological concepts connection.

The authors propose an algorithm to solve this problem. It includes six operations running consecutively to display the ontology (Fig. 1).

1. Feature Engineering—is relay function of ontologies (ontology analysis of the elements), i.e., converting one format of presentation of the initial ontology, usually result in the format RDF (S), as it is considered standard when working with ontologies.
2. Selection of Text Search Steps is selection of the next step to find a candidate. Choosing an expert semantic proximity search algorithm and semantic distance between pairs of concepts, depending on the goal.
3. Similarity Computation is calculation of similarity, determination of similarity among pairs of ontologies' concepts. It is calculated in step of comparison among ontologies.
4. Similarity Aggregation is aggregation of similarities, i.e., association of entities into one total value confirmation for mapping of connections. Among the pairs of matched entities, we choose one that has a measure of semantic similarity is the highest. Similarity limit is chosen heuristically.
5. Interpretation is formation of the mapping between ontology elements based on similarities, in other words, it is comparison of the names of concepts, assigning the selected class name synthesized from the other two.
6. Iteration is repeating of several steps of the algorithm consists of several stages and stops when it cannot find the new maps.

The proximity and coherence are the main criteria of mapping algorithm of ontologies integral elements.

Evaluation model of semantic proximity for ontologies elements was presented in [6].

4 Evolutionary Approach to the Problem of Knowledge Integration

Hybrid measure takes into account the differences between the compared objects on various grounds and is determined by the dominant value (a measure of Euclid), which allows you to increase the weight of the measures that are important and

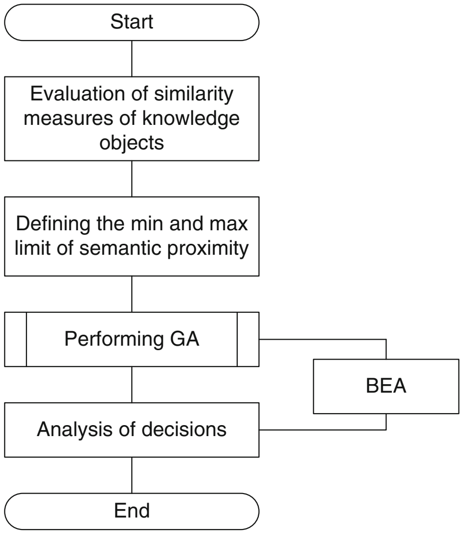


Fig. 2 Generalized architecture of genetic search

virtually ignore measures with small values [9–12]. In this regard, the most promising approach to determine a quantitative measure of proximity is the automatic determination of the weighting coefficients using GA. Generalized structure of the genetic search of weighting coefficients shown in Fig. 2.

The process of genetic research is a sequence of transformations of a finite set of alternatives to another using the mechanisms and principles of genetics and evolution of wildlife [13, 14].

The weighting factors *t*, *r*, *a* allow you to adjust the process of calculating of the semantic proximity of the two concepts. According to the formula

$$\mathrm{C}(\mathrm{k}_i, \mathrm{k}_j) = (\mathrm{t} \cdot \mathrm{C}^{T^{\mathrm{ax}}}(\mathrm{k}_i, \mathrm{k}_j) + \mathrm{r} \cdot \mathrm{C}^{R^{\mathrm{el}}}(\mathrm{k}_i, \mathrm{k}_j) + \mathrm{a} \cdot \mathrm{C}^{A^{\mathrm{tr}}}(\mathrm{k}_i, \mathrm{k}_j)) \quad (1)$$

the evaluating task of concepts semantic proximity of ontology has some limitations: $minf_{t,r,a}(\bar{x}), \bar{x} = (t, r, a) \in F \subseteq S;\ t, r, a \in [0; 1];\ t + r + a = 1$ where $\bar{x}$ is an admissible solution, *F* is the range of permissible values, and *S* is the search area.

The objective function (OF) is based on the search for Euclidean distance and has the form:

$$f_{t,r,a} = \sum_{k_i \in ONT, k_j \in ONT'} \left(t * C^{Tax}(k_i, k_j) + r * C^{Rel}(k_i, k_j) + a * C^{Atr}(k_i, k_j) - 1\right)^2 \quad (2)$$

Thus, the search for the weighting factors is to perform four steps.

Evaluation of the similarity of data and knowledge integral ontologies produced in the first stage. The method of calculating the semantic proximity of concepts allows quantifying the similarity between the concepts [6, 14]. We define a values measure limit of proximity to rank the elements of the result set (Fig. 3).

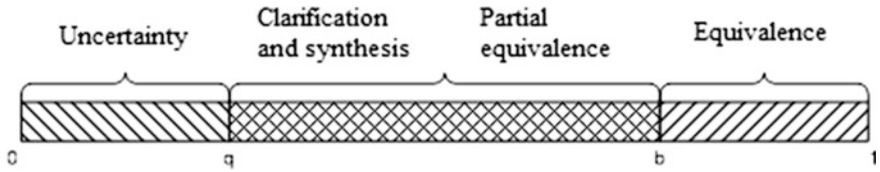


Fig. 3 Levels of threshold values for semantic proximity measures

The parameter b is the minimum limit, which determines the value where it is possible to map the full ontology. The limit where the concepts are accepted in part equivalent is calculated as follows:

$$q = \max(\mathrm{C}\ (\mathrm{k}_i, \mathrm{k}_j)\ |\forall \mathrm{k}_i \in \mathrm{ONT}, \forall \mathrm{k}_j \in \mathrm{ONT}')\ \times p_2/100, \qquad (3)$$

where p_2 is the percentage where q is taken as similarity limit for establishing a partial equivalence concepts.

The concepts are different in the case when the measure value of semantic proximity is less than the threshold q.

Further GA runs to find the weighting limit and compare with limit values. Block of evolutionary adaptation (BEA) has been introduced to implement feedback search architecture, which is based on interaction with the external environment (the decision-maker) manages the process of finding and configuring the GA [13]. It affects the rearrangement of the current population of alternative solutions and the creation of a new population.

5 GA of Weighting Coefficients Search

The block diagram of GA is shown in Fig. 4. The first stage is the input weighting factors $w_i\{t_i, r_i, a_i\}$, which determine the importance of proximity measures $C^{T^{ax}}(k_i, k_j)$, $C^{R^{el}}(k_i, k_j)$, $C^{A^{tr}}(k_i, k_j)$.

Further, the initial population of alternative solutions P is generated with $t, r, a \in [0; 1]$, $t + r + a = 1$. Each alternative corresponds to the chromosome, which is a $w_i = \langle t_i, r_i, a_i \rangle$, n is the number of chromosomes in the population.

After the step of generation we calculate OF's value for each chromosome according to the formula (2). Selecting pairs of chromosomes for subsequent crossing is made based on the data obtained. Such a choice is made according to the principle of natural selection, where the greatest chance to participate in the creation of new species belongs to chromosomes with the highest values of OF.

The method of the roulette wheel has been chosen for the implementation of operator selection [10]. In spite of the random nature of this procedure, the parent individuals are selected in proportion to the values of OF: each chromosome is

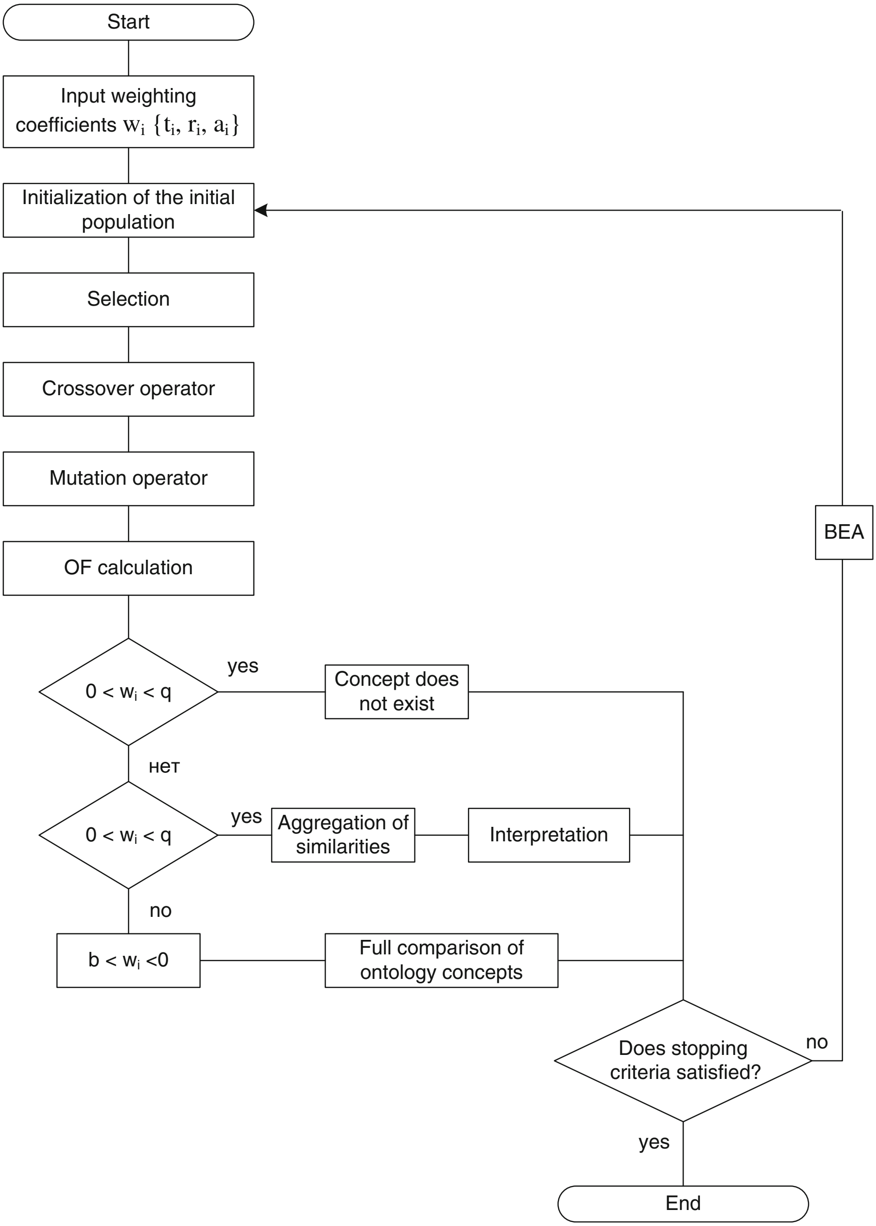


Fig. 4 GA block diagram

associated sector of the roulette wheel, and its value shall be proportional to the value of OF given chromosome, therefore, the greater the value OF, the more sectors on the roulette wheel.

The next step of GA is to perform crossover operator, where its main task is to provide ultimately the most functional features that were present in the set of source solutions.

Ordered crossover operator implemented as a part of the solved problem. Ordered crossover is implemented in transformation stages of genetic material and enables the only real solution.

Break point is selected randomly. Next we make up the left segment of the parent chromosomes P1 in the left segment of the chromosome, the descendant of P1′. The remaining genes P1′ is taken from the second parent chromosome P2 left in an orderly manner. The second descendant P2′ formed in a similar manner.

Break point is selected randomly. Next we make up the left segment of the parent chromosomes P1 in the left segment of the chromosome, the descendant of P1′. The remaining genes P1′ is taken from the second parent chromosome P2 left in an orderly manner. The second descendant P2′ formed in a similar manner.

Gene mutations have the greatest importance in the solutions of the current task. Mutations lead to the appearance of qualitatively new properties of the genetic material. The essence of the developed mutation operator consists of the following. We randomly select a random number in the studied chromosome. Factor mutation determines the intensity of mutations. It determines the fraction of genes undergoing mutations at the current iteration, based on their total amount. If the mutation rate is too small, we get the situation in which a plurality of useful genes simply will not exist in the population. The magnitude selected according to the value after the change of *i*-gene it was in the interval [0; 1].

Using this strategy, the implementation of the operator increases the search space, which is a prerequisite for finding the optimal solution [14].

OF calculation results leads to different operations with ontology concepts presented in the algorithm (Fig. 4).

In this algorithm as a stopping criterion is proposed to use a certain number of iterations. As long as the stopping criterion has not been reached, evolutionary adaptation is performed in the transition to the next iteration.

6 Experimental Results

The problem of integration of ontologies as well as mapping of the ontological concepts involves finding commonalities and differences in the specifications of multiple ontologies for future interoperability and representation of data and problem-oriented knowledge in IIS [5–8, 12].

Ontologies of integral IIS initially have nothing to do; therefore, we need to find semantically similar elements to confirm the correctness of ontologies and semantic relationships established between ontological concepts. Therefore, the purpose of the analysis of algorithms assess the similarity measures we have considered projects that implement the methods of calculating of the semantic proximity of concepts integral ontologies. To evaluate the developed algorithm authors have

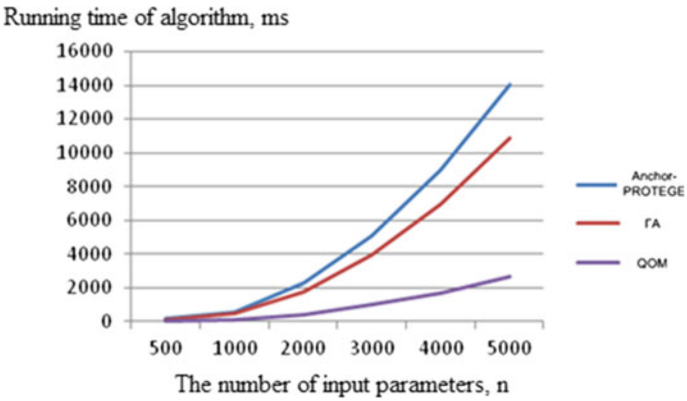


Fig. 5 Graph of dependence of time decisions from the number of input parameters

performed a comparative analysis of such systems using Anchor-PROMPT [5] and QOM [9]. The software supports process of unification and ontology mapping.

The software environment for finding weighting factors has been developed for experimental research work of the proposed GA t, r, a, that allow you to adjust the process of calculating of the semantic proximity of the two concepts. This problem is reduced by solving a system of linear algebraic equations.

The experimental results allowed determining the dependence of the algorithm on the input parameters (weight coefficients): n js the number of chromosomes in the population; chromosome is a tuple $w_i = \langle t_i, r_i, a_i \rangle$.

The graphs of working time of the algorithm, as well as the Anchor-PROMPT and QOM, the number of input data shown in Fig. 5.

The proposed approach presents the original mechanism for mapping and integration of ontologies using a genetic algorithm to determine the proximity of ontology elements according to (weights) copies of concepts. Ontologies are defined as taxonomic concepts with attributes. During the search process, the rules GA matching items ontologies. Precision performance depends on the quality obtained at each iteration GA-effective solutions, weighing the results of determination of similarity of concepts of integral ontologies.

The time complexity of the algorithm is approximately $O(n^2)$.

7 Conclusion

The main advantages of an evolutionary approach to solving the problems of integration and processing of domain-specific knowledge are to identify the key concepts for the construction of the resulting ontology, as well as elimination of subjective descriptions of concepts ontology and depending on perspective development of ontologies. To solve the problem of semantic conflicts we have proposed the assessment model of semantic proximity based on harmonization of the attribute, the taxonomic and relational similarity measure. Authors have developed GA of determining similarity criterion for the classification of concepts maps in the following groups: equivalence of partial equivalence, summarizing, clarification, uncertainty. The proposed approach allows finding an effective solution to the problem of data integration and knowledge using modified genetic operators and the process of evolutionary adaptation. A distinctive feature of the approach is the automatic calculation of weight coefficients using GA.

Acknowledgment The study was performed by the grant from the Russian Science Foundation (project # 14-11-00242) in the Southern Federal University.

References

1. Bova, V.V., Kureichik, V.V., Legebokov, A.A.: The integrated model of representation model of representation oriented knowledge in information systems In: 8th IEEE International Conference In: Application of Information and Communication Technologies—AICT 2014, pp. 111–115. IEEE Press, Astana, Kazakhstan (2014)
2. Kravchenko, Y.A., Kureichik, V.V.: Knowledge management based on multi-agent simulation in informational systems. In: 8th IEEE International Conference “Application of Information and Communication Technologies—AICT 2014”, pp. 264–267. IEEE Press, Astana, Kazakhstan (2014)
3. Bova, V.V., Kravchenko, Y.A., Kureichik, V.V.: Decision support systems for knowledge management. In: Advances in Intelligent Systems and Computing, pp. 123–130. Czech Republic (2015)
4. Rodzin, S.I., Podzina, L.S.: Mobile learning systems and ontology. In: Advances in Intelligent Systems and Computing, pp. 45–54. Czech Republic (2015)
5. Skvortsov, N.A.: Questions harmonize heterogeneous ontological models and ontological context. Ontol. Model. J. 149–166 (2008)
6. Bova, V., Kureichik, V., Zaruba, D.: Data and knowledge classification in intelligence informational systems by the evolutionary method. In: Proceedings of the 6th International Conference Cloud System and Big Data Engineering (Confluence’2016), pp. 6–11. India (2015)
7. Bova, V.V., Kravchenko, Y.A., Kureichik, V.V.: Development of distributed information systems: ontological approach. In: Advances in Intelligent Systems and Computing, pp. 113–122. Czech Republic (2015)
8. Gavrilova, T.A.: The ontological approach to knowledge management in the development of corporate information systems. News Artif. Intell. J. **1**, 24–30 (2003)
9. Tuzovskiy, A.F.: Working with ontologies in the knowledge management system the organization. In: Abstracts of the Second International Conference on Cognitive Science CogSci-2006, pp. 581–583. SPb: SPbGU (2006)
10. Zaporozhets, D.Y., Zaruba, D.V., Kureichik, V.V.: Hybrid bionic algorithms for solving problems of parametric optimization. J. World Appl. Sci. J. **23**, 1032–1036 (2013)
11. Rodzin, S., Rodzina, L.: Theory of bioinspired search for optimal solutions and its application for the processing of problem-oriented knowledge. In: 8th IEEE International Conference “Application of Information and Communication Technologies—AICT 2014”, pp. 142–147. IEEE Press, Astana, Kazakhstan (2014)
12. Bova, V.V., Kureichik, V.V., Legebokov, A.A.: Integration of ontologies in scope of model and conceptual semantics: modified approach. In: 9th IEEE International Conference

"Application of Information and Communication Technologies—AICT 2015", pp. 111–115. IEEE Press, Rostov-on-Don, Russia (2015)
13. Gladkov, L.A., Gladkova, N.V., Legebokov, A.A.: Organization of knowledge management based on hybrid intelligent methods. In: Advances in Intelligent Systems and Computing, pp. 107–112. Czech Republic (2015)
14. Bova, V.V., Legebokov, A.A., Gladkov, L.A.: Problem-oriented algorithms of solutions search based on the methods of swarm intelligence. J. World Appl. Sci. J. **27**, 1201–1205 (2013)

Design Intelligent Lifecycle Management Systems Based on Applying of Semantic Technologies

Alena V. Fedotova, Irina T. Davydenko and Anne Pförtner

Abstract Design of management systems is an actual task. Nowadays design methods of management systems in the context of complex technical systems lifecycle management are underdeveloped. The article considers the approach of construction of intelligent lifecycle management systems by Open Semantic Technologies for Intelligent Systems (OSTIS). The knowledge base of the system is based on a system of domain ontologies, supporting management of design, as well as further development of complex technical system. In this paper, the problem of making intelligent of product lifecycle management systems is considered in the context of the development lifecycle management as a knowledge management system, circulating at all stages of lifecycle. It is the particular novelty of the paper. The paper presents a methodology for designing component OSTIS, an architecture intelligent lifecycle management system, considered a fragment of lifecycle domain ontology and domain of project management by languages SCn, SCg.

Keywords Intelligent system · Semantic technology · Ontology · Ontological modeling · Knowledge management · System lifecycle · PLM-system

---

This work was supported by the BRFFR (The Belarussian Republican Foundation for Fundamental Research) and RFBR (The Russian Foundation for Basic Research) “Methods and Tools of Ontological Modeling for Semantic Technologies of Intelligent Systems Designing” (F15RM-074, 15-57-04047).

A.V. Fedotova (✉)
Bauman Moscow State Technical University (BMSTU), 2nd Baumanskaya St., Bldg. 5, Block 1, 105005 Moscow, Russian Federation
e-mail: afedotova.bmstu@gmail.com; aluon@mail.ru

I.T. Davydenko
Belarusian State University of Informatics and Radioelectronics (BSUIR), P.Brovki St., Bldg. 6, 220013 Minsk, Belarus
e-mail: ir.davydenko@gmail.com

A. Pförtner
Technische Universität Berlin (TUB), Straße des 17. Juni 135, 10623 Berlin, Germany
e-mail: anne.pfoertner@tu-berlin.de

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 450, DOI 10.1007/978-3-319-33609-1_22

1 Introduction

Today complication of production, processes, value of development networks and information technologies (IT), which are the main body of the systems of product lifecycle management, is increasing steadily. Management of this information becomes more and more problematic. How does it influence on the engineers? More than 1400 engineers of German industry gave the feedback with alarming but expected results in 2014. Almost two thirds of respondents spend on the average on the main tasks such as development, design or validation only 20 % of their time. The research confirmed the absence of time for creative engineering work. Is it possible to release engineers and designers from routine and administrative Product Lifecycle Management (PLM) tasks using present-day intelligent technologies? Experts explored the demand for intelligent Product Lifecycle Management systems [1] in constant dialogue with PLM industry. This demand shows the necessity of development of the new generation of intelligent PLM solutions.

Product Lifecycle Management is an integrated approach to manage of the information, related to the product lifecycle, based on the right set of organization preconditions, processes and IT-solutions [2].

Up-to-date tendencies of construction of PLM system are their integration and making them intelligent [3, 4]. One of the approaches to development of intelligent systems of complex product lifecycle management is the using of Open Semantic Technologies for Intelligent Systems (OSTIS) [5, 6].

The main goal of the work is design of the intelligent lifecycle management systems with applying of semantic technologies, what increase using system efficiency, its continues development (improvement) when it is used and essentially increase the term of its obsolescence.

2 Methods of Component Design OSTIS

The authors of OSTIS project suggest the methods of component design, which are based on continuous expansion of reusable components libraries (standard technical solutions). The development of component design technology includes:

- providing of compatibility of intelligent system components on the base of their common ontological conception;
- ontology development of intelligent system component design [7, 8];
- creation of the libraries of intelligent systems standard components and more precise definition of such components (upper level ontologies, domain ontologies, tasks and applications ontologies, usable fragments of knowledge bases, the output machines, interface components, etc.);
- expansion of capabilities of ontological modeling languages in favor of complex knowledge representation, development of computer support solutions of intelligent system synthesis from available components.

Every intelligent system, which is designed on the base of OSTIS technology, will be interpreted as the result of integration of the following subsystems [9]:

- a domain (basic) intelligent system;
- a subsystem of adapted management of the dialog with the user;
- an intelligent help-system for information service and training the domain intelligent system users that should not have high qualification when they begin to work with the system;
- an intelligent lifecycle management system of the domain intelligent system, which coordinate the activity of the domain intelligent system implementers and processes of its perfection;
- an intelligent management system of information security of domain intelligent system.

A general architecture of the integrated intelligent system, which is designed on the base of the OSTIS technology, is presented in Fig. 1.

The enumerated systems must be designed on the base of one and the same technology to ensure their interoperability (semantic compatibility).

In this way, when every intelligent system is designed, it is necessary to design at the same time:

- a subsystem, which provide the information service and training of the ultimate users of this intelligent system, i.e. which is the execution of manual paper for this system in the form of intelligent inquiry and training system. It will help to broaden the ultimate user contingent, to increase the efficiency of the using system and streamline the exploration;
- a subsystem, which coordinate the implementers of the designed intelligent system, so long as system development (improvement) continues when it is used, and needs development of special methods and computer solutions of continuous domain intelligent system improvement in using [10]. It essentially increase the term of its obsolescence;

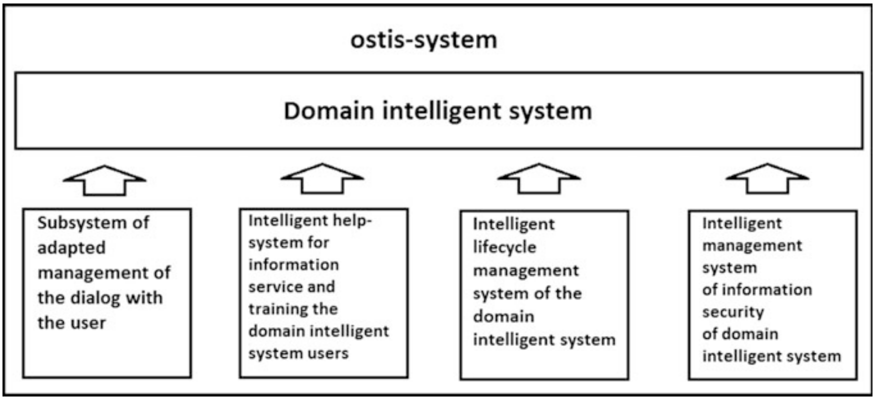


Fig. 1 General architecture of the integrated intelligent system, which is designed on the base of the OSTIS technology

- a subsystem, which manage the information security of designed intelligent system.

The subsystem of design management of the intelligent system have to be developed as the intelligent system integrated with the main (domain) intelligent system. Then eventually it can be not only the coordinator of developer activities, but also the independent design solution that can test, diagnose and analyze both the main design intelligent system and itself [9].

In this study, the problem of making intelligent of complex technical system lifecycle management is considered in the context of design of the management system of knowledge, which circulates at every lifecycle phase [11, 12].

3 The Structure of the Lifecycle Management System of Complex Technical System

As any system, built on the OSTIS technology, the architecture of intelligent lifecycle management system has the following structure:

- *system knowledge base*—includes the system of ontologies, describing the domain area. These ontologies are needed to organize the process of system design and improvement (Domain Ontology: lifecycles of systems, activities, maintenance, project management, etc.);
- *knowledge processing machine*—includes the agents that provide support for the performers' activity for the technical system development, design tasks management and terms of their execution, etc.;
- *system user interface*—represents the subsystem, that provide interconnection of the user and the product lifecycle management system.

Knowledge base is an integral set of the knowledge sufficient for the operation of some intelligent system. It is principle at this point that this knowledge is enough so that this intelligent system can cope with the proper class of tasks. Knowledge base can be interpreted as some semantic space in which mentioned intelligent system works. Different kinds of documents, such as documents of different projects, technical systems, domains, can be represented in the knowledge base. The term "knowledge" is both the resource and the managed object in the tasks of knowledge management at the enterprise [13].

Usually the information resources of an enterprise includes well-defined database, different documents, implicit employee knowledge. Knowledge management is a creation of the enterprise consolidated information area and its gradual transformation into the knowledge space [14]. It ensures flexible distributive approach to generating, collecting, representation, replenishment, spreading, actual usage of the enterprise collective knowledge and organization of access to them.

Initial knowledge structuring is the most important aspect of knowledge management throughout complex technical system lifecycle and its phases. Knowledge at the enterprise can be divided into groups according to different criteria, for example: (1) lifecycle stages and phases—designing, industrial, operating knowledge; (2) components of enterprise activity—knowledge about products and services, processes and technologies, equipment and rigging, organizing structures and staff of enterprise; (3) components of surrounding microenvironment of the enterprise—knowledge about customers, suppliers, partners, competitors [15].

Let us consider the fragment of lifecycle domain ontology [16] in the SCn-language.

system lifecycle
∈ *key sc-element'*:
  o
  ∈ *explanation*
  <= *sc-text translation* *:
    o
    ∋ *example':*
      [Lifecycle of technical system is determined by the time interval from the beginning of its creation to the end of its usage; moreover, a birth of system idea is believed to be the beginning of the lifecycle and the system recycling (utilization) – the end]
      ∈ *English*

system lifecycle model
⊃ *standard lifecycle model*
⊃ *software lifecycle model*

standard system lifecycle model
<=standard structure:*
  {
  - *concept*
  - *design*
  - *manufacturing*
  - *usage*
  - *maintenance*
  - *end of the usage*
  }

software lifecycle model
⊃ *waterfall lifecycle model*
⊃ *iterative lifecycle model*
⊃ *spiral lifecycle model*

lifecycle stage «Design»
⊃ *specification*
⊃ *preliminary design*
⊃ *technical design*
⊃ *working design*
⊃ *production planning*

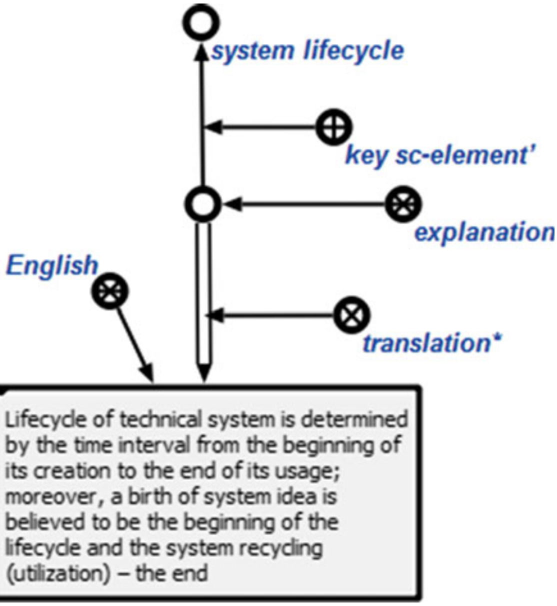


Fig. 2 The fragment of system lifecycle ontology

Let us consider the fragment of lifecycle domain ontology in the SCg-graphical language (Figs. 2 and 3).

Management of design is the fundamental part of lifecycle management system. There are methods of structuring of project works and orientation of management process according to the purpose at the heart of modern methods of design management. The main task of design management is to reach the predetermined project purposes under in advance known implementers, time limits, appropriate use of opportunities and reacting to risks [17].

Usually a project is the scale, technically difficult activity, aimed at the achievement of the appointed purposes. It is necessary to organize the management to obtain more effective and qualitative project result. Project management is a set of knowledge, experience, methods and which that should be applied to project processes to satisfy requirements made to the project, and expectations of project participants. It is necessary to find optimum combination between purposes, terms, costs, quality and other project characteristics to meet the requirements.

It is necessary to understand clearly what every criterion represents and how the interaction inside the project goes on according to these criteria to find the optimum

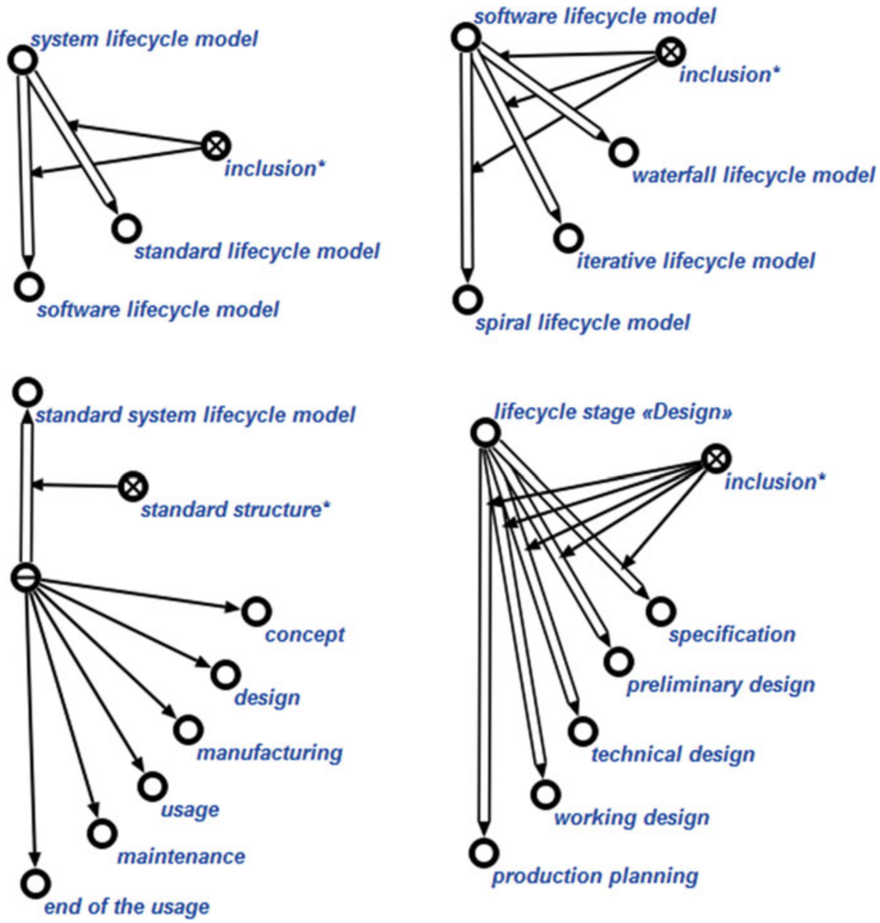


Fig. 3 The fragment of lifecycle domain ontology in the SCg-graphical language

combination between these project criteria. It is necessary to have the subsystem of design management to organize such interaction.

The domain of project management is marked out for the formal describing of knowledge belonging to the design management of intelligent systems. The concepts *project, product*, project task** are the fundamental key concepts of this domain.

project
∈ *key sc-element'*:
o
∈ *explanation*
<= *sc-text translation**:
o
∋ [***project*** is a purposeful activity, the result of which is a certain product that has not existed before]
∈ *English*
∈ *key sc-element'*:
o
∈ *comment*
<= *sc-text translation**:
o
∋ [Every product description (specification) includes:
– indication of product, which is the result of this project;
– indication of project task (project purpose, technical task) – requirements, that developed product must meet;
– the period of project realization;
– the team of project developers (with role indication of every developer);
– the plan of project realization (decomposition on the subprojects for unatomic project, and decomposition on the project tasks for atomic project)]
∈ *English*

product*
= *to be the product of the appropriate activity (including clearly marked, clearly declared project)*
⊃ *basic product**
= *to be basic product of this activity (including this project)*

project task*
= *to be the project task of the appropriate project**
⊃ *basic project task**
= *to be the basic project task of this project**
⊃ *busy project task**
= *to be the urgent project task of this project**

project documents*
∈ *specification*
∈ *key sc-element'*:
o
∈ *explanation*
<= *sc-text translation**:
o
∋ *example'*:
[***project documents**** are the input data for management process, describing the project content]
∈ *English*

Every of developed ontologies represents reusable component of knowledge base included in the Library of knowledge base reusable components [18].

4 Conclusion

The approach to construction of intelligent lifecycle management systems of complex technical system, applying the OSTIS technology is considered in the article.

The knowledge base of such a system is based on the system of domain ontologies, providing support of design management and further improvement of complex technical system. The use of this technology will expand the user contingent, increase the efficiency of system usage and simplify the usage, ensure continuous improvement of the domain intelligent system during usage, that essentially increase the term of its obsolescence.

References

1. Stark, R., Damerau, T., Hayka, H., Neumeyer, S., Woll, R.: Intelligent information technologies to enable next generation PLM. In: The IFIP WG5.1 11th International Conference on Product Lifecycle Management—PLM14. Yokohama, Japan, 7–9th July (2014)
2. Stark, J.: Product Lifecycle Management: 21st Century Paradigm for Product Realization, 2nd edn. Springer, London (2011)
3. Evgenev, G.B.: Intelligent Systems of Design. Bauman MSTU (2012)
4. Evgenev, G.B.: An ontological methodology for developing intelligent systems in mechanical engineering. In: News of Higher Educational Establishment. Engineering, № 6 (651), pp. 39–47. M.: Bauman MSTU (2014)
5. Golenkov, V.V., Gulyakina, N.A.: Principles of construction of mass semantic technology of intelligent system componential design. In: Open Semantic Technologies for Intelligent Systems. Proceedings of International Technological Conference (OSTIS-2011) Minsk, BSUIR, pp. 21–59 (2011)
6. Open Semantic Technologies for Intelligent Systems—OSTIS. http://ims.ostis.net
7. Evgenev, G.B.: An ontological methodology of application intelligent system development. Integrated models and soft computing in artificial intelligence. In: The 8th International Technological Conference, 2 volumes, vol. 1. Fismathlit, pp. 252–258 (2015)
8. Petrochenkov, A.B., Bochkarev, S.V., Ovsyannikov, M.V., Bukhanov, S.A.: Construction of an ontological model of the life cycle of electrotechnical equipment. Russ. Electr. Eng. **86**(6), 320–325 (2015)
9. Golenkov, V.V., Gulyakina, N.A.: Graphodynamical models of parallel knowledge processing: concepts of design, implement and development. In: Open Semantic Technologies for Intelligent Systems (OSTIS-2012): Proceedings of International Technological Conference, Minsk. BSUIR, pp. 23–52 (2012)
10. Kolchin, A.F., Strekalov, A.F., Ovsyannikov, M.V., Sumarokov, S.V.: Product Lifecycle Management, Anaharsis (2002)

11. Tarasov, V.B., Fedotova, A.V., Cherepanov, N.V.: Lifecycle ontologies of complex technical system. In: Open Semantic Technologies for Intelligent Systems. Proceedings of the 4th International Technological Conference OSTIS-2014, Minsk, BSUIR, pp. 471–482 (2014)
12. Jun, H.B., Kiritsis, D., Xirouchakis, P.: A primitive ontology model for product lifecycle MetaData in the closed-loop PLM. In: Goncalves, R.J., Muller, J.P., Mertins, M., Zelm, M. (eds.) Enterprise Interoperability II. New Challenges and Approaches, pp. 729–740. Springer, Heidelberg (2007)
13. Tarassov, V.B., Fedotova, A.V., Stark, R., Karabekov, B.S.: Granular meta-ontology and extended Allen's logic: some theoretical background and application to intelligent product lifecycle management systems. In: Proceedings of the 4th International Conference on Intelligent Systems and Applications (INTELLI'2015), St. Julians, Malta, Copenhagen, pp. 86–93. IARIA XPS Press (2015)
14. Khoroshevsky, V.F.: Knowledge space in the Internet and Semantic Web (Part 1). In: Artificial Intelligent and Making Decisions, № 1, pp. 80–97 (2008)
15. Fedotova, A.V.: Intelligent system of planning maintenance and equipment repair at the management of the high valuable product. In: Intelligent Systems and Technologies: Up-to-date State and Perspectives. Edited volume of the International Summer School on Artificial Intelligence for Students, Postgraduate Students and Young Scientists, Tver, pp. 153–161 (2013)
16. Fedotova A.V., Vetrov A.N., Tarassov V.B.: Information Granulation in Life Cycle Modeling for Complex Technical Systems. In: On-line Journal "Naukovedenie", № 5 (18) (2013). http://naukovedenie.ru/PDF/53tvn513.pdf
17. Gavrushin, S.S., Evgenev, G.B.: Multicriteria optimization in the life cycle of products. In: Information Technologies, № 2, pp. 37–42 (2014)
18. Shunkevich, D.V.: Support tools of knowledge-based systems component design. In: Open Semantic Technologies for Intelligent Systems (OSTIS-2015): Proceedings of the 5th International Technological Conference, Minsk, BSUIR (2015)

Application of Swarm Intelligence for Domain Ontology Alignment

Alexandra Semenova and Viktor Kureychik

Abstract In the given paper Alexandra Semenova we propose an approach for ontology alignment. To find the optimal ontology alignment a combination of two techniques is presented. In the suggested approach definitions of informal concepts of the ontologies are mapped to formal normalized concepts of intermediate ontology WordNet. Then swarm intelligence approach is applied where semantic similarity weights are used as particles. Task of particle swarm optimization is initialized by random particles population following by search of optimal solutions through updating of generations. Our approach provides several benefits in terms of good accuracy, refinement of alignments over time and minimizing of man subjectivity. Here, the results of ontology alignment are shown for the domain of transport system.

Keywords Ontology alignment · Swarm intelligence · Knowledge bases · Transport system · Swarm optimization · Semantic integration

1 Introduction

Nowadays ontologies have been realized as a key technology for industrial computer systems. Nowadays there are a lot of specific domain ontologies under Semantic Web. For example, ontologies are used in applications dealing with knowledge management, business processes modeling, intellectual data integration, information retrieval, data base integration, bioinformatics, e-commerce, and etc. Cooperative accumulation of ontologies provides boundless collection of knowledge about a particular universe. Though ontologies reflect a conceptual vision of their developer's word model, the increase in the amount of models and their

A. Semenova (✉) · V. Kureychik
Southern Federal University, Taganrog, Russia
e-mail: alexaforum@rambler.ru

V. Kureychik
e-mail: kur@tgn.sfedu.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 450, DOI 10.1007/978-3-319-33609-1_23

heterogeneity is becoming a global scale integration problem. Heterogeneous data models in computer systems results in integration problem when two or more of these systems need to interact and exchange information. To enable collaboration between different ontologies from the same domains, ontologies developers require the semantic alignment of the different formalisms.

Ontology alignment is a key aspect in the field of knowledge exchange so it perhaps provides a solution for data heterogeneity problem. The alignment process identifies the equivalences between some concepts and theirs properties of the ontologies as well as different levels of confidence. So ontology alignment is an urgent problem both for development of new ontological engineering branches and for development of intellectual systems for natural language processing.

An alignment is a specification of set of correspondences between discrete entities (which can be classes, properties, rules, predicates, etc.) of two heterogeneous ontologies. The process of ontology alignment may be represented as a function based on the set E of all entities $e \in E$ and based on the set of possible ontologies O is a partial function [1]

$$align: E \times O \times O \rightharpoonup E \quad (1)$$

It should be noted that $align_{O_1 O_2}$ holds for $align(e, O_1, O_2)$. Once an alignment between two ontologies O_1 and O_2 is established, then entity e is aligned with entity f if $align(e) = f$. A pair of entities (e,f) that is not yet aligned and for which appropriate alignment criteria still need to be tested is called a candidate alignment.

To find a unique best alignment is an inherently difficult task. The weighted average is very often used as a fuzzy aggregate [2]. But these proposals use, in the best of the cases, weights determined by an expert. To avoid the expert intervention we suggest not using weights from an expert, but compute those for obtaining the optimum alignment function applying swarm intelligence. This allows solving ontology alignment problem accurately and without human intervention.

Therefore, the ontology alignment problem can be stated as follows. Given two source ontologies find the most reasonable alignment between them by mapping the ontologies to an intermediate ontology, and then, maximize the number of correspondences in the obtained alignment.

Generally ontologies feature such semantic components as concept names, properties and relations between them. The main aim of ontology mapping is to find the concepts that represent the meaningful connection within similar concepts of different ontology. Different measures, allowing checking the similarity of the concepts (string similarity, semantic similarity and similarities between the concepts) exist. Once the similarity features of two given entities from two ontologies are selected and measured, they will be aggregated. One of the wide-spread techniques to compute the optimal aggregation for different types of similarity measures is weighted sum [3]. However choosing the optimum parameters of these techniques such as thresholds and other constraints is difficult. Weights may be defined for example by expert knowledge or by preliminary domain knowledge. The main problems of supervised approaches for ontology alignment are relatively big run-time and complexity of obtaining data samples having enough volume and quality [4].

Ontology alignment using swarm intelligence is being used as an alternative to more conventional ontology alignment technique. Particle swarm optimization (PSO) is a population-based search algorithm. It is initialized with a population of random solutions—particles. Each particle is associated with a velocity with which it flies through the search space. The main advantage of PSO against the other evolutionary computation techniques is that it has less parameter to adjust. Hence PSO does not have any complicated evolutionary operators (mutation, crossover, and etc.) as in genetic algorithm.

Thus, the alignment problem is transformed into a task of finding optimal weights of semantic measures between concepts. Following these premises, we propose a new approach to generate ontology alignments. The key difference of the suggested approach is that it relies on the combination of mapping ontologies to an intermediate ontology, and then, getting an optimal solution for ontology alignment via swarm intelligence [5]. In the suggested approach definitions of informal concepts of the ontologies are mapped to formal normalized concepts of intermediate ontology WordNet. Then swarm intelligence approach is applied where semantic similarity weights are used as particles. Task of particle swarm optimization is initialized by random particles population following by search of optimal solutions through updating of generations. Our approach provides several benefits in terms of good accuracy, refinement of alignments over time and minimizing of man subjectivity. Here, the results of ontology alignment are shown for the domain of transport system.

The paper has the following structure. In Sect. 2 combined ontology alignment based on swarm intelligence is depicted. Then in Sect. 3 implementation of ontology alignment for the transport system domain is performed.

2 Combined Ontology Alignment Based on Swarm Intelligence

As was early said we are going to present a combined approach for ontology alignment based on swarm intelligence. Visualization of the suggested approach for optimal ontology alignment is depicted on the Fig. 1.

Firstly, an alignment between two input ontologies O_1 and O_2 is computed with the help of WordNet. For this purpose definitions of informal concepts of the ontologies are mapped to formal normalized concepts of intermediate ontology WordNet. To do this we suggest a well-known technique that is called Latent Semantic Indexing [5].

Let's consider a sequence of operations to optimize ontology alignment by maximization of a number of correspondences in the obtained ontology applying swarm intelligence.

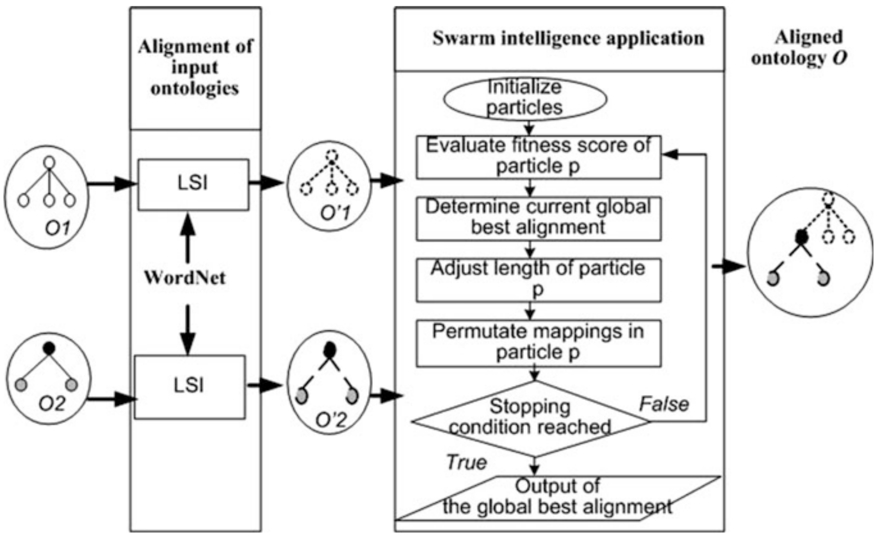


Fig. 1 Stages of optimal ontology alignment

Particle Swarm Optimization is based on a set of search techniques, inspired by the behavior of natural swarms, for solving optimization problems. In Particle Swarm Optimization, particles represent candidate solutions (in our case candidate alignments). Population of these particles is called a swarm. A swarm includes N particles moving around a D-dimensional search space (in our approach particles have different dimensionality, according to [6]). By dimensionality we understand the number of correspondences in the alignment it currently represents. If dimensionality equals 0 then a particle is an empty alignment.

Particle Swarm Optimization is adapted for ontology alignment to achieve the following two goals:

- Identify the most reasonable alignment.
- Maximize the number of correspondences in the alignment.

The first objective is what is most widely understood as the ontology alignment problem. The second objective arises out of the observation, that there is most likely not a complete overlap of ontologies to be aligned, i.e. not all ontological entities of either ontology will be involved in the desired alignment. However, one still does not want to miss any of the valid correspondences, so the goal is to maximize the number of correspondences contained in the alignment.

The population in Particle Swarm Optimization is initialized with random particles and the candidate solutions or particles move around the search space with the goal to achieve optimal fitness.

Initially, each particle has a position and velocity and the position and velocity of each particle are updated. The updated velocity vector is used to generate a new particle X.

Let's $\vec{X}_p$ be a particle that is represented as a vector, $p \in \{1, \ldots, N\}$.
where $c_{(p,k)}$ (for each $j \in \{1, \ldots, k\}$) is a correspondence. By correspondence $c_{(p,k)}$ we understand a pair of entities (e_l, e_2) meaning that e_l and e_2 refer to the same real-world entity.

Total number of correspondences in an alignment with mapping 1:1 be at most n where $n = \min\{|C_1|, |C_2|\} + \min\{|P_1|, |P_2|\}$.

Particle dimensionality (k) is variable and $k \in \{1, \ldots, n\}$. Let $F(\vec{X}_p)$ be the fitness of an alignment A for a particle $\vec{X}_p$. Each particle holds the configuration of the best alignment of dimensionality $l \in \{1, \ldots, n\}$ (pBest) according to F that is designated as [7]

$$\vec{B}_p = \left\{d_{(p,1)}, d_{(p,2)}, \ldots, d_{(p,l)}\right\} \tag{2}$$

where for each $j \in \{1, \ldots, l\}$, $d_{(p,l)}$ is a correspondence. According to [7], the dimensionality l of the best alignment (pBest) does not need to coincide with the dimensionality k of its current configuration.

The best performing parameter configuration any particle in the swarm has ever represented according to F is denoted by [7]

$$\vec{G} = \{d_1, d_2, \ldots, d_m\}, \tag{3}$$

where for each $j \in \{1, \ldots, m\}$, d_j is a correspondence. Its dimensionality is $m \in \{0, \ldots, n\}$.

Let's F be a fitness vector of a particle and it is denoted by a 2-by-k array [7]

$$\vec{F}_p = \begin{pmatrix} f_{(p,1)} & f_{(p,2)} & \cdots & f_{(p,k)} \\ c_{(p,j1)} & c_{(p,j2)} & \cdots & c_{(p,jk)} \end{pmatrix} \tag{4}$$

mapping a fitness $f_{(p,\mu)}$ to each correspondence $c_{(p,j\mu)}$. The vector is ordered by its fitness values.

A velocity vector is defined as another 2-by-k array [7]

$$\vec{V}_p = \begin{pmatrix} v_{(p,1)} & v_{(p,2)} & \cdots & v_{(p,k)} \\ c_{(p,j1)} & c_{(p,j2)} & \cdots & c_{(p,lk)} \end{pmatrix} \tag{5}$$

mapping a proportional likelihood $v_{(p,\mu)}$ to each correspondence $c_{(p,l\mu)}$. Note that the vector is ordered by its proportional likelihoods. Proportional likelihoods are used to raise the probability of those correspondences to be preserved in a particle that are also present in the personal and global best alignments. Initially, for each $c_{(p,l\mu)}$, $v_{(p,\mu)}$ is set to 1. This initialization is also done for new correspondences joining the particle during its evolution. The update of the proportional likelihoods $v_{(p,\mu)}$ is then done in two steps, using two parameters $\beta \in \mathbb{R}^+$ and $\gamma \in \mathbb{R}^+$. Firstly, if

$c_{(p,l\mu)}$ is present in $\vec{B}_p$, add β to $v_{(p,\mu)}$. If it is present in $\vec{G}_p$, add $v_{(p,\mu)}$ to γ. After this, each $v_{(p,\mu)}$ is multiplied by a uniform random number $\phi_\mu \in (0,1)$.

Let's calculate constant set which will not be replaced during a random reinitialization, $k \in (0,1)$ [7]

$$F_{(p,k)} = \left\{ c_{(p,j\mu)} | \mu \in \{1, \ldots, k \cdot k\}, j_\mu - \text{ordering as in } \vec{F}_p \right\}, \tag{6}$$

$$V_{(p,k)} = \left\{ c_{(p,l\mu)} | \mu \in \{1, \ldots, k \cdot k\}, l_\mu - \text{ordering as in } \vec{V}_p \right\} \tag{7}$$

Consequently, the sets $F_{(p,k)}$ and $V_{(p,k)}$ contain such matching particles that $k \cdot k$ are well appreciated and have the highest rank according to their values $v_{(p,\mu)}$, respectively. The unchanging set is defined as a set

$$K_{(p,k)} = F_{(p,k)} \cap V_{(p,k)}, \tag{8}$$

containing such matching, which are part of both sets $F_{(p,k)}$ and $V_{(p,k)}$.

For more strict convergence with respect to optimal alignment additional secure set of correspondences, which will never be replaced in the particle, is introduces as follows:

$$S_{(p,\sigma)} = \left\{ c_{(pj\mu)} | f_{(p,\mu)} < \sigma \right\}, \tag{9}$$

where $\sigma \in (0,1)$—is a threshold for alignment for correspondences which are to be included in a secure set. Since there is a possibility of getting stuck in the local optimum for the alignment, it is necessary to choose a sufficiently small value of σ. Each update algorithm at first calculates the new length of the particle k' in accordance with the process of self-adaptation. Then, the particle retains the set $S_{(p,\sigma)} \cup K_{(p,k)}$ and replaces the remaining matching $k' - |S_{(p,\sigma)} \cup K_{(p,\sigma)}|$ by new random values. This behavior ensures the convergence of each particle to the optimal, as an keep-set of continuously increasing and deviation due to random re-initialization are less drastic with the evolution of the swarm [8].

Using this strategy increases search dimension, which is essential for finding the optimal solution. In the presented combined approach based on swarm intelligence the dimension of each particle is updated at each iteration. The population size is much larger than the number of possible particle length. Regarding the problem of ontology alignment, the value of the possible length of the particle can be much more, because it depends on the size of the input ontologies, i.e. the number of classes and their properties. Thus, the problem may be significantly more complicated, if you increase the size of the population. In order to obtain optimal size of alignment, it is necessary to consider the dynamic correspondence of particle lengths. To this end, during the initialization of each particle is necessary to set a random number of correspondences. As a rule, the actual number of matches is unknown, so initially introduced the assumption of a uniform distribution of zero and the maximum possible correspondences of two ontologies, considering only the alignment of 1:1 [9].

3 Implementation of Ontology Alignment for the Transport System Domain

According to Fig. 1 the proposed approach for ontology alignment consists of two stages. During the first stage mapping to WordNet according to the LSI method is performed. This stage comprises the following steps [10]:

- Choose a concept from the ontology. Let C be the concept name.
- Get all WordNet senses lexicalized by C', where C' is a linguistic variation of C. These senses provide the focus of the algorithm for C.
- Get the hyperonyms and hyponyms of all C' senses.
- Build the "association matrix. An $n \times m$ matrix that comprises the n more frequently occurred terms in the vicinity of the m WordNet senses found in step 2.
- Build a query string using the terms in the vicinity of C. The query string is a sequence of digits, each digit taking value 0 if a term in the vicinity of C does not exist in the set of n, and 1 if a query term exists in the set of n.
- Find the ranked associations between C and C' senses by running the Latent Semantics Analysis (LSA) function and consider the association with the highest grade.

The second stage is calculation of swarm intelligence. The algorithm for swarm intelligence computation comprises three parts: initialization step, the swarm iteration, and an update procedure to determine the new configuration of each particle [11].

The computation of an alignment starts with an initialization. In this initialization step, each particle is initialized with a random number of correspondences. It also encompasses evaluation, i.e. computation of the fitness value of each correspondence and the initial assertion of the personal best alignment.

The execution of the algorithm is an iterative, guided evolution of the particle swarm. The globally best alignment is updated iteratively, if a new best performing particle is seen. The guided evolution of particles behaves according to the update procedure. Note, that each particle can be evaluated and updated in parallel, which is one of the main advantages of this approach [12].

Now we will consider an example of application the suggested approach. On Fig. 2 two input ontologies are shown.

Suppose that alignment correspondences comprise five alignments (k = 5):

- $c_{(p,1)}$ = (Product_Model, Category) with fitness value of 3.64;
- $c_{(p,2)}$ = (Body_Style_Value, Size) with fitness value of 1.25;
- $c_{(p,3)}$ = (Fuel_Type, Fuel) with fitness value of 0.39;

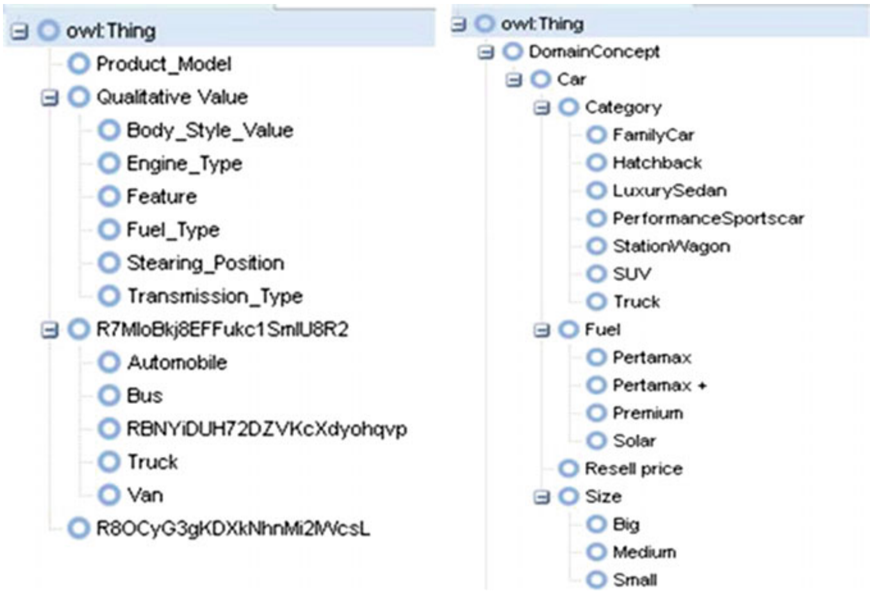


Fig. 2 Input ontologies for alignment

- $c_{(p,4)}$ = (Van, StationVagon) with fitness value of 7.39;
- $c_{(p,5)}$ = (Qualitative_Value, RecelPrice) with fitness value of 1.54.

Then fitness values of single aligns were defined as follows:

$$\vec{F}_p = \begin{pmatrix} 0,04 & 1,23 & 1,55 & 3,65 & 7,54 \\ c_{(p,3)} & c_{(p,2)} & c_{(p,5)} & c_{(p,1)} & c_{(p,4)} \end{pmatrix}$$

Note that the values in the matrix are sorted in ascending order, and the lowest value is the best estimate. Velocity vector Vp initialization has been performed, for all $v_{(p,k)}$ the value of 1 was assigned:

$$\vec{V}_p = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ c_{(p,1)} & c_{(p,2)} & c_{(p,3)} & c_{(p,4)} & c_{(p,5)} \end{pmatrix}$$

Assume that the correspondences $c_{(p,2)}$, $c_{(p,3)}$ and $c_{(p,5)}$ also belong to $\vec{B}_p$, and $c_{(p,3)}$ belongs to $\vec{G}_p$. Then, add the parameters β and γ, respectively (for example, $\beta = 0{,}4$ and $\gamma = 0{,}5$):

$$\vec{V}_p = \begin{pmatrix} 1 & (1+\beta) & (1+\beta+\gamma) & 1 & (1+\beta) \\ c_{(p,1)} & c_{(p,2)} & c_{(p,3)} & c_{(p,4)} & c_{(p,5)} \end{pmatrix}$$

After adding the parameters, we multiply each $v_{(p,k)}$ on a universal random number $\phi_j \in (0,1), \forall\{1,\ldots,5\}$. The array will then be sorted in descending order, as the highest value corresponds to the highest likelihood. Thus, we obtain the following

$$\vec{V}_p = \begin{pmatrix} 1,34 & 1,12 & 0,88 & 0,76 & 0,32 \\ c_{(p,2)} & c_{(p,5)} & c_{(p,4)} & c_{(p,3)} & c_{(p,1)} \end{pmatrix}$$

Suppose, $k = 0.6$, then the keep-set $K_{(p,k)}$ is defined as the intersection of the first $k \cdot k = 3$ corresponding arrays $\vec{F}_p$ and $\vec{V}_p$. The result is:

$$K_{(p,k)} = \left\{\mathrm{c}_{(p,2)}, \mathrm{c}_{(p,5)}\right\}.$$

Let σ = 0,1 and secure the set is given as $S_{(p,\sigma)} = \left\{c_{(p,3)}\right\}$.

The result of the algorithm is the following set:

$$S_{(p,\sigma)} \cup K_{(p,k)} = \left\{c_{(p,2)}, c_{(p,3)}, c_{(p,5)}\right\}.$$

Also, the algorithm will replace the remaining corresponds to a random new.

To sum it up the proposed approach helps to achieve the following two goals: identify the most reasonable alignment and maximize the number of correspondences in the alignment.

4 Conclusion

In the given paper we propose a combined approach for ontology alignment that consists of two stages.

During the first stage mapping to WordNet is performed. In the suggested approach definitions of informal concepts of the ontologies are mapped to formal normalized concepts of intermediate ontology WordNet. So definitions of informal concepts of the ontologies are mapped to formal normalized concepts of intermediate ontology WordNet. For alignment of input ontologies with WordNet we've used a well-known technique that is called Latent Semantic Indexing. Latent Semantic Indexing (LSI) tries to overcome the problems of lexical matching by using statistically derived conceptual indices instead of individual words for retrieval.

Then swarm intelligence approach is applied where semantic similarity weights are used as particles. Particle Swarm Optimization (PSO) is based on a set of search techniques, inspired by the behavior of natural swarms, for solving optimization problems. In Particle Swarm Optimization, particles represent candidate solutions (in our case candidate alignments). Task of particle swarm optimization is initialized by random particles population following by search of optimal solutions through updating of generations.

This strategy increases search dimension, which is essential for finding the optimal solution. In the presented combined approach based on swarm intelligence the dimension of each particle is updated at each iteration. The population size is much larger than the number of possible particle length. Our approach provides several benefits in terms of good accuracy, refinement of alignments over time and minimizing of man subjectivity. Here, the results of ontology alignment are shown for the domain of transport system.

Acknowledgement This work is partially supported by the State Order "Management of scientific researches" 213.01/OPNI 012 and by RFBR grant 15-07-05523.

References

1. Euzenat, J., Shvaiko, P.: Ontology Matching, 2nd edn. Springer, Heidelberg (2013)
2. Semenova, A.V., Kureychik, V.M.: Domain ontology development for linguistic purposes. In: 9th International Conference on Application of Information and Communication Technologies (AICT), pp. 83–88. IEEE Press, New York (2015)
3. Euzenat, J., Ferrara, A., van Hage, W.R., Hollink, L., Meilicke, C., Nikolov, A., Ritze, D., Scharffe, F., Shvaiko, P., Stuckenschmidt, H., Trojahn dos Santos, C.: Results of the ontology alignment evaluation initiative 2011. In: 6th OM Workshop (2011)
4. Sowa, J.: Knowledge Representation: Logical Philosophical, and Computational Foundations, 608p. Pacific Grove (2000)
5. Gladkov, L.A, Kureychik, V.V., Kureychik, V.M.: Bioinspired Optimization Methods, 384p. Fizmatlit, Moscow (2009)
6. Lapshin, V.: Ontology in Computer Science. Nauchniy mir, Moscow (2010)
7. Gladkov, L.A., Kureychik, V.V., Kureychik, V.M.: Diskretnaya matematika: Teoriya mnozhestv, algoritmov, algebra logiki. TTI UFU, Taganrog, p. 312 (2009)
8. Gladkov, L.A., Kureychik, V.V., Kureychik, V.M.: Genetic Algorithms, 2nd edn. Fizmatlit, Moscow (2006)
9. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks, vol. IV, pp. 1942–1948. Perth, Australia, IEEE Service Center, Piscataway, NJ (1995)
10. Bock, J., Hettenhausen, J.: Discrete Particle Swarm Optimisation for Ontology Alignment. Information Sciences (Special Issue on Swarm Intelligence and Applications) (2010)
11. Kopaygorodskiy, A.N.: Primenenie ontologiy v semanticheskih informacionnih sistemah. Ontologiya proektirovaniya, № 4(14), pp. 90–98 (2014)
12. Sánchez, D., Batet, M.: Ontology-based semantic similarity: a new feature-based approach. J. Biomed. Inf. **44**, 118–125 (2012)

Mathematical Models of Learning Materials Estimation Based on Subject Ontology

M.A. Shpak, E.V. Smirnova, A.P. Karpenko and A.V. Proletarsky

Abstract Two mathematical models are under consideration in this paper: the ontology usage as a tool of automation has become popular nowadays. There are comparisons between the subject ontology's part and students' cognitive map which are being developing during the test. It gives a possibility to assess student's knowledge skills as well as some new characters of the educational outcomes. The core competencies form the basis of the student's ability to learn, as well as interdisciplinary concepts or metaconcepts. The first and second sections of the article present a subject ontology model. The third section is devoted to meta-concepts testing based on metasubject' ontology. In conclusion we formulate our main results and prospects of its development.

Keywords Mathematical model · Cognitive style · Way of thinking · Student metacompetences

1 Introduction

Russian Federal State Educational Standard identifies three levels of educational outcomes: personal, objective and interdisciplinary. The ontology usage as a tool of automation has become popular nowadays. In particular the comparison between the

M.A. Shpak · E.V. Smirnova (✉) · A.P. Karpenko · A.V. Proletarsky
Bauman State Technical University, Moscow, Russia
e-mail: evsmirnova@bmstu.ru2; elena.galiamova@mail.ru

M.A. Shpak
e-mail: krolex@mail.ru1

A.P. Karpenko
e-mail: apkarpenko@mail.ru3

A.V. Proletarsky
e-mail: pav@bmstu.ru4

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 450, DOI 10.1007/978-3-319-33609-1_24

objective ontology's part and students' cognitive map which is developing during the test, gives us a possibility to assess his knowledge skills as well as some new characters of the educational outcomes. The educational outcomes as a development of universal educational actions ensuring mastery of core competencies form the basis of the ability to learn, as well as interdisciplinary concepts (metaconcepts) [1]. Control technique of the student conceptual knowledge is reviewed, for example, in [2]. This technique based on the student cognitive map of the learner (SCM[1]) [3], which formalizes his notion about a fragment of a subject semantic network (SSN[2]) and, ideally, coincides with this fragment. Control of students mastering the conceptual structure of this fragment is reduced to comparing sub graph SSN corresponding to the fragment, and the graph that corresponds to the SCM [3]. This article can be considered as the development of the articles [2] in the direction of learning materials and complexity testing based on metaconcepts ontology.

The first and second sections of the article present a subject ontology model and the SCM. The third section is devoted to metaconcepts testing based on metadisciplines ontology which includes methods to construct local and extended Cognitive Map Learning (CML[3]). In conclusion we formulate our main results and prospects of its development.

We assume further that $|A|$ is a number of elements (power) of a countable set of elements A; a_i—i-th element of the set.

2 The Subject Ontology Model

Let O—an ontology of the subject. Semantic network (SN[4]) or ontology O defines a tuple $SS = \langle C,\ R,\ W_C,\ W_R \rangle$, where the following notation is present: $C = \{c_i\}$ a set of concepts SN; $R = \{r_i\}$ a set of relations between the concepts of the set $C; W_C = \{w_{C,i}\}$. $W_R = \{w_{R,i}\}$-values the importance of measures set C concepts and relationships R, respectively. Methods for assessing the value of these measures are considered, for example, in article [4].

If the concepts $c_{i_1},\ c_{i_2},\ i_1 \neq i_2$ in the SN (1) are linked to each other with some relation out of the number of relations R, then we say that these concepts are *related information*.

Let's collate the SN (1) weighted directed graph without loops G whose vertices correspond to the concepts of ontology O, and the arc—communication links of these concepts between each other. The weights of the vertices of the graph G are equal to the concept importance, and the weights of the arcs—the importance of appropriate relationships, respectively.

[1]SCM—Student cognitive map.

[2]SSN—Subject semantic network.

[3]CML—Cognitive map learning.

[4]SN—Semantic network.

Let $h(c_{i_1}, c_{i_2})$ be a measure of the distance between concepts $c_{i_1}, c_{i_2} \in C$. Value $h(c_{i_1}, c_{i_2})$ can be determined in several ways, e.g., the number of arcs of the shortest path in the graph between the vertices corresponding to these concepts can be used as its value [5, 6].

Let us introduce the sub graph $G_i(h)$, where $h > 1$ of the graph G corresponding to the concept $c_i \in C$. Sub graph $G_i(h)$ includes all the concepts $C_i(h) \subset C$ of the graph G, where the distance between them and the concept c_i does not exceed the value h. A set of relations linking the concepts of a set $C_i(h)$ is denoted by $R_i(h)$; the corresponding fragment of the SN-$SS_i(h)$.

Obviously, the sub graph $G_i(1)$ is a set of concepts associated with metaconcept's information c_i.

3 Model of Cognitive Map Learning (CML)

Let's select the local and extended CML. Extended CML corresponds to a fragment of the semantic network $SS_i(h)$ and the local one—to the fragment $SS_i(1)$.

The model of the extended CML is defined by tuple

$$CM_i(H) = \langle \tilde{C}_i(h), \tilde{R}_i(h) \rangle, \; h > 1, \tag{1}$$

where the following notation are present: $\tilde{C}_i(h)$—a set of concepts of C, which is determined as a part of SN $SS_i(h)$ by the students; $\tilde{R}_i(h)$—a set of relations of R between the concepts in $\tilde{C}_i(h)$.

Let's present CML (2) as a directed, possibly disconnected weighted graph without loops $\tilde{G}_i(h)$ whose vertices correspond to the concepts $\tilde{C}_i(h)$ and arcs—to relationships $\tilde{R}_i(h)$.

Local CML is defined by a similar tuple

$$CM_i(1) = CM_i = \langle \tilde{C}_i, \tilde{R}_i \rangle \tag{2}$$

and is present in the form of a graph $\tilde{G}_i(1) = G_i$ similar to the graph $\tilde{G}_i(h) = G_i(h)$. Here are the definitions: $\tilde{C}_i = \tilde{C}_i(1)$, $\tilde{R}_i = \tilde{R}_i(1)$.

4 Testing Technique Based on Metasubject Ontology

Metasubject ontology model coincides with the model discussed in claim 1, with the replacement of the term concept with the term metaconcept. The respective semantic network is defined as MSN. Expansion and local CML corresponding to MSN[5] are $MSS_i(h)$, $MSS_i(1) = MMS_i$ means $MCM_i(h)$, $MCM_i(1) = MCM_i$; where $h > 1$.

[5]MSN—Metasubject ontology.

4.1 Methods to Construct the CML

Local CML MCM_i is constructed by using a local test T_i. Test task for the test T_i consists generally of two parts and at the level of content is as follows:

1. From the list of possible answers, create a set of metaconcept $\tilde{C}_i$, information related to metaconcept c_i;
2. From the list of possible answers, create a set of types of relationships $\tilde{R}_i$ linking metaconcept's of the set $\tilde{C}_i$ with metaconcept's c_i.

List of possible answers to the first question contains the following metaconcepts:

- all metaconcept's of a set C_i are correct answers;
- specified number of randomly selected without repetition metaconcept's c_{i_l}, $l = 1, 2 \ldots$ From among metaconcept's set C for which the measure $h(c_i,\ c_{i_l})$ is equal to two (incorrect answers);
- the required number of randomly selected without repetition metaconcept's c_{i_m}, $m = 1,\ 2,\ \ldots$ from among metaconcept's set C for which the measure $h(c_i,\ c_{i_m})$ is equal to three (wrong answers);
- and so on until the predetermined number of incorrect answers.

For the second question, a list of possible answers for each pair "metaconcept of the set $\tilde{C}_i$-is a metaconcept" contains all of the relation R.

To construct the CML $MCM_i(h),\ \ h > 1$ using a set of tests $\mathbf{T}_i(h) = \{T_{i_j}\}$, where a local test T_{i_j} is designed to assess the level of assimilation of student's metaconcept $c_{i_j} \in C_i$.

Value $\eta_i(h) = \frac{|\mathbf{T}_i(h)|}{|C_i|} \in (0;\ 1]$ we call coverage level test $\mathbf{T}_i(h)$ SN set $MSS_i(h)$. Note that the value $|\mathbf{T}_i(h)|$ is equal to the number of local tests in the set $\mathbf{T}_i(h)$ and, at the same time, the number of test metaconcept set C_i.

Test T_{i_j}, $j \in [1\!:\ |\mathbf{T}_i(h)|$ formed by the following rule:

1. randomly without repetition choose $|\mathbf{T}_i(h)|$ metaconcept c_{i_j} from the set C_i;
2. for each of metaconcept's c_{i_j} discussed above rules by creating tests T_{i_j}.

Answers to student test items test $\mathbf{T}_i(h)$ suite generate $|\mathbf{T}_i(h)|$ sets of responses $\{\tilde{C}_{i_j},\ \tilde{R}_{i_j}\}$, CML is based on these responses to the following rules:

- metaconcept's form a set $\tilde{C}_i$ of a set $\{\tilde{C}_{i_j}\}$ by combining same name metaconcept's in one;
- create a set $\tilde{R}_i$ of relations by combining a set $\{\tilde{R}_{i_j}\}$ of relations linking the corresponding metaconcept's.

4.2 Evaluation Metrics of Students' Metaconcept Knowledge

All metaconcept metrics of students' assimilation level presented further have different value and scale. Therefore program realization requires normalization of the values of these metrics, for example, according to the scheme [2]. Construction of normalized metrics based on different linear and nonlinear scales of assessments were reviewed, for example, in the work [2].

Metrics $p_1(c_i)$. Let us introduce into consideration subsets $\tilde{C}_i^t$, $\tilde{R}_i^t$ appropriate for sets $\tilde{C}_i$, $\tilde{R}_i$, which contain the correct students' answers. Relative numbers of correct answers in this sets we believe to be

$$\bar{n}_{c,i}^t = \frac{\tilde{C}_i^t}{C_i}, \ \bar{n}_{R,i}^t = \frac{\tilde{R}_i^t}{R_i} \tag{3}$$

respectively. Similarly, the definition of a subset of wrong answers $\tilde{C}_i^f$, $\tilde{R}_i^f$ and respective value $\bar{n}_{c,i}^f$, $\bar{n}_{R,i}^f$. In calculating these values does not match the values in certain elements should be included as elements of the sets C_i, R_i, not included into sets $\tilde{C}_i$, $\tilde{R}_i$, so as elements from second set, not included into first set.

5 Conclusion

There are cases of ontology usage in mobile language learning, in self-organizing neural networks, etc. In middle and high schools metasubject training can be provided by the so-called interdisciplines, as well as through appropriate modification methods of teaching traditional subjects. Accordingly, we consider two approaches to proficiency learners' metaconcept's testing. The paper authors focus their efforts on new test algorithms development with new methodology of testing based on subject ontology structure. The aim of investigation is to create new type of tests as well as new computer program as a tool for e-learning education.

Proposed original methods of control metaconcept's student's knowledge on the basis of metasubject ontology and based on the ontology of traditional academic subjects.

In development work, the authors plan to develop appropriate algorithms and software, as well as testing of the developed methods, algorithms and software for solving practical problems of estimating metasubject students' knowledge.

The work was supported by the Russian Ministry of Education (project ID: RFMEFI57714X0135).

References

1. Cruz, B.A., Eremeeva, E.V.: The definition of competencies metasubject junior student. Mod. Probl. Sci. Educ. **6** (2013). http://www.science-education.ru/ru/article/view?id=11014
2. Belous, V.V., Karpenko, A.P., Smirnova, E.V.: Evaluation of conceptual knowledge of the student based on hierarchical clustering role. Educ. Sci.: Sci. Tech. Electron. Ed. **9** (2014). http://technomag.bmstu.ru/doc/726237.html
3. Avdeeva, Z.K., Kovriga, S.V., Makarenko, D.I., Maksimov, V.I.: The cognitive approach in the management. Probl. Upr. **3**, 2–8 (2007)
4. Ontology conceptual data models. http://www.cambridgesemantics.com/resources/blog/ontologies-conceptual-models
5. Greshilova, A.V.: The content of interdisciplinary competences of secondary professional education students. Magister Dixit **1**, 13 (2014). http://md.islu.ru/sites/md.islu.ru/files/rar/greshilova_statya_md_0.pdf
6. Scharnhorst, A., Ebeling, W.: Evolutionary search agents in complex landscapes—a new model for the role of competence and meta-competence (EVOLINO and other simulation tools). The Virtual Knowledge Studio: website. http://virtualknowledgestudio.nl/documents/_andreas-charnhorst/arxiv_final.pdf

Hybridization of Fuzzy Inference and Self-learning Fuzzy Ontology-Based Semantic Data Analysis

Nadezhda Yarushkina, Vadim Moshkin, Ilya Andreev, Victor Klein and Ekaterina Beksaeva

Abstract The article covers a method of describing the domain by means of fuzzy ontology. This paper describes fuzzy ontology self-learning algorithms based on extracting terms from text corpora. The method of knowledge inference based on fuzzy ontology and a set of SWRL rules (Semantic Web Rule Language) is also presented in the paper. In addition, the paper presents the results of experiments on ontology self-learning. The developed algorithms have been used in the experiments and a comparative analysis of inference algorithms was made.

Keywords Fuzzy ontology · Self-learning ontology · Fuzzy inference

1 Introduction

The use of ontology allows unifying the analysis methods and to create a specific semantic context. This context provides the perception of its results in the natural human form. Based on this, one of the important components of excellent quality of the conducted analysis is to create expert ontology.

N. Yarushkina (✉) · V. Moshkin · I. Andreev · V. Klein · E. Beksaeva
Ulyanovsk State Technical University, 32, Severny Venetz Str.,
432027 Ulyanovsk, Russia
e-mail: jng@ulstu.ru

V. Moshkin
e-mail: PostForVadim@ya.ru

I. Andreev
e-mail: ares-ilya@ya.ru

V. Klein
e-mail: vikklein93@gmail.com

E. Beksaeva
e-mail: ekaty@clukva.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 450, DOI 10.1007/978-3-319-33609-1_25

The main feature of domain ontology is the need for its permanent development and updating due to the change of the represented domain. The main method of ontology development as a form of knowledge representation is an intelligent extraction of semantics from large volumes of unstructured data, mostly from texts [1–3].

Extracting semantics from texts includes technologies, directed to extract semantic units, facts, attributes, concepts and events in order to end up the analyst to receive a structured set of data with which they can conduct further analysis of the domain [4]. The analysis of the domain includes the following steps:

1. Formation of the ontology core.
2. Expansion of the ontology core using semantic algorithms of terminology extraction from domain texts.
3. The inference of new knowledge through the integration of ontologies and a set of rules.
4. Go to step 2 using the existing ontology as core (Fig. 1) [5].

Thus, ontology engineering, using an automated extraction of knowledge from corpus, involves developing intelligent algorithms in order to obtain new knowledge through inference.

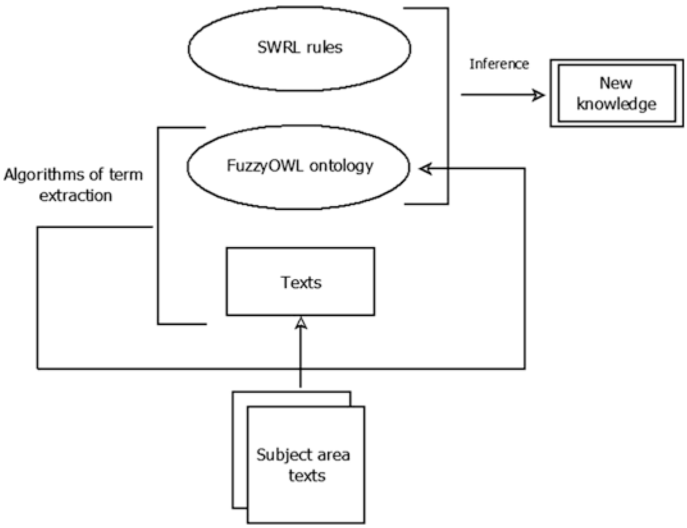


Fig. 1 Diagram of ontological analysis algorithm

2 Models and Algorithms of Ontologically-Based Data Analysis

2.1 Semantic Metrics for Evaluating Termhood

In the course of solving the problem of automated ontology expansion we developed two criteria for terms extraction from domain texts using existing ontology core:

- Thesaurus-based criterion;
- Internal linkage criterion.

Thesaurus-based criterion. A thesaurus is a reference work that lists words grouped together according to the similarity of meaning (containing synonyms and sometimes antonyms), in contrast to a dictionary, which provides definitions for words, and generally lists them in alphabetical order. Any ontology is a complicated version of the thesaurus.

Thesaurus approach assumes search of lemmas from the input words and their combinations among the terms defined in the ontology. For this purpose, each ontology class has a "HasLemma" property, which has a string value obtained by object name lemmatization.

The supporting ontology object, used in further analysis, has the degree of proximity in relation to the input word/word combinations, is calculated by the following formula:

$$k_t = \max_{i=1} \frac{n_i}{p_i} \tag{1}$$

where m is the number of all ontology objects, n_i is the number of words from the input sequence, contained in the lemma of the current ontology object, p_i is the number of words in the current ontology object.

The process of assessing the proximity of the input words to the subject area terms is shown on Fig. 2.

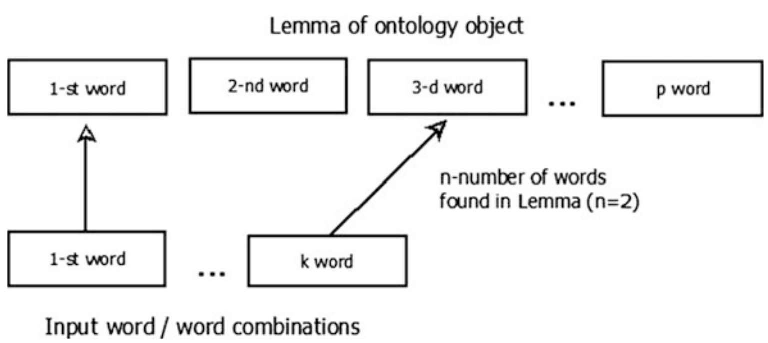


Fig. 2 Finding the supporting ontology object

Each object in the ontology has an “IsTerm” property of boolean type.

The degree of proximity of input words to the terms of domain, according to the Thesaurus criterion, is calculated by the following formula:

$$k_{Ont} = \frac{k_t}{c+1} \tag{2}$$

where k_t is the result of the first step of the analysis, c is the number of relations between the supporting ontology object and the nearest object with the true “IsTerm” value.

Internal linkage criterion. The developed metrics allows extracting terminology by not only defining the termhood of single words, but also comparing the terms from the text with ontology objects and lemmas combinations of those objects, using R_{add} relations. The Internal linkage criterion is the implementation of this.

$$t_1 + R_1 + t_2 + R_2 + \cdots R_m + t_n, \tag{3}$$

where $R_i \in R_{add}$, $t_{\mathrm{j}} \in T$, R_{add} is a set of relations that allow expanding the set of objects of the described domain through a combination of related objects lemmas. For example: properties “IsRelatedWith” and “IsPartOf”.

Thus, extracted terms which are part of other terms, consisting of more words, are not considered as terms in order to avoid redundancy.

2.2 Fuzzy Ontology Model and Inference Algorithm

The concepts and the relationships between those concepts in natural language are used to create ontologies. In the majority of cases, they are ambiguous, imprecise and have no rigid frontiers so that the transition from conventional ontology to fuzzy looks natural.

Any FuzzyOWL ontology could be described as follows:

$$I = \left(I_f, C_f, P_f, D_f, Q_f, L_f, Mod_f\right), \tag{4}$$

where $\boldsymbol{I_f}$ is the set of classes in ontology, $\boldsymbol{C_f}$ is the set of fuzzy classes in ontology, $\boldsymbol{P_f}$ is a set of properties, $\boldsymbol{D_f}$ is a set of axioms in ontology, $\boldsymbol{O_f}$ is a set of values of degrees which can be added to fuzzy axioms, $\boldsymbol{L_f}$ is a set of operators corresponding to the fuzzy logic operators, $\boldsymbol{Mod_f}$ is a set of fuzzy modifiers. The fuzzy modifier is a function which applies to a fuzzy set to change its membership function. The functions can be line hedges and tent functions [6, 7].

Inference based on fuzzy ontology has significant differences compared to the extraction of knowledge from conventional ontology (Fig. 3).

SWRL rules are used as a form of knowledge. These allow inferring from the data contained in the ontology and adding to it the new [8].

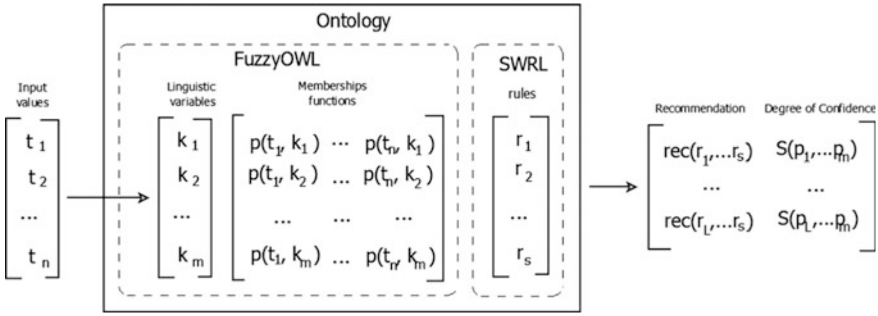


Fig. 3 Inference based on fuzzy ontology

The description of some situations by SWRL rules are as follows:

`Situation(?x) ∧ include-The-Index (?x, net)∧ has-Amount-Collision (net, large) ∧ include-The-Index (?x, net)∧ has-Background-Load (net, low) → mean (?x, Architectural-problems)`

The exponent of confidence level is a function of the incoming values of membership functions:

$$S_{Rule} = \min \mu(t_i, k_j), \tag{5}$$

where t is a term of problem domain that is present in SWRL-rule, k is the linguistic variable, j = [1…m], $\mu(t,k)$ is the value of the corresponding membership functions.

Eventually, the user of the expert system gets a set of recommendations in a natural language, each of them is assigned some value S = [0…1].

3 Implementation of Ontology Oriented Data Mining Algorithms

3.1 Ontology Oriented System for Extracting Terminology

In the course of solving problem were formed the following steps:

1. The expert created OWL-ontology corresponding problem domain;
2. Ontology oriented system for extracting terminology has been developed. This system uses described above metrics.

The developed OWL-ontology has a hierarchical organization, and includes 261 instances and about 47 relations (ways in which classes and individuals can be related to one another).

The instance of developed OWL-ontology:

```
<owl:Class rdf:ID="Butt_mill">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    > Butt_mill</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#Mill"/>
    </rdfs:subClassOf>
  </owl:Class>
```

Ontology oriented system for extracting terminology has been developed for the implementation of the described algorithms.

3.2 Inference System Based on Fuzzy Ontology

Software system TSAnalyzer 2.1 was developed in order to solve the problem of providing the new data inference based on FuzzyOWL ontology and has been modified OWL ontology analysis of the local data-processing network according to the methodology FuzzyOWL. To provide access to the ontology was used FuzzyDL (Fuzzy Description Logic) library in addition to Java-framework Apache Jena. The architecture of the developed system is shown in Fig. 4.

The Pellet reasoner is the core of TSAnalyzer 2.1. It provides inference and allows processing the data in OWL.

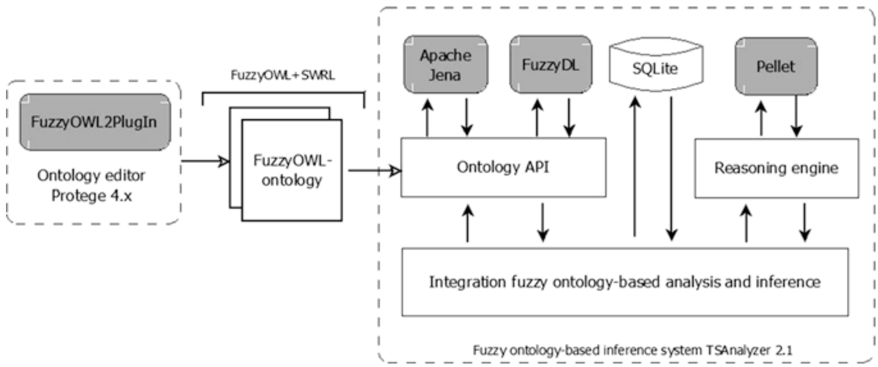


Fig. 4 The architecture of inference system based on fuzzy ontology

4 Analysis of Experiments on Extraction of Terms and Inference of New Knowledge

4.1 Experiments on Extraction of Terms

The results of experiments are intended to study the efficiency of the developed algorithms. We reviewed the results of the two indicators: using the thesaurus and the internal linkage criterion; as well as four categories of word usages: a single word, two words, three words and four words.

For the experiment was used a text of about 62,000 words from the operating guide to Computer Numerical Controlled Lathe Machine.

To evaluate the efficiency of the developed criteria the next notations were used: *Precision*, *Recall* and F_1 [9, 10].

The experimental results through the use of the thesaurus are shown in Table 1. The results of internal linkage criterion are shown in Table 2. Thesaurus criterion results were achieved considering a threshold value of k_{Ont} = 0,5.

Based on the above, the second criterion has given the best results in extracting single-word terms. The second criterion also shows better values when extracting terms consisting of two words.

Extraction of three-word terms by the thesaurus criterion may be considered satisfactory: extracted more than 50 % of the three-word terms of the domain with an average Precision.

The final results show that the extraction of terms which consist of four words are identical in both cases. Recall exceeds 1 for both criteria, which shows the extracting of terms that were not previously identified. Although similar results, the first criterion is losing to the second criterion due to the lower Precision.

Table 1 The experimental results through the use of the thesaurus

Amount of words	Terms	$k_{Ont} > 0.5$	Right	P	R	F_1
1	294	120	88	0.73	0.29	0.42
2	631	305	133	0.43	0.21	0.28
3	361	379	214	0.56	0.59	0.57
4	107	196	120	0.61	1.12	0.79

Table 2 The experimental results of the internal linkage criterion

Amount of words	Terms	Candidates	Right	P	R	F_1
1	294	168	154	0.91	0.52	0.66
2	631	431	372	0.86	0.58	0.69
3	361	370	327	0.88	0.9	0.89
4	107	159	129	0.81	1.2	0.97

Table 3 Comparison of inference methods based on plain (OWL + SWRL) and fuzzy (FuzzyOWL + SWRL) ontologies

Recommendations	OWL + SWRL	FuzzyOWL + SWRL
No decisions	2	–
Several decisions (including correct or close to correct)	4	26
Several incorrect decisions	–	–
One correct (or close to correct) decision	23	4
One incorrect decision	1	–

4.2 Inference Based on Fuzzy Ontology

A series of experiments, involving data about potential problem situations that occur during parallel LAN computing, was conducted. The aim of the experiments is to compare the two inference methods, proposed in the paper.

The object of the experiments was the local area network (LAN) of the Center of Multimedia Technologies of Ulyanovsk state technical university. Comparative characteristics of methods of inference in the simulation of problem situations is shown in Table 3.

As can be seen from the results of the experiments, the inclusion of fuzziness in the description of domain has several advantages:

- significantly reduces the risk of loss of possible outcomes of the unit inference;
- increases the flexibility of the output process;
- users are given more wide variety of options, ordered according to the relevancy to the combination of expert knowledge, embedded in the ontology and a set of rules.

5 Conclusions

Thus, experimental results suggest a high efficiency of the described methods. These methods were developed by combining linguistic algorithms of terminology extraction from large text corpora and inference based on fuzzy ontology.

Furthermore, combining the two inference methods forms a production storage and knowledge representation. Due to the fuzziness in the ontology elements, it is possible to reduce the risks of losing possible outcomes that may be important for the process.

In future work, described method will be modified by using case-based reasoning (CBR). Although in a CBR system the main source of knowledge is the set of previous experiences, our approach to CBR is towards integrated applications that combine case specific knowledge with ontology models of general domain knowledge.

References

1. Ahmad, K., Gillam, L.: Automatic ontology extraction from unstructured texts. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS, vol. 3761, p. 1330 (2005)
2. Liu, K., Hogan, W.R., Crowley, R.S.: Natural language processing methods and systems for biomedical ontology learning. J. Biomed. Inform. **44**(1), 163 (2011)
3. Borisova, N.: An approach for ontology based information extraction. Inf. Technol. Control. **12**(1), 15–20 (2015)
4. Schadow, G., McDonald, C.J.: Extracting structured information from free text pathology reports. In: Proceedings of AMIA Annual Symposium, pp. 584–588 (2003)
5. Yarushkina, N., Moshkin, V.: Methods of construction of fuzzy ontologies of complex subject areas. In: Open Semantic Technologies for Intelligent Systems. VI International Scientific and Technical Conference, pp. 401–406 (2015)
6. Bobillo, F., Straccia, U.: Fuzzy description logics with general t-norms and datatypes. Fuzzy Sets Syst. **160**(23), 3382–3402 (2009)
7. Lee, C.S., Jian, Z.-W., Huang, L.-K.: A fuzzy ontology and its application to news summarization. IEEE Trans. Syst. Man Cybern. PartB **5**, 859–880 (2005)
8. Straccia, U.: Towards a fuzzy description logic for the semantic web. In: 2nd European Semantic Web Conference, pp. 167–181 (2005)
9. Pazienza, M., Pennacchiotti, M., Zanzotto, F.: Terminology extraction an analysis of linguistic and statistical approaches. In: Proceedings of the NEMIS 2004 Final Conference, pp. 255–279 (2004)
10. Nenadic, G., Ananiadou, S., McNaught, J.: Enhancing automatic term recognition through variation. In: Proceedings of 20th International Conference on Computational Linguistics COLING'04, pp. 604–610 (2004)

Models for Supporting of Problem-Oriented Knowledge Search and Processing

Yury Kravchenko, Ilona Kursitys and Victoria Bova

Abstract In terms of search procedures and problem-oriented knowledge processing, complexity of identification and usage of key information is increasing constantly. A suggested hypothesis is based on the following statement: one of the ways to solve this problem is the improvement of semantic models for interpretation and using metadata of already-existing search profiles, pursuing similar aims, as prior data. We researched case-based reasoning in semantic search relating to knowledge filter. Concrete scientific results are: agent model, metamodel and case-model of knowledge filter which can solve problems of semantic identification of key information and processing of heterogeneous knowledge resources on the basis of ontology-based structures.

Keywords Semantic models · Knowledge search and processing · Agent models · Metamodels · Case-models

1 Introduction

Information flows from a priori and computer-generated sources are growing rapidly. In such conditions the progressive methods of knowledge management, data mining and ontology-based search cannot guarantee effective work of decision support systems [1–3]. The main problem is the complexity of identification and usage of key information, which is increasing constantly [3, 4]. One of the ways to solve this problem is the improvement of semantic modeling for interpretation and applying users' search profiles pursuing similar aims as prior data [5].

Y. Kravchenko (✉) · I. Kursitys · V. Bova
Southern Federal University, Rostov-on-Don, Russia
e-mail: krav-jura@yandex.ru

V. Bova
e-mail: vvbova@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 450, DOI 10.1007/978-3-319-33609-1_26

2 Agent Model of Knowledge Filter

Let us consider the development of agent model of knowledge filter, which can solve problems of semantic identification of key information and heterogeneous knowledge resources processing [6–8]. An important problem that prevents the improvement of information search queries efficiency in inhomogeneous distributed knowledge sources is the semantic mismatch of researcher's (user's) vision of querying and its form of expression with the rules of query form on the basis of concrete interface of used information systems [8–11].

The suggested hypothesis is an assumption that semantic identification of search query key information can be effectively carried out with the use of pre-processing filter—a knowledge filter, running on the basis of the semantic concepts taxonomy tree as an systematization of complex areas of the reality and hierarchical knowledge in order to define and arrange terms and its synonyms with a further query transformation into the most effective form. The quality of the modified query can be evaluated on the basis of a multicriteria decision support model, which determines query 'weight' from syntax, semantic, categorical and priority criteria.

Agent model of such filter (Fig. 1) allows us to coordinate knowledge search in inhomogeneous distributed sources which contain unstructured or semi-structured

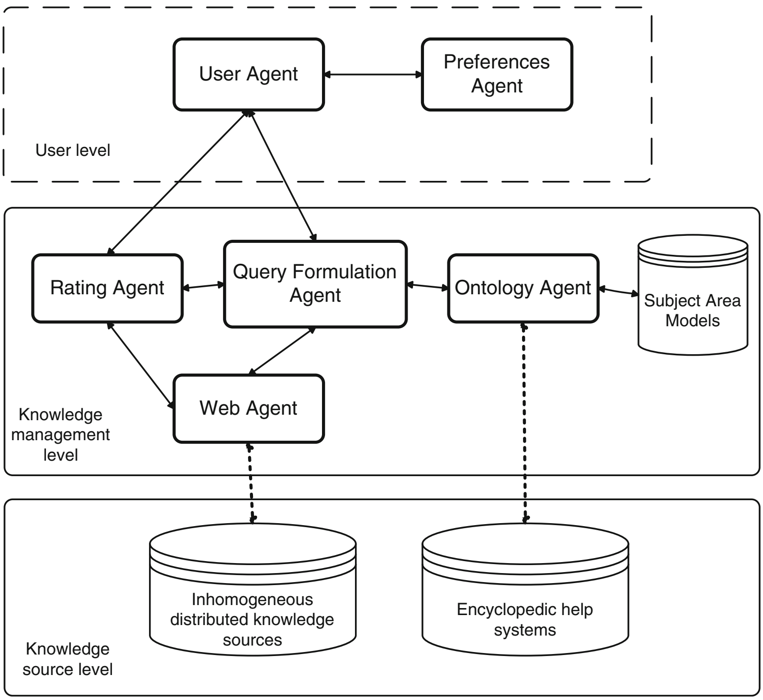


Fig. 1 Multilevel agent model of the knowledge filter

data. Patterns, that agents are making during search queries formulation, improvement, processing and arranging, can be indexed and saved in a repository as precedents. The justification of agent model usage is that agents can be considered as autonomous and preventive subjects.

The main benefits of suggested knowledge filter are:

1. to support the semantic search of relevant knowledge based on ontological models;
2. to use of informational encyclopedic help systems that have different functions for improvement of search query form;
3. to increase of user's query efficiency on the basis of precedents repository usage.

The suggested agent model of the knowledge filter consists of three levels:

- an user level;
- a knowledge management level;
- a knowledge sources level.

Special agents are located on different levels and have clearly defined functions. These functions support interactive creation and processing of query, which includes formulation, modification, ranging and representation of results. The architecture of knowledge filter is open and modular which allows us to define and configure new ontologies, knowledge sources and help systems in appropriate time.

User Agent is responsible for interaction and application of user settings which are controlled by Preferences Agent. These preferences involve term weight which determines its relative value; significance of query results; preferences of different knowledge sources, etc. Preferences Agent can also study the users' preference on the basis of practice and feedback.

Ontological Agent has an access to model images of investigated subject areas with the use of open ontology construction standard—Web Ontology Language (OWL). An encyclopedic help system allows Ontological Agent to use generalizing subject terms and synonyms. There are help systems with expanded functions which enable user to translate the geographical name of a certain place into latitude and longitude co-ordinates needed for some knowledge sources. Additions and specifications can be added in a modular way, which gives an opportunity to refresh the model of concerned subject area, defining new concepts and relations.

Initial query form is sent to Ontologial Agent which obtains needed synonyms after access to the help system. With the use of Query Formulation Agent user is offered to choose one or several synonyms. Ontological Agent receives the solution and updates corresponding attribute in OWL ontology scheme. Depending on the data type of changed attribute a given encyclopedic help system is involved. Obtained information is sent to Query Formulation Agent, which pass reformulated query to Web Agent for processing.

So, while getting to Query Formulation Agent, the initial user's query is improved by Ontological Agent on the basis of semantic processing, with available ontological models of subject areas and encyclopedic help systems. Reformulated query is divided on subqueries intended for appropriate knowledge sources by Query Formulation Agent. This provides semantic terminology mediation used in ontology model of subject area and help system with local sources. Beyond that, query translation is needed to obtain data from suitable inhomogeneous knowledge sources.

The main role of Web-Agent is to obtain the reformulated query after its processing by Ontological Agent and dividing on subqueries by Query Formulation Agent. Web-Agent sends subqueries and adjusts inhomogeneous distributed knowledge sources in terms of the following aspects: user settings of resources; resources authority and reputation; agreement on the conditions of use; sizing of subquery answers; quality measure of network traffic maintenance and resource dynamic work load [6].

Rating Agent is responsible for collecting results of subqueries from different knowledge sources, estimating it by semantic criteria in accordance with Preferences Agent demands. Preferences are defined in terms of authority weight of the using resource (from 0 to 10) and term weight which is involved in query.

3 Meta-Model of the Knowledge Filter

Considering the issue about the pattern of knowledge filter behavior we should define the semantics of through workflow which controls the whole search process, including query specification, query reformulation, query dividing, web-service selection, knowledge source selection, arranging results and providing advices. Abstracting from agent classes, its relations and properties, we created a knowledge filter static meta-model (Fig. 2).

In suggested meta-model each agent manages corresponding object classes, processes specifications and web-services. For instance, a user determines User Preferences, which can be specialized for Search Preferences and Resource Selection Preferences. The user submits a User Query, which has several Query Terms interacting with Concept Ontology one at a time. The Ontology Agent also manages both The User Query and The Concept Ontology, which are made available by Source Ontology. Source Ontology and Source Data are specializations of Source. Source is controlled by Web-Agent.

Users query contains several specified queries, each of them is set out to several data sources, which provides one or more of data elements as query results. On the basis of returned query results the user can use feedback in terms of result relevance and other comments.

Search results affect the development of metadata connected to user preferences, query formulation, data sources usage and results arranging. Knowledge filter meta-model can be implemented as a relational database.

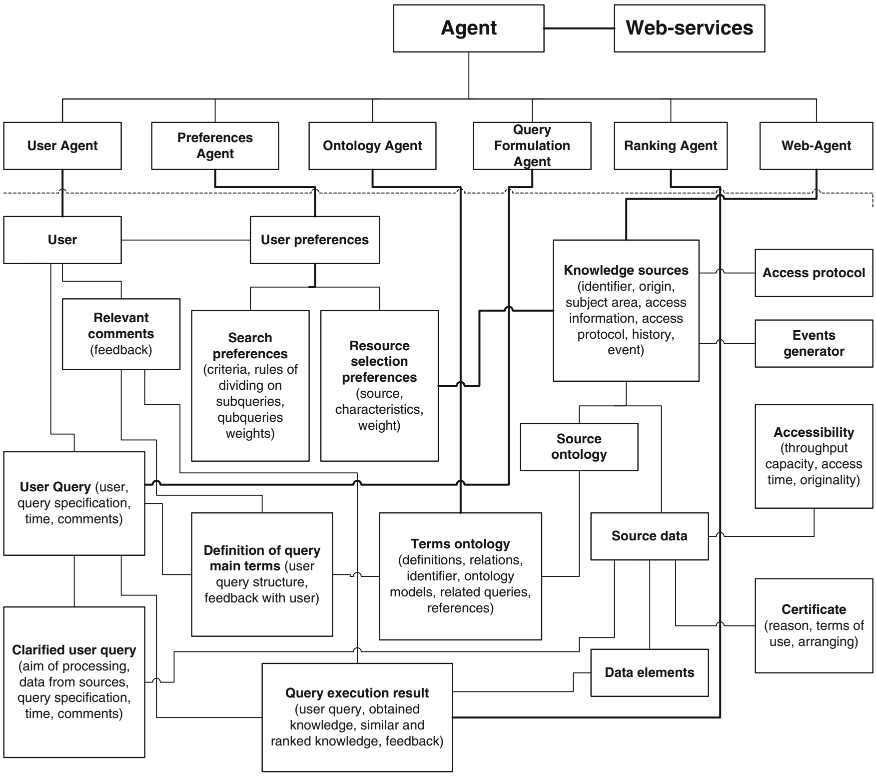


Fig. 2 Meta-model of the knowledge filter

4 Case-Model of the Knowledge Filter

Decision support for search query improvement will be more effective in the case of consideration of precedents (previously saved successful user queries). We suggest a case-model of knowledge filter (Fig. 3), which allows to modify query specification and to improve it through previously saved in repository queries. Query modification is successful only if the user highly evaluates the compliance of search results to initial targets [5, 7].

In this context a Case-agent controls the connection of the User Agent and the Query Formulation Agent with query case-model base and repository of successful user queries precedents. Such collective filtration allows the knowledge filter to choose knowledge sources and to adjust to them dynamically.

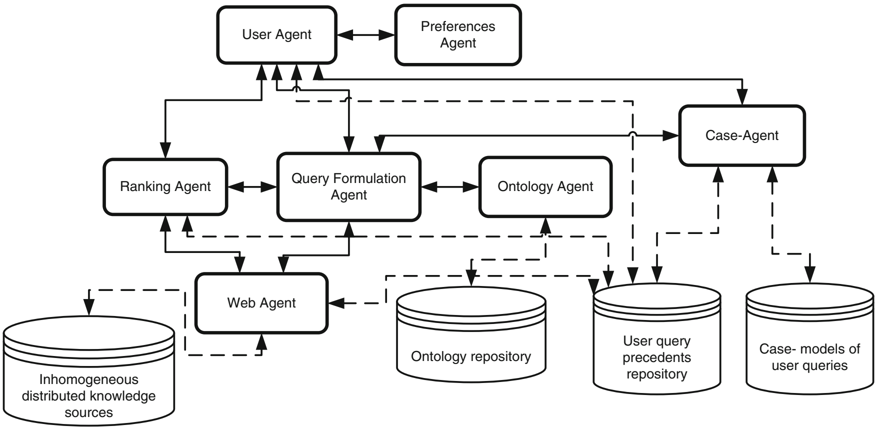


Fig. 3 Case-model of the knowledge filter

5 Experimental Research

We based on the semantic sifter theory to build the knowledge ontology model. It is according to XML, RDF, OWL standard for life cycle-based methodology ontology's creation. We used the ontology language OWL, based on the XML syntax to add semantic sifter for knowledge ontology model by RDF. Construction of knowledge ontology model is to achieve the complex structure relations representation and reasoning between the resource knowledge and topics.

The basic parameters of the experiment: T_i—topic of a resource; V—vector environment; V_1—vector of resource R_1; χ_i—the numbers that the topic T_i entrances in R_1; V_2—vector of resource R2; ψ_i—the numbers that the topic Ti entrances in R2; sim()—the similarity of word characteristics by word meaning distance; simd()—the conception of distance vector similarity of two resources R1 and R2; {A1, A2, …, An}—the conception set of resources R1; {B1, B2, …, Bm}—the conception set of resources R2; dis(Ai, Bj)—the distance between the conceptions R1 and R2 $(i=\overline{1,n};\ j=\overline{1,m})$; RES()—resource similarity computation; x = 0, 7—the coefficient of the conception characteristic vector; y = 0, 3—the coefficient of the conception distance vector.

Algorithm. Resource similarity computing.

Input: Resources R1 and R2. Output: RES(R1, R2) similarity of R1 and R2.

```
Begin
```

4. $sim(R_1, R_2) = \overline{V_1} \cdot \overline{V_2} = \frac{\sum_{i=1}^{n} \chi_i \psi_i}{\sqrt{\sum_{i=1}^{n} \chi_i^2} * \sqrt{\sum_{i=1}^{n} \psi_i^2}}$
5. $sim_d(R_1, R_2) = \frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{m} dis(A_i, B_j)}$
6. $RES(R_1, R_2) = x * sim(R_1, R_2) + y * sim(R_1, R_2)$

```
End
```

Table 1 Knowledge sifter results in agent and keywords models

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7
Group 1	94	90	87	91	97	95	95
Group 2	85	83	81	88	94	81	89

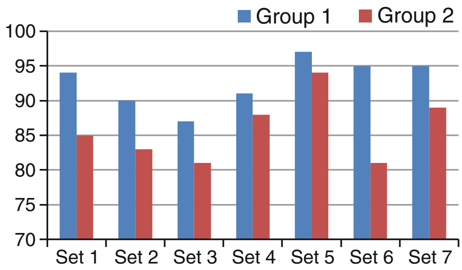


Fig. 4 Evaluation of 100 users in 2 groups

To evaluate the performance of semantic method in knowledge sifter, we carried out some experiments. More than hundred users were divided into two groups. Described method was applied only to the first group. The keywords knowledge sifter is used only in second group. We ask the users to judge the precise of knowledge sifter results and ranking. The ranking scores are in the scope 0–100. The best score is 100. To ensure the accurate of the evaluation, we choose 7 different knowledge sets for the 2 groups. Results are shown in Table 1.

As can be seen in Fig. 4, from the extensive user studies, using more than 100 comparisons, we demonstrate preference for the knowledge sifter results generated by our agent method, compared to a large number of existing keywords models knowledge extraction.

In the described algorithm, the requests' ordered queue is always sorted in ascending order according to reachable distances. Therefore, the algorithm always selects to deal with the points with the smallest distance.

6 Conclusion

Knowledge filtration assumes the usage of knowledge filter agent model, which coordinates knowledge search in inheterogeneous distributed sources, containing unstructured or semi-structured data. Patterns that agents are making during search queries formulation, improvement, processing and arranging can be indexed and saved in a repository as precedents. The justification of agent model usage is that agents can be considered as the autonomous and preventive subjects.

The main benefits of the developed model are: to support the semantic search of relevant knowledge based on ontological models; to use of informational encyclopedic help systems that have different functions for improvement of search query form; to increase of user query efficiency on the basis of precedents repository usage.

To define semantics of the through workflow, which controls the whole search process, including query specification, query reformulation, query dividing, web-service selection, knowledge source selection, arranging results and providing advices, we developed the knowledge filter static meta-model. Each agent of suggested meta-model controls corresponding objects classes, processes specifications and web-services.

In order to optimize search query we developed a knowledge filter case-model, which allows to modify query specification and to improve it through previously saved in repository queries. Concrete scientific result is solving the problem of semantic identification of the key information and processing the heterogeneous knowledge resources.

Acknowledgments The study was performed by the grant from the Russian Science Foundation (project # 14-11-00242) in the Southern Federal University.

References

1. Amerland, D.: Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and Amplifies Your Online Presence. D.Amerland. Que Publishing, 230 p (2013)
2. Bova, V.V., Kravchenko, Y.A., Kureichik, V.V.: Decision support systems for knowledge management. Software Engineering in Intelligent Systems. In: Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015), vol. 3, pp. 123–130. Springer International Publishing AG, Switzerland (2015)
3. Bova, V.V., Kravchenko, Y.A., Kureichik, V.V.: Development of distributed information systems: ontological approach. Software Engineering in Intelligent Systems. In: Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015), vol. 3, pp 113–122. Springer International Publishing AG, Switzerland (2015)
4. Eberhart, R.C., Shi, Y.: Particle swarm optimization: developments, applications and resources. In: Proceedings of the IEEE Congress on Evolutionary Computation (CEC), Seoul, Korea (2001)
5. Gangeni, A.: An Overview of the ONIONS Project: Applying Ontologies to the Integration of Medical Terminologies. In: Gangeni, A., Pisanelli, D.M., Steve, G. (eds.) Data and Knowledge Engineering, vol. 31. pp. 183–220 (1999)
6. He, Q., Zhao, X.R., Luo, P., Shi, Z.Z.: Combination methodologies of multi-agent hyper surface classifiers: design and implementation issues Second international workshop. AIS-ADM 2007. Proceedings, pp. 100–113. Springer, Berlin, Heidelberg, (2007)
7. Hu, X., Shi, Y., Eberhart, R.C.: Recent Advances in Particle Swarm. In: Proceedings of Congress on evolutionary Computation (CEC), Portland, Oregon, pp. 90–97 (2004)
8. Kerschberg, L., Jeong, H., Kim, W.: Emergent Semantic in Knowledge Sifter: An Evolutionary Search Agent based on Semantic Web Services. In: Spaccapietra, S., Aberer, K., Cudre-Mauroux, P. (eds.) Journal on Data Semantic VI. LNCS, vol. 4090, pp. 187–209. Springer, Heidelberg (2006)
9. Kravchenko, Y.A., Kureichik, V.V.: Knowledge management based on multi-agent simulation in informational systems. Conference proceedings. In: 8th IEEE International Conference "Application of Information and Communication Technologies, AICT 2014", pp 264–26715–17 Oct 2014, Astana, Kazakhstan

10. Shi, Y., Eberhart, R.C.: A modified particle swarm optimizer. Proceedings of the IEEE Congress on Evolutionary Computation (CEC), pp. 69–73. Piscataway, NJ (1998)
11. Sousa, T., Silva, A., Neves, A.: Particle Swarm based Data Mining Algorithms for classification tasks. Parallel Comput. **30**(5–6), 767–783 (2004)

A New Approach for Software Development in Terms of Problem-Oriented Knowledge Search and Processing

Andrey Lezhebokov, Bogdan Shkalenko and Elmar Kuliev

Abstract The paper deals with a development of a new approach of knowledge representation, which can be used in design the software to support of search and processing of problem-oriented data and knowledge. The authors formulate fundamental principles of a object-oriented knowledge repository building. Also there is a software architecture on the basis of object approach and object-oriented knowledge base, which describe knowledge in a subject area with maximum precision. The paper describes current criteria of estimation modern software development systems, presents the analytical review of existing programming aids and techniques. To design modern easy-to-use software environment the authors determine the most appropriate development tools. On the basis of chosen development tools we implement a cross-platform software and describe users' interface. Further research is directed to determine the possibilities of high-level flexible frameworks to create modern cross-platform programming solutions.

Keywords Software · Information system · Knowledge search · Problem-oriented knowledge · Object approach · Database · Knowledge base

1 Introduction

At the present time a lot of companies and agencies rapid accumulate huge amounts of data (BigData) in different fields. There are plenty of BigData sources in modern world. For example, time-line data from measuring devices, events from radiofrequency identifier, social network message flows, meteorological data, rEarth Observation (EO) constellation, data flows from mobile cellular about subscribers

A. Lezhebokov · B. Shkalenko · E. Kuliev (✉)
Southern Federal University, Rostov-on-Don, Russia
e-mail: elmar_2005@mail.ru

A. Lezhebokov
e-mail: legebokov@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 450, DOI 10.1007/978-3-319-33609-1_27

location, devices of audio and video registration [1, 2]. Mass distribution of technologies and new models of different devices and internet-services became a starting point for BigData entry into all spheres of human life. Primarily, there are scientific research, business sector and government.

Accumulated information increased so much that there is need to develop modern approaches, principles, technological and practical tools for its processing and analysis. Stored data gives an opportunity to extract new knowledge. With the use of this knowledge we can obtain analytical information and control the companies' development strategy, predict market trends and find new optimal solutions, which provide a successful development in a competitive climate. In modern world such information analysis is essential part of daily activity for many organizations [3, 4].

Today, an optimal data storage problem is the most promising. It allows to extract critical information from accumulated data most effectively. This paper studies the representation method and data base management system (DBMS) capable to approximate the data storage structure to the real subject area.

2 A New Approach for Problem-Oriented Knowledge Representation

The development of modern Data Mining technologies makes claims on data repositories, which popular relational models and systems cannot respond. For instance, one of recent claims is multiple reorganization of data model during information requests [5].

The structure of obtained problem-oriented knowledge is reflected in data model used in the chosen physical storage. Usually the model reflects structures of different application spheres or subject areas and is a fundamental model for describing the suggested knowledge representation method. In general, the data model is determined by the complex of elements descriptions (entities), connection (relations) and the aggregate of these elements properties and behavior descriptions. The term "data structure" is commonly connected to its external representation. It is very important to use data structure that corresponds to the object structure for optimization of input data transformation and for its further processing effectively [6].

The suggested knowledge representation method involves static model representation and dynamic successive approximation to the user's aim. An important difference of suggested method is built-in tools of controlling the data models consistency, constructed within it during the storing and processing (Fig. 1).

The subject area has a great importance for knowledge structuring. In real world an object or a group of objects can be represented by other dissimilar objects. Usually such objects combine into a single hierarchical structure. Complex objects are implemented as constructions with simplified objects. Thus, the concept of hierarchy is essential for representation of structural relations in objects. Often hierarchy is a multilayer structure, where structural elements of the object are

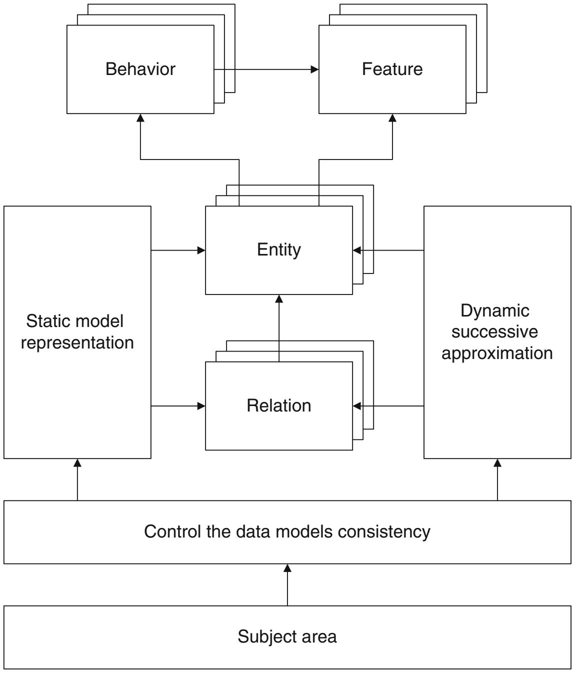


Fig. 1 Knowledge representation based on object approach

constructions composed of simpler entities. In suggested knowledge representation method data hierarchical structure reflects abstract structural relations that correspond to objects being in hierarchical relations in the real world. So, a systematics is assigned in data structure, which determines the semantics, and knowledge processing software must have capabilities for these semantic relations control [7].

So, apparently, new means and methodologies for knowledge database structuring need to have objective orientation related to object selection. To create applications, reflecting subject area objects, close to real ones, it is need to research in post-relational direction.

The studied "object" philosophy is mostly popular now, because the term "object" in any object-oriented system means the combination of data and software, which represent some entity (description) of the real world. Each "object" data are composed of components of random type called attributes. Each program that describes the "object" behavior and changes its attributes is called a method. Users can not always see what the object has inside, but they can use it addressing to its program component which is assigned by abstract interface.

Recent studies results show that the most popular relational technology is not suitable for the work with complex objects. Let consider the example of the work with a complex object. This example is described in different literature sources in terms of object-oriented programming: "Hardly anyone comes with an idea to disassemble a car (complex object) into parts (entities) before placing it in the garage (database), rearrange them and collect it backwards the next morning".

Object-oriented approach as a knowledge representation method gives an opportunity to work with complex structured data and to effective overcome constraints connected to relational DBMS technologies. In relational databases hierarchical data is represented as many-to-many relation tuples. During extraction of data with attributes the relational database must perform "expensive" connecting operations i.e. to spend computation resources. Hierarchical data in object-oriented knowledge bases are represented, stored and processed by natural way due to the fact that attribute value can be an object.

3 Principles of Problem-Oriented Knowledge Extraction and Structuring for Object-Oriented Knowledge Database

As previously noted, the development of object-oriented knowledge database controlling systems is caused by increasing amount of stored and processed information and with the need to satisfy requirements of applications, that differ from data processing applications, which are typical for relational database systems. Attempts to use relational database technologies in such complex applications as computer aided design (CAD) systems; computer aided manufacturing (CAM); programming technologies; knowledge-based systems and multimedia systems have met a set of constraints of relational database systems. Under conditions of appearance of new database applications, new demands are occurred, which are satisfied with the use of object-oriented databases (OODB) [7–9].

The reason of object-oriented database system appearance was a need for more suitable real world entities representation and modeling, because OODB provides much more advanced data model, than traditional relational databases. The paradigm of OODB is based on fundamental concepts, like objects, identifiability, classes, inheritance, overloading and late binding [7, 10].

In object-oriented data model each real world entity is represented by one term—an object. The object has condition and behavior. Object condition is determined by its properties values—attributes. Properties values may be primitive (such as strings or integers) and non-primitive. Non-primitive ones involve a set of properties. Thus, objects may be defined in terms of other objects recursively. Object behavior is determined by methods that control object condition. Each object has a unique identifier which is defined by the system. Objects that have the same properties and behavior are grouped in classes. The object can be an instance of a single class or several classes.

The majority of OODB implementations are based on following knowledge extraction and structuring principles (Fig. 2).

Abstraction. Each real "thing" in database is a class member. A class is defined as an aggregate of properties, methods, public and private data structures and

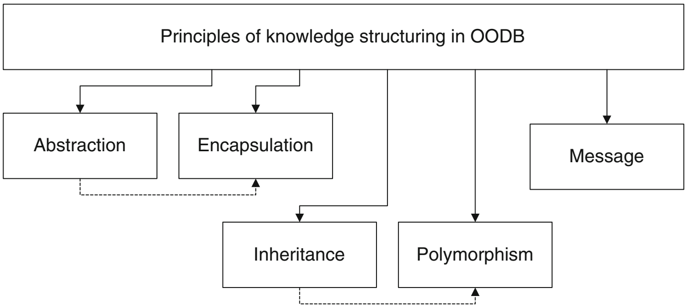


Fig. 2 Principles of knowledge structuring in OODB

managing program applied to objects (instances) of a given class. Classes are essentially abstract data types. Methods are procedures called to produce actions with the object (for instance, to print itself or to copy itself). Properties are data values related to each class object characterizing it in some way (for example, color or age). Properties do not occur in each class realizations, they are just short entries without parameters.

Encapsulation. Internal representation of data and details of public and private methods (functions or programs) is a part of a class definition and is known only inside this class. Access to class objects is permitted through properties or methods of this class or its "parents", rather than by the use of internal realization specialities knowledge.

Inheritance (may be single or multiple). Each class is defined as a part of class hierarchy. The definition of a lower-level class inherits properties and methods of its "parent", unless they are declared non-heritable fairly or was modified by a new definition. In the context of single inheritance a class can have only one parent class, i.e. class hierarchy has tree-like structure. In the context of multiple inheritance a class can derive from multiple parents, i.e. class hierarchy has a structure of direct non-cyclic graph, unnecessarily tree-like. Not all modern object-oriented DBMS supports multiple inheritance.

Polymorphism. Several classes can have the same names of methods and properties, even if they are considered as different. This allows us to write access methods which will properly work with objects of different classes, if only corresponding methods and properties are identically defined in these classes. For example, the "Print" method can be defined in many classes, but do different work depending on object class it is applied to.

Message. The interaction between objects is realized through the message sending and receiving responds. This is different from procedure calling for other models. To apply a method to an object we need to send a following message: "apply this method". The paradigm of the message sending is not always used in object-oriented databases due to its complexity.

4 The Software Architecture

As a result of conducted studies we developed the software for problem-oriented knowledge search and processing on the basis of a new approach of problem-oriented knowledge representation and principles of object-oriented data storages building. Information and logical model of software environment shows the subject area data as an aggregate of information objects and relations between them (Fig. 3). The input data in top level contains mobile visualization device, physical data storage, subject area, requirement to knowledge representation results.

Creation of informational materials, 3D models and benchmarks is performed by following users' categories: programming engineer, subject area expert and designer-visualizer. Form technical functioning point of view, components needed for system work are: mobile operation systems for the application launch; free 3D models editor called Blender and benchmarks needed for control block creation. The output information for the software is results of knowledge search and processing—new problem-oriented knowledge.

The software environment architecture (Fig. 4) includes following modules: Access Management, User Management, Back-Up Master, Mark Recognition, 3D Models, Tasks Generator, Mark Calculation and Control Block. Administrator, expert and user interact with informational tool environment through corresponding graphic interfaces.

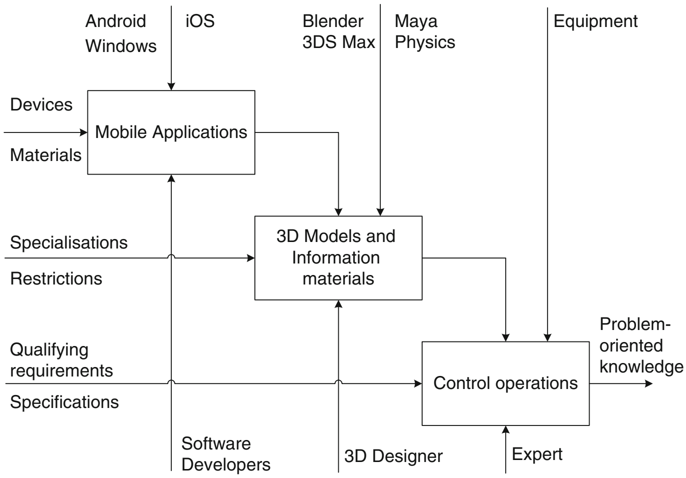
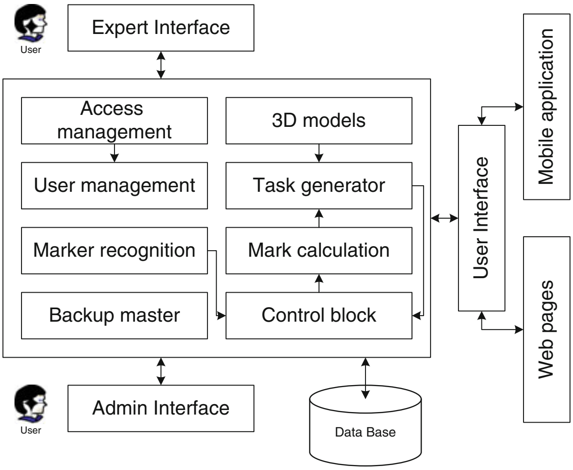


Fig. 3 Figure 3 Decomposition of information and logical model

Fig. 4 Software architecture



5 The Developed Software

The main criterion to choose a program feature implementation is a cross-platform support. This criterion exactly represents general tendencies in the development of applicable problem-oriented software. The authors chose Unity design environment. It is a design tool for 2D and 3D applications and games running Windows and OS X. Applications created using Unity can be run Windows, OS X, Windows Phone, Android, Apple iOS, Linux and Wii, PlayStation 3, PlayStation 4, Xbox 360, Xbox One. This program package allows to design applications for running in browsers with the use of special Unity plug-in (Unity Web Player) and the WebGL technology implementation. Earlier there was experimental project realization support in terms of Adobe Flash module, but later the Unity design team refused it.

Applications designed with the use of Unity support the DirectX and the OpenGL technologies that are used for the visualization of 3D interactive scenes and models. This software is used by major developers (Blizzard, EA, Ubisoft), as well as by individual developers (small project team) due to its free version, user-friendly interface and the ease of use.

The Unity has a simple Drag&Drop interface (Fig. 5), which is easy to adjust, composes of various windows that allows debugging applications in the editor. The software supports three script languages: C#, JavaScript (modified), Boo (Python dialect). The editor supports DirectX 11 and HDR. Physics calculations are performed by physics engine PhysX by NVIDIA.

The example of the software implementation is a mobile applications complex and a cloud server platform (knowledge and data storage), which allows to solve local knowledge management problems in education field (education and academic performance rating). The authorized user is asked to choose a section (subsection) and a theme (information unit) for knowledge search. If expert is not going to add a

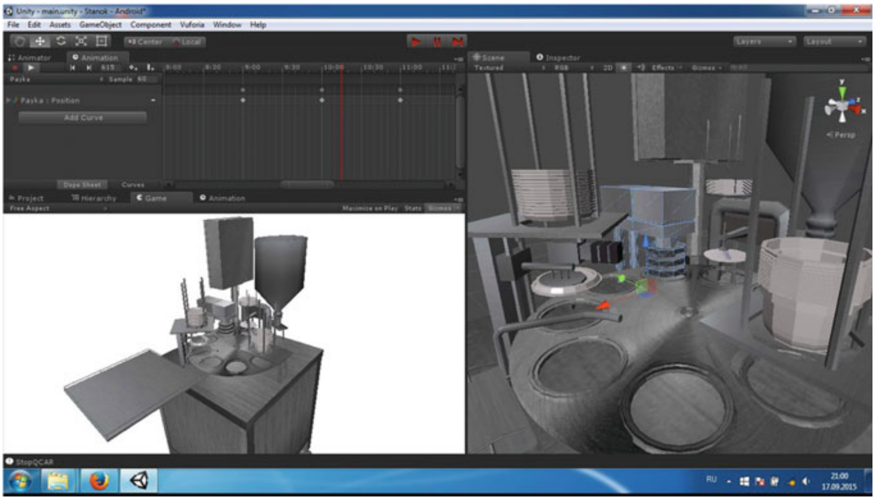


Fig. 5 Unity interface

new data of subject area (education content, 3D models, tests), the chosen section can opened in rendering view. While viewing, the user may edit the embedded content, and then the refreshed content is accessible to review. Then expert can shut down the application or open other sections for search and visualization.

The developed software is implemented for mobile devices. It is intended for practical lessons based on mobile applications in the education environment. The environment contains an augmented reality module permitting to increase the visualization level of studied objects and materials. The developed solution supports different user categories: the expert, the consumer (client) and the administrator. The environment gives an opportunity to improve the quality of problem-oriented knowledge representation in the current teaching content, to organize knowledge search and processing at time suitable for user through a single cloud storage and local client mobile interfaces. The software functioning with mobile devices excludes user reference of time and place of problem-oriented knowledge search and processing.

6 Conclusion

Considered principles of problem-oriented knowledge extraction and structuring are fundamental for computer-aided engineering of knowledge databases. On the basis of these principles we can develop distributed information system that have a set of benefits. For example, it is the maximum proximity to the subject area data structure, that allows to increase the efficiency of information search and extracting requests executing for its further processing.

The benefits of the suggested object-oriented approach for problem-oriented knowledge representation are apparent, but relational technology has important advantages—theoretical validity and reinforcements by standards. The majority of object-oriented databases have a lack of requests executing aids. Object-oriented knowledge databases permit now to create flexible database schemes and to fill it with objects, but they cannot give us powerful tools for extracting objects from database for shared access with other users.

The developed software environment for problem-oriented knowledge search and processing in the form of mobile application complex and cloud server platform gives an opportunity to solve the local knowledge management problems effectively. The environment has shown its high reliability and efficiency while designing informational subsystems for education processes support.

In future we intend to develop new principles and approaches for creation of distributed information systems of access to problem-oriented data and knowledge and object-oriented represented knowledge.

Acknowledgments The study was performed by the grant from the Russian Science Foundation (project # 14-11-00242) in the Southern Federal University.

References

1. Gladkov, L.A., Gladkova, N.V., Legebokov, A.A.: Organization of knowledge management based on hybrid intelligent methods. In: Advances in Intelligent Systems and Computing, vol. 349, pp. 107–112 (2015)
2. Wu, G., Pedrycz, W., Ma, M., Li, H., Liu, J.: A particle swarm optimization variant with an inner variable learning strategy. In: The Scientific World Journal, vol. 2014, Article number 713490 (2014)
3. Dukkardt, A.N., Lezhebokov, A.A., Zaporozhets, D.: Informational system to support the design process of complex equipment based on the mechanism of manipulation and management for three-dimensional objects models. In: Advances in Intelligent Systems and Computing, vol. 347, pp. 59–66 (2015)
4. Kureichik, V.V., Kravchenko,Y.A., Bova, V.V.: Decision Support Systems for Knowledge Management, Software Engineering. In: Intelligent Systems Advances in Intelligent Systems and Computing. vol. 349, pp. 123–130 (2015)
5. Kravchenko, Y.A., Kureichik,V.V., Zaporozhets, D.Y., Zaruba, D.V.: Information and knowledge integration based on simulation modeling. In: 9th IEEE International Conference on Application of Information and Communication Technologies, AICT 2015. pp. 22–25
6. Popov, L.M., Ustin, P.N.: Formation of psychological readiness for higher educational institutions student professional activity. Soc Sci (Pakistan) **10**(4), 418–420 (2015)
7. Gao, Q., Wang, D., Gao, F.: Impact of knowledge representations on problem-oriented learning in online environments. Int J Human-Comput Interact. **31**(12), 922–938 (2015)
8. Haitzer, T., Zdun, U.: Semi-automated architectural abstraction specifications for supporting software evolution. In: Science of Computer Programming 90 (PART B), pp. 135–160 (2014)
9. Bova,V.V., Kureichik, V.V., Lezhebokov, A.A.: The integrated model of representation of problem-oriented knowledge in information systems. In: 8th IEEE International Conference on

Application of Information and Communication Technologies, AICT 2014—Conference Proceedings, pp. 111–115 (2014)

10. Lezhebokov, A.A., Kravchenko, Y.A., Bova, V.V.: Support system for QR-code-based educational processes. In: 8th IEEE International Conference on Application of Information and Communication Technologies, AICT 2014—Conference Proceedings, pp. 482–486 (2014)

The Concept of Construction Methodology Notion for Intelligent Systems

Yuri Rogozov

Abstract In the paper discussed the concept of constructing a methodology notion in sign form, which is a schematization in form of action mechanism of normalization of system creation rules. The proposed idea is based on the following principles: the concept of the methodology is a system that represents a set of rules for creating interpenetrating particular paradigm creation systems. Rules are represented as cognitive action mechanism for generating a specific cognitive means acts to create an object symbolic form. Cognitive action mechanisms can be objectified to the properties (attributes) of the object being created and presented in the form of a specific mechanism of action. The methodology as a system is formed from a system of methods of organizing mechanisms of action, forming a common mechanism of cognitive activity (objective and methodological knowledge) and the logic of motion content (via attributes) mechanisms of action.

Keywords Methodology · System presentation · Knowledge

1 Introduction

Design and construction of objects of different types have become a mass profession in modern conditions. To provide a standard format for the result presentation it require a detailed regimentation of developers' labor at different levels. Both of these circumstances decisively stimulate research development in the field of methodology as "in depth", i.e. in the direction of more detailed disclosure of basic principles and forms for its structure formation and as "broad wise"—in the direction of a rigorous and specialized design of tools for scientific cognition. On the base of proper per se thesis that the true knowledge can be obtained only by true

Y. Rogozov (✉)
Department of System Analysis and Telecommunications,
Southern Federal University, 44 Nekrasovsky St., Taganrog, Russian Federation
e-mail: yrogozov@sfedu.ru; rogozov@tsure.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 450, DOI 10.1007/978-3-319-33609-1_28

method, it is necessary to learn how to create such a methodology that would allow to get desired result for a specific task in a short time. We proceed from the fact that the task is to develop such an understanding of methodology, which should provide the obtaining of specific paradigm and fundamental rules of systems construction theories. To do this, hold a brief analysis of existing methodology definitions and find out what functions it should serve to create a paradigm.

2 Problem Statement

The methodology has long been regarded as the doctrine of systematization of knowledge about the modus operandi. This definition includes in structure a plurality of knowledge on methods, approaches, techniques, etc., that a person can use in all its activities. The aim of methodology is also to develop rules of implementation of actions and their conversion: “The methodology of pedagogy—is the study of pedagogical knowledge and the process of getting it that is the pedagogical cognition. It includes:

1. the doctrine of structure and function of pedagogical knowledge, including educational issues;
2. original, key, fundamental, philosophical, general scientific and pedagogical position (theories, concepts, hypotheses), with methodological sense;
3. the doctrine of pedagogical knowledge techniques (methodology in the narrow sense of the word)”.

Methodology is:

1. a set of research methods used in any science;
2. the doctrine of the method of cognition and transformation of the world [1].

From the above definitions, it follows that the methodology is the study of structure of rules system, of properties of formation or obtaining specific rules (norm) the creation of structures, logical organization (interconnection, interoperability) rules for cognitive activity, which results in the creation of rules to build process models. Using such definition of the methodology for a specific task is very difficult. It is because of the search from a huge field of knowledge methodology concrete rules needed for this particular task. Moreover, it is often required not only well-known knowledge, but new special systems of rules that should be created for a particular purpose. These rules must have the properties required for the task (goal). Therefore, the use of the methodology, in the sense that it has nowadays, presents the following problems:

- Uncertainty of the methodology adaptation process in general to solve specific problems (i.e., the processes of adaptation and approximation is performed out of the methodological field of rules);

- Complexity of the adaptation process. The process of approximation is the "cognitive process", the implementation of which requires skill and time, such as problem-solving method of Optner [2]. The methodology is structured weakly (according to the "general" purpose) but is applied to concrete results;
- Inadequacy of the real problem and methodology, as means of knowledge search and development of rules to solve this task. It is necessary to set the same goals as for the concrete methodology construction and for the real object and create an appropriate methodology for a particular task as a set of interrelated specific rules.

Academicians A. M. Novikov and D. A. Novikov made the step to methodology approximation to solve a particular problem. Thye defined methodology as the organization of concrete activity [1]. By creating a methodology for specific activities—games, science, and so on we can reduce the gap between the object and the means of cognitive activity. To some extent, the proposed by A. M. Novikov concept generalizes the known concepts: "If we consider methodology as the study of the activity organization, then, of course, we need to consider the concept of "organization". In accordance with the definition given in [3], the organization is:

3. Internal order, interaction consistency of more or less differentiated and autonomous parts of a whole, due to its structure;
4. A set of processes or actions, leading to the formation and improvement of relationships between parts of a whole;
5. Association of people sharing a particular purpose, and acting on the base of certain rules and procedures. … The methodology considers the organization of the activities (activity—is a purposeful human activity)" [1].

The last definition say that concrete methodology is a structured system of concrete rules for binding differentiated and autonomous parts of a whole and of processes of rules binding by which we are build a model out of the parts of the final result. Moreover, the structure of form of rules representation is constant, but the content is determined by the purpose—the type of activity.

However, this "approximation" does not solve the basic problem—the development of a specific methodology as a system and tool by means of which we must form, structure and logically link entities. Systems of rules should be transformed into a means of knowledge (methods, etc.) for each specific subject of study.

In order to obtain specific methodology for a specific task, the general methodology structure has to be dynamic and able to be configured for a specific task (target) and specific processes. The methodology in general should be a "form of forms" for creation of specific methodologies.

Such representation of the methodology form in general, as "form of forms" or as a system of "implicit" rules is consonant with the concept given by Bela Banath: "In the most general sense, the system is a configuration of parts that are associated and connected together (to each other) through a network of relations" [4].

The methodology is the study of organization rules of the creating forms process, of rules formation or creation properties, rules structure and logical organization (relationship and interaction) for the cognitive activity. Organization is a process of relation formation or a form of the cognitive process form interrelated rules. Then it turns out that the organization is an action which results in the need to have a system of interrelated processes. Methodology as a tool should allow to obtain the structure of logical organization of related (interacting) specific methods or processes, with the help of which we can organize cognitive activity of a particular object of study. The methodology is a tool for obtaining rules of interconnected system processes for a particular study, the organization of these rules for constructing a process model of creating an object.

The methodology in a broad sense (general methodology) must be converted to a particular methodology as a set of agreed rules of the organization process of building a model of a particular object. Methodology in general is an organization form of universal system (for the entire class of rules systems), which is converted to a specific methodology during the creation a particular system, for example information system. In this case, the structure and content of general methodology of systems must be configured with respect to target into specific methodology for the construction of information systems.

Methodology "in general" is independent and has a certain structure, in which rules are not defined and is an interconnected dynamic system of relationships of meaning. Relationships of meaning do not have objectivity and consist of the operations and procedures of selection or choice. In broad terms, the methodology in general is irrespective to subject of study. If there is the goal, the methodology is structured. The goal as usual structures the methodology, reduces the set of rules and methods. It forms the structure of a particular methodology out of methodology "in general" to create a specific object of study. Widely formed methodology structure (in general) does not correspond to a specific purpose.

If to form and structure rules, adequate to research objectives, they will interact and create whole integrated systems of knowledge, which in the literature are called system of rules [5]. With all the uncertainty of structure content of methodology "in general" it must have a certain form. What requirements should satisfy this form?

3 Proposed Approach

The selection procedure itself is the procedure of understanding. Then the methodology can be represented in the form of non-objective forms of meaning understanding procedure. Methodology form as a meta-system, based on a system approach, is integrate metasystem of relationship, consisting of a plurality of control (configurable) dynamic mechanisms of actions (parts) of the cognitive processes that we need to learn how to build and connect with the help of the dynamic relationship between them formed by the rules of meanings understanding.

If the priority is to "construct a theory of "system object"", when the first step is usually the task of the synthesis of the various knowledge systems, i.e., multidisciplinary research task, which eventually should give some whole theoretical framework, that "removes" these particular (relative to the global problem) subjects and their results [6, p. 149]. V.N. Borisov proposes to solve this problem by constructing a "transition" or the motion from a particular subject knowledge of one domain to the defining of subject knowledge of another field [7, p. 5]. The transition from one sign form to another, as specifies G.P. Schedrovitsky, should be carried out with the help of other knowledge [8], for example, meaning coordination and understanding of knowledge.

Knowledge can be obtained only as a result of action performing only by specific rules. But then the question arises—what symbolic forms of system rule representation corresponds to this thesis, what existing symbolic form, presented in the form of work on rules creation, how to show generation of new rule with the help of other rule. And if that statement is true, then we introduce the term "symbolic form of system rule" (SFSR)—it is a motion (activity, vehicle of knowledge) that depicts in a symbolic form the process of change and getting (generation) new rule with the help of other rule. It is needed to answer the question —with the help of which specific cognitive means (activities, works, vehicle of knowledge), consisting of system rules, you can obtain the required definitions of system approach and what should be the sign form—SFSR?

Let's proceed from the fact that any sign form of system concept definition should reflect the specifics of the system approach, state of the system, i.e., should have the most important property of system concepts—to change, constantly evolve, be not a noun, but a verb [9] and should be presented as a whole organization rule of methods changing mechanism of constructing the system.

But before moving to rules (for example, principles), and process of forming the worldview as the ideal image of process of creating sign forms of system concepts, let's define what we shall understood under these concepts, i.e., how we get what we should get. Philosophical encyclopedia provides the following definition of concept given by Schelling [10]: "The concept as action, as activity is the opposite of perceivable, but the concept (which coincides with the self-conscious, with "I"), on its own creates such opposite thing… Concept is objectified not in the same language form, but also in the creation of man, in action, in which they are objectified." This definition should be interpreted so that the system concept as knowledge is a whole, consisting of two interpenetrating activities of knowledge—knowledge of the action (note that concepts are not divided into parts, they are objectified) and knowledge of the concept of "action". Together they form a "matrioshka": knowledge of action generates a certain knowledge of the concept of "action", knowledge of the concept of "action" is contained in the knowledge of the action. Changing the content of action knowledge will change (objectify) the knowledge definition of the concept of "action" (quality). To objectify the concept it is needed to learn how to build these cognitive actions and change their contents. Knowledge of action is a rule of action over rules.

Definition of concept as a rule should explain the process of formation and changing rule with the help of other rule. This other rule must be presented in the sign form of knowledge machine and this machine is to be built on the basis of knowledge change process—evolutionary worldview.

Thus, any systemic definition of concept, as symbolic form, is generating activity as knowledge machine. This is the most important specific component of systemic approach—its symbolic form should be presented in the form of knowledge machine, that generates the systemic representation of object's properties changes. Building the machine of rules procreation is the key problem of system approach, the solution of which would solve all its problems. The essence of the proposed new ideology is that the system approach will study not objects, but techniques and methods, as construction and changing activities of the object. Object will be contained inside these rules: "what" should be obtained must be in "how" we will receive it, in the tool of rules creation. More detailed—it should be in the way as the activity producing rules and this process activity as a form of knowledge of the rules, to be presented in symbolic form.

The first and most important requirement is caused by the multidisciplinary system knowledge. How to solve this problem, how in the same symbolic form, on the one hand to take into account the different properties of the domain object, and on the other—to abstract from the domain object. It is necessary to invent such SFSR, which would enable to abstract from the subject areas, and in the process of its objectification and actualization, presentation in the form of a specific action, it was filled with knowledge (content) of the research object properties in a given subject area. Building a SFSR as activity or knowledge machine, that is abstracted from the subject areas with the possibility of further updating of subject knowledge to the object properties (attributes), will allow to solve the problem of synthesis of system concepts and then build a general theory of systems. The idea is to find such a rule of symbolic form of the machine of rules procreation, which will allow, firstly, to present it as a symbolic form of activity, secondly, to link the knowledge of the various domains. This activity should be with a higher level of abstraction, such as a cognitive tool consisting of pointless rules create objects with specified properties, not containing specific actions subject knowledge.

To implement the stated ideas it is needed to create a different principle to the construction of symbolic form of system rules as knowledge machine. This principle should organize systematic rules into a completely different new structures and formations. It is necessary, as was noted above, to split structure and process of structures organization. Structure as matrioshka should be within the process of creation (organization) of the object. To perform the schematization into symbolic form of not an object, not the research subject as an act of action of developing the object properties, but knowledge about rules of object creating as a cognitive tool of rules procreation. It is necessary to represent a cognitive tool (method) in the form of rules (normalization) of action performing to obtain knowledge about the object. It is needed to move to the next level of abstraction—represent knowledge as a symbolic form of activity as cognitive means of rules creation. What principles should be used for construction of methodology symbolic form of concrete forms of methodologies?

Besides the symbolic form of reproduction as a single mechanism of rules, activity includes normalization as a single process of creating norms as ways and means of action implementation to create the object. Symbolic form of norm is a rule the action implementation. Consequently, it is needed to build a symbolic form of transition from rules to norm, and then move from one norm to another, and from the latter to the other rules and vice versa. From [7] we know that this transition is a process of cognition. In [7] cognition is defined as form of the motion from one knowledge to another. In fact, the motion is a connection between rules and implementation of interpenetration of processes of direct and inverse transformation of rules. Logic of conversion (motion) must be presented in the certain sign form. Sign Form (SF) is a methodology of motion, which should reflect the logic of the cognition construction as the interpenetration of rules conversion processes. Motion is a set of rules, which are organized in methods of organizing, then any rule must be represented in the SF of organization techniques of action mechanisms, which should generate each other: "Adequate image of "development" of any object, including the "design", necessarily includes two opposite and at the same time interconnected systems. One should depict the process of changing the initially given subject, the other—the mechanism of this changing…" [11]. "We can say that the process is the application of the method in a specific time interval (process step)" [12]. But "In contrast to the nature objects, the action does not exist, it is—carried out" [13]. The cognition motion is carried out by content of the SF [14]. SF reflects knowledge about conversion methodology (conversion method), but its content implements (carries out) logic of rules transformation: "Thus a method of cognition is not external to the cognitive activity of precept, and, in the words of Hegel, a form of self-motion of content knowledge" [7, p. 7]. The content of concrete action's SF defined by the rule or methods of organizing mechanisms of action that have the necessary knowledge to achieve goals.

Knowledge transformation logic is based on the principle of mutual-generation and self-generation, which is the implementation of the principle of transition from general rules to individual and vice versa. Logic of conversion of one rule into another and vice versa is carried out as follows. We take the first rule and present it as a technique of the organization of the first mechanism of action, as shown in [15]. The content of the symbolic forms of organization techniques of action mechanisms is rules that generates first knowledge or meaning. Each of SF's rules of action mechanism of the first knowledge we present in a symbolic form of organization technique of action mechanisms, consisting of target systems of rules. Then, from the obtained target systems of rules we generate the action mechanism of second knowledge, the latter generates the second knowledge itself. Similarly, after going way back from second rule you can get first rule. Under the relationship and interpenetration of rules should be understood the use of the same rules for the construction of various cognitive mechanisms of actions.

Now we proceed to the creation of SF of descripted logic of transformation, which is expressed in symbolic form methodology of the cognition process. We need to build from the transformation logic a symbolic form of the cognitive process. This requires from system knowledge to construct a symbolic form of organization technique of individual cognitive action mechanism, such as it is described in [15, 16]. Then, from the sign form of organization technique of individual cognitive mechanisms of action it is necessary to generate the sign form of motion as a set of interpenetrating sign forms of individual action mechanisms, i.e. from the organization together individual action mechanisms techniques created a method of organizing the movement mechanism. The content of these mechanisms is knowledge, the choice of these knowledge is determined by goal of building the logic of the cognitive process. Knowledge about cognition (rules) interpenetrate into knowledge about the motion (rules of changing) and knowledge about target object properties and their motion.

There is a schematic representation in symbolic form of not action act of obtaining object properties, but of rules (knowledge, normalization) of performing actions for creating an object with the specified properties in the form of activity. In practice, there is a transition to the next level of abstraction—level of modeling methods as cognitive means (rules) of object creating. This makes it possible to solve the most important problem—the problem of multidisciplinary research and relation of multidisciplinary knowledge. It is implemented the interconnection of methods of knowledge organization as means of object creation, but not of objectless knowledge of object properties. Since we are talking about the creation of rules with the help of rules, then as means of knowledge creating it has to be used the rule, presented in the form of the relationship of his party—the rules and methods of organization mechanisms you, complements the action. It is a knowledge of the rules of object creation. We talk about knowledge, about rules of object creation. The creation of symbolic form of normalization will solve the problems of systemic concepts formation. The main difference between the form of organization of rules as a scheme of the mechanism of cognitive action and the act of action is that in cognitive action the methodological knowledge (rule) is used as means (instrument) of action performing, the action itself conveys this knowledge to the object on which it is made, and the result of action as matrioshka is within the rules of action implementation.

To understand the essence of the proposed idea and to distinguish it from the well-known, let's paraphrase the task and allocate within it differences from the known approach. Adequate image of "development" in the form of rules, as a way of creating any object, including the "design", necessarily includes two opposed and, at the same time, interconnected systems.

One system should depict the process of changing the method of creating the original object in the form of rules, the other system should depict the mechanism of this change of rules as a way of creating an object with given properties. There is representation, in symbolic form as activity, not of action of object model, not the research subject as a process of production the object property, but in a symbolic form of system of interrelated rules (knowledge) as cognitive means of object creating. It is almost modeled the cognitive process of matching the methods of object creation.

Let's move to the principles. The main idea of the concept of building the concept of methodology in symbolic form, satisfying the above requirements, is a schematic representation in the form of action mechanism of normalization. The proposed idea can be expressed by the following principles. The first principle is—the system concept is a generating system of rules. The generating system is organized set of interpenetrating activities of rules as ways of creating a system concept of the object. Activity is normalization, provided in the form of cognitive mechanism of action as cognitive means of rules generating for creating a symbolic form of the object. Cognitive tool of rules for creating the object concept can be updated to the properties (attributes) of the object being created, at the same time transforming into an act of activity. Principle characterizes the system concept as a system of combined methods of the organization of cognitive mechanisms of action and knowledge within their content and the content motion. System concept of the methodology will be formed from the system of organization methods of mechanisms of action and knowledge, included in the mechanism of cognitive activities (objective and methodological knowledge) and logic of content motion of (from means to attributes) mechanisms of action.

The basic process of transition from object knowledge to the sign form of methods of organization of cognitive action mechanisms, as the means of their creation, is performed through the second principle—generating activities as ideal model of interpenetration and intergeneration of various rules. The relationship of rules is usage of the same rules for building mechanisms of cognitive actions as generating or generated action mechanisms.

4 Conclusion

The report presents developed methodology framework "in general" as a relation of dynamic procedures of meanings understanding, and mechanisms for structuring the methodology "in general" into the specific methodology for a particular purpose are shown. It was shown that the methodology should be understood as a tool of specific rules (standards) perform actions on the object creation. The methodology is not the norm, but a dynamic process of changing the norm of creating the object, that has a predetermined functionality.

Acknowledgements The reported study was partially supported by RSF, research project No. 16-11-10021.

References

1. Novikov, A.M.: The methodology of education. M.: Egves (2002)
2. Peregoudov, F.I., Tarasenko, F.P.: Basis of System Analysis, 3rd edn. Rev. Tomsk: Publishing house of the NTL (2001)
3. Optner, S.L.: System analysis for business and industry issues. Trans. from English. Soviet Radio, Moscow (1969)
4. Banath, B.: A Taste of Systemics. The Primer Project (1997). http://www.isss.org/taste.html
5. Systematic knowledge and systematic way of thinking. http://www.superinf.ru/view_helpstud.php?id=1499

6. Yudin, G.E.: The Systems Approach and the Principle of Activity, 350 p. M.: Nauka, (1978)
7. Borisov, V.N.: To the definition of cognitive activity method. Problems of methodology of scientific knowledge. Scientific works. Philosophical series, vol. 2, pp. 3–9. Novosibirsk (1968)
8. Schedrovitsky, G.P.: About the difference between original concepts of “formal” and “substantive” logic. Methodology and logic of science. Scientific notes of Tomsk University, № 41. Tomsk (1962). http://www.fondgp.ru/gp/biblio/rus/8
9. Schedrovitsky, G.P.: Knowledge Synthesis: problems and methods. On the Way to the Theory of Scientific Knowledge, pp. 67–109. M. Science (1984). http://www.fondgp.ru/gp/biblio/rus/51
10. Filosofskaya entsiklopediya. http://dic.academic.ru/dic.nsf/enc_philosophy/952
11. Schedrovitsky, G.P.: Categories “process—mechanism” in the context of the development study. Development and implementation of automated systems in the design (theory and methodology). App. III to it. Computer-aided design and objectives of designing activity. Development and implementation of automated systems in the design (theory and methodology) (1975). http://www.fondgp.ru/gp/biblio/rus/40/GP75e.doc
12. Levenchuk, A.I.: Situational Engineering of Methods (2009). http://ailev.livejournal.com/750878.html
13. Dubrovsky, V.Y.: Articles, lectures and performances. http://berezkin.info/?page_id=238
14. Schedrovitsky, G.P.: Philosophy, Science, Methodology. Logic and Methodology of Science, 656 p. M. Kult.Politiki (1997)
15. Rogozov, Y.I.: The approach to determining the metasystem as the system. In: Proceedings of ISA RAS, vol. 64, №. 4. S. 96–115 (2013)
16. Rogozov, Y.I.: Concept of metasystem as a methodological basis for system creation. Ind. ACS Control. № 2, pp. 38–46 (2013)

The Model of Subject-Oriented Storage of Concepts Sense for Configurable Information Systems

S. Kucherov, Y. Rogozov and A. Sviridov

Abstract Existing object concept of systems building where in center stage is put software object that implements the functions assigned to it, leading to a mismatch of the solutions obtained and domain. This applies both to information systems in general and in particular databases. The object approach uses concepts with a concrete sense, which greatly limits the variability of systems, including configurable systems. To solve this problem requires a transition to subject-oriented concept, in center of which is placed the process of obtaining the object. This concept requires working with concepts with implicit sense, which is specified according to the situation in different ways. In turn, the storage in a subject-oriented concept should provide advanced as compared to the database functionality. The paper proposes a model of subject-oriented storage of concepts sense for configurable information systems and the method of its implementation.

Keywords Configurable information system storage · The concept · The concept sense · Data variability · Database

1 Introduction

Consideration of existing systems in terms of their construction elements and principles of design leads to the conclusion that its creation is made with ontology of concepts with concrete sense [1]. Each concept with concrete sense is embodied

S. Kucherov (✉) · Y. Rogozov · A. Sviridov
Department of System Analysis and Telecommunications, Southern Federal University, 44 Nekrasovsky St., Taganrog, Russian Federation
e-mail: skucherov@sfedu.ru; s.a.kucherov@gmail.com

Y. Rogozov
e-mail: yrogozov@sfedu.ru

A. Sviridov
e-mail: asviridov@sfedu.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 450, DOI 10.1007/978-3-319-33609-1_29

in the form of program objects: modules, functions, procedures, libraries, etc., as well as their different compositions. In turn, each implemented software object is associated with the concepts from a subject domain, which user creates or uses in its daily activities. Databases in such systems store concrete senses of concepts (to be precise, the properties of concrete senses) in the form of individual data facts (factual database) [2].

From the viewpoint of the possibility to making changes to existing databases and information systems as a whole, concepts must be defined with an implicit sense, which is to be variability. This is easily illustrated by the following example. For example, there is the concept of the subject domain “gradebook”. Its name defines the type of object, used to account for the marks obtained by the students or pupils. At the same time implementation of this concept, its sense, depending on the situation and the particular educational institution may be quite different. Existing approaches to development of configurable information systems suggest a solution to this problem by standardizing concepts and translate them into separate software objects with knowingly more sense than is needed in a particular situation [3–5]. Such introduction of redundancy and standardization on the one hand, complicates the system, and on the other does not solve the problem of adapting a system for a particular task.

For this reason, construction of databases for configurable information systems and configurable systems themselves, unlike existing classical systems must be based on using as storage elements concepts with concrete and an implicit sense. It is possible due to transition from the object conception (the cornerstone is program object) to the subject-oriented conception (the cornerstone is user work to create a sense of the concept). Means of representing concepts’ sense may be the mechanism of action [6, 7]. It is important to note that both subjective and objective conceptions of building systems basically use the concept of action, but in the object conceptions, this action have specific content, and in the subject—content is implicit.

To solve the task of storing concepts’ sense in the subject-oriented conception of information systems development it is necessary to meet following requirements:

- variable amount of mechanism’s characteristics with supporting replication on structures of all mechanisms available in the storage;
- variable content of mechanisms available in the storage with the ability to use as characteristics of other mechanisms (in subject-oriented conception the communication is also an action and is represented as a mechanism);
- availability of tools for manipulation with mechanisms: modification of structure, content modification, binding, etc.;
- simultaneous storage of concrete and implicit senses for one concept
- availability of data modeling tools which uses terms close to the user.

This tool will be not a database that performs only storing function. It will be a storage that can perform actions on their own content and create new content. This article proposes the approach to construction of subject-oriented storage of concepts sense for configurable information systems. In the first section of the article the objective and subjective conceptions are compared. The second section proposes a model of subject-oriented storage. In a third section method of its implementation in NoSQL database are described. The main conclusions are set out in the conclusion.

2 The Object and Subject-Oriented Concepts of Information Systems and Data Warehouses Development

The vast majority of modern research in the field of data storage are focused on finding solutions for high-performance horizontal scale-out [8]. As well as fundamental view on the storage (database) remains unchanged—the database is represented as electronic card-index, which contains individual facts regardless of the nature of their origin [2]. For this reason, in information systems, including a configurable systems, is still used the classic relational database. Figure 1 shows the design process from the viewpoint of data semantics reducing and its distribution to system components.

The conceptual view of the idea of creation of information systems allows to generate the following picture:

- A potential user until the system deployment performs some work (consisting of individual activities) over a set of domain concepts;

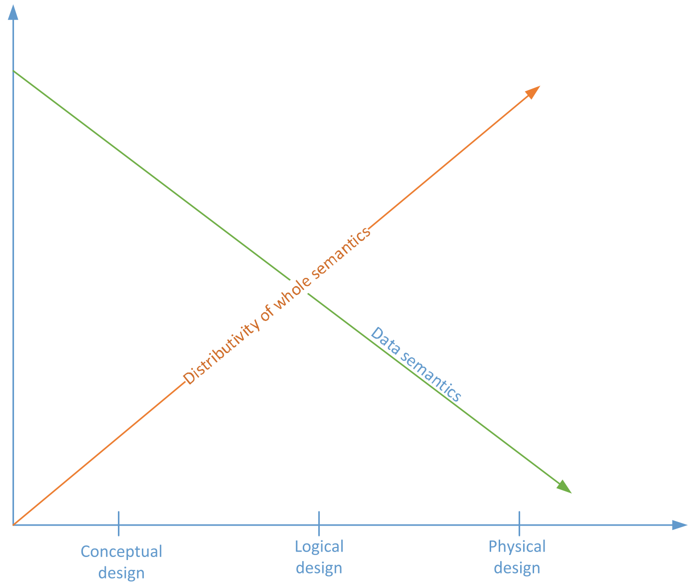


Fig. 1 Semantic reducing in the process of creating the classic database

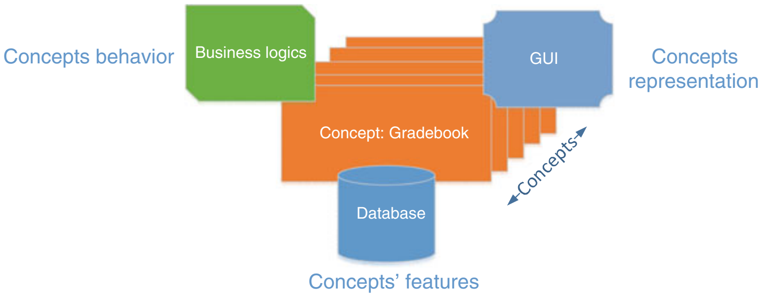


Fig. 2 Presentation of concepts in classical systems

- Each domain concept has a name that allows us to define it, and its sense, which determines an implementation of concept in specific environments;
- By existing techniques, such as requirements engineering, initial sense of user activities and domain concepts with which he operates, are transmitted to third parties (developers) who are create information system.
- Initial sense is converted as many times as the stages takes place in the system's life cycle, beginning from the moment of its receipt to the working prototype.

This corresponds to an object conception of building systems used to date (not object-oriented), in which the cornerstone is placed ready software object [9]. This leads to loss of complete representation by concrete sense of domain concept (see Fig. 2). The reasons of this are next:

- Properties that reflect the sense of concept are placed in a database;
- A visual representation of the sense of concept contained in the user interface;
- The behavioral aspect of concept is reflected in the application business logic;
- Domain specialist does not make reflecting the concept into the system. It is made by technical specialist, often without knowledge of domain specifics;
- Different specialists create each of aspects on which concept sense are shared.

All this leads to a mismatch between the domain concept sense and its realization in the system, and inside the concept because of sharing its parts to system objects. Often this problem is called the problem of misunderstanding between customer and developer [10].

When considering the design process of classic databases it can be clearly noticed that in the transition from domain to implementation occurs gradual reduction of semantics, and the data is actually taken out of context. This can be illustrated by the following expression:

$$tCells \subseteq Entts \subseteq DObj \subseteq Domain \tag{1}$$

where:

- Domain—The set of domain concepts;
- DObj—The set of data objects representing concepts;
- Entts—The set of entities representing the domain;
- tCells—The set of table cells, representing the entity;

Thus, one domain concept, which is operated by an user, is implemented in the system as a set of different types of parts (see Fig. 2). It should be noted that the methods of dealing with this problem now reduced to creating feedback between lifecycle stages and subsequent adjustment decisions in accordance with the original problem.

Another problem lies in the variability of existing systems and databases. If we talk about the possibility of making changes to existing databases and information systems as a whole, the concept should have variable sense. In databases of configurable information systems to obtain a qualitatively new level of variability, transition from the object to subject conception is required. The subjective (subject-oriented) conception, in contrast to the object, models not separate system parts (in the form of concepts with a concrete sense) but works on concept sense production that make up the system (in the form of concepts with an implicit sense).

When storing data within the subject-oriented conception of configurable information systems construction another task arises—the memorizing of the creation process of concept sense structure with the possibility of variable changes its content. Unlike today used factual databases, configurable information system storage should be based on a model, which allows to “remembering” the data acquisition process.

Subject-oriented storage of concepts sense for configurable information systems can be conceptually represented as follows (see Fig. 3).

According to Fig. 3, the basis of the configurable information systems is a subject-oriented storage of concepts sense, which combines the following functions:

- Storing concepts with implicit sense;
- Storing concrete senses (data facts);
- Specifying tools (objectification of concepts);
- Storing concepts with a concrete sense.

In addition, the storage should interact with the external environment through the following instruments:

- Case-tool for the formation of concepts senses, which realizes operation of storage content manipulation;
- Tools for working with concepts with a concrete sense, which realizes functions of graphical user interface.

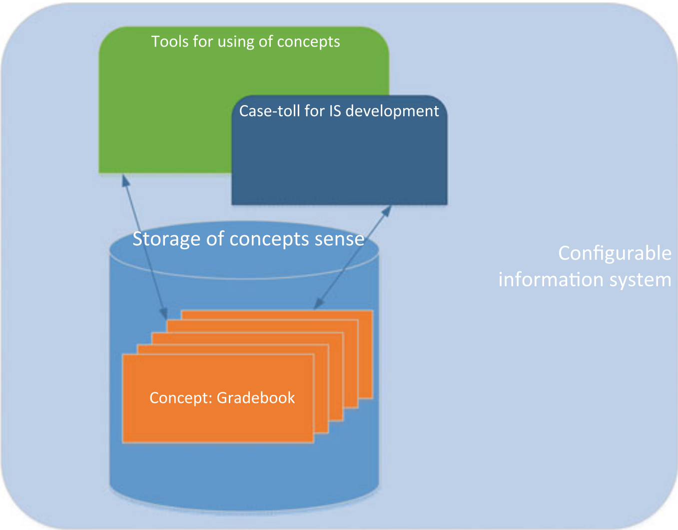


Fig. 3 Subject-oriented storage concept

3 The Model of Subject-Oriented Storage of Concepts Sense for Configurable Information Systems

The main difference between subject-oriented storage of concepts sense and classic databases is that this storage contain most part of configurable information system while the database is designed to store only the results of the information system utilization. Ideally, the subject-oriented storage of concepts sense is configurable information system as whole, and it should allow:

- Repeatedly create concepts with the concrete sense from of the existing concrete senses and concepts with the implicit sense;
- Create new concrete senses for solving the task of objectification concepts with the implicit sense;
- Create a new concepts with the implicit sense;
- Perform arithmetic, logic and other functions.

The possibilities of modern technologies do not allow going to the process of automatic generation of concepts with the concrete sense, but using a subject-oriented conception, this process can be automated.

Since every sense is nothing more than work of subject to create the concept, then the work can be described, stored and then continue to play if necessary. For describing the concepts' senses the mechanism of action can be used [6, 7].

Representation of configurable information as a set of concepts with the implicit sense allowing to model not system parts (as it does a lot of existing methodologies), but user activities and is a fundamentally new view of the information system

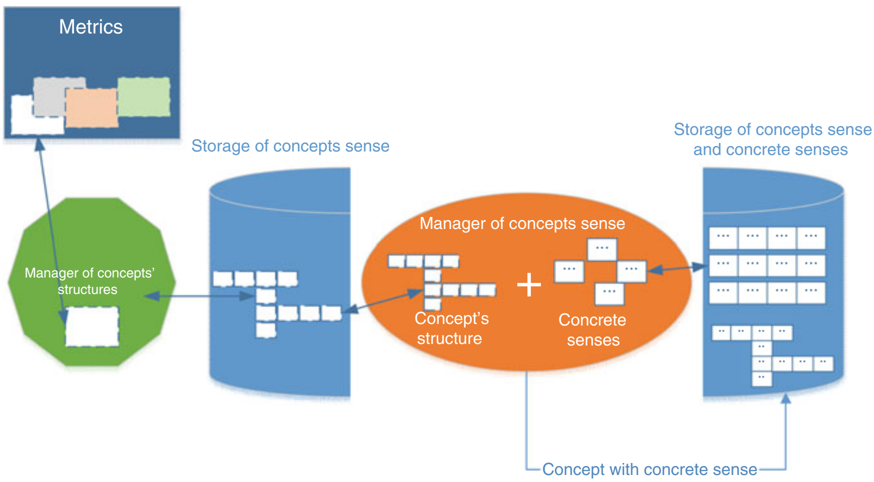


Fig. 4 Architecture of subject-oriented storage of concepts sense

[11–17]. To describe the subject operation the mechanism applies a simple set of characteristics—elements, functions, tools and results. Regardless of abstraction level, this set of characteristics set remains fixed and is a form, which filler content (specification of elements, functions, etc.) is a process of designing steps. Having common characteristics, linking through the substitution result as the element or as nested separate mechanisms of action form the description of user activity to create a sense of the whole system.

Therefore, the model of data representation in an subject-oriented conception being adequate to single representation model of configurable information system should use as a basic unit of storage user work.

Architecture of subject-oriented storage of concepts sense can be represented as follows (see Fig. 4).

From the viewpoint of the structural aspects, storage is divided into the following elements: concepts with the implicit sense, concrete senses, concepts with the concrete sense and results of the work with the concepts with the concrete sense.

The concept with the implicit sense—is a specification of object types that make up the domain. The concept with the implicit sense can only have the structure of the mechanism. Any content of mechanism is not provided. It is the basic building block of configurable information system as a whole.

The concrete senses—are elements, features and tools, which is filled mechanism structure during the objectification of concept with the implicit sense.

The concept with the concrete sense—is a description of the subject work on concept creation expressed by used elements, applied functions, tool that regulates the rules of interaction between elements and functions, and a result, which can be obtained by carrying out the work. It is worth noting that for the concept with the concrete sense only result type is specified, while the result value appears after the operation with concept. In other words, the concept can be created at any desired moment. The concept with the concrete sense—is the basic unit of storage and presentation of user data model in the configurable information system.

The results of the work with the concepts with the concrete sense—it is a means to implement the required functionality for the storage of factual data with the extended semantics. It contains and the result (a reflection of user experience with the system) and the process of its production, reflected in concrete elements, functions and instruments used in the user's work. It is obvious that such a model is beyond the scope of traditional relational technology. And for its implementation need to be able to define your own basic storage elements. This opportunity have NoSQL databases that implement the concept of the free data scheme determination.

4 NoSQL Implementation of Subject-Oriented Storage of Concepts Sense for Configurable Information Systems

Storage is being implemented on the basis of database GlobalsDB [18]. A key feature of this solution is fully supporting schema-free approach [19]. The storage permanently contains global variables—globals.

The approach to the implementation of storage model based on the creation of a special data structure. This structure is a multi-dimensional sparse array. The structural elements of the array are the elements of the metric chosen to describe the domain. In our case, this metric—the mechanism of action.

In accordance with the subject-oriented concept, the mechanism is divided into an element, a function, a tool and a result. The sense concepts described by using the mechanism of action, is the main pattern for configurable information system modelling in the form of subject works delegated to the information system. Next, we build a model using the selected metric (Fig. 5).

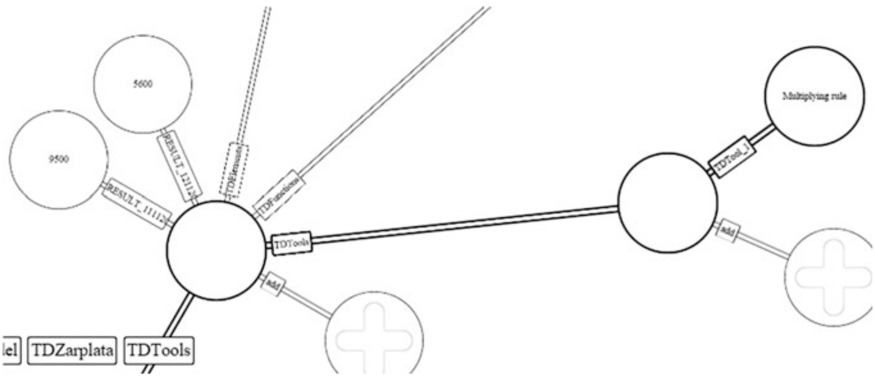


Fig. 5 User data model specification

Communication of concepts senses is made by means of a special system Global IS_A_RESULT. The value of this global is a link to another Global that contains the desired value. Building a model of the subject work is a process of sequential creation of tree from instances of the pattern "sense of the concept".

This tree contains both a specification of the subject work in the form of the concepts senses, and the results of work in the form of conventional factual data with extended semantics. Layer configuration in the form of concepts with the concrete senses described using globals with the prefix name TD*. The results of the system utilization stored in globals with the prefix name RESULT_*, where * is used as an absolute time, set at the time of receipt of the result.

5 Conclusion

The paper proposes the model of subject-oriented storage of concepts sense for configurable information systems, which solving the problem of variability existing configurable information systems and conformity between application systems and initial user tasks.

This tool allows to lead storage to a new level—the semantics of the data, the nature of their occurrence are clearly documented in a readable form for the user. This in turn has a positive effect both on the modelling process and the process of operation and improve system by the end user.

Proposed in this article model implements basic requirements for storage of configurable information system set out in the introduction of paper, as well as set forth in claim 3 functional features.

Subject-oriented storage of concepts sense for configurable information systems allows to remember the structure of the system, its properties, a custom data model and the results of the system in an adequate form with respect to the domain model and the information system.

The use of NoSQL technologies, in turn, will allow to realize high-performance solutions, devoid of drawbacks structure-independent databases [20–22].

Acknowledgement The reported study was partially supported by RFBR, research project No. 15-07-04033

References

1. Rogozov, Y.: Methodological approach to the construction of the concept of system. In: Proceedings of ISA RAS. 2015. T. 65, №1, pp. 89–109 (2015)
2. Date, C.J.: Database in Depth, 240 p. O'Reilly (2005). ISBN 0-596-10012-4

3. Savinov, A.: ConceptMix—self-service analytical data integration based on the concept-oriented model. In: Proceedings of 3rd International Conference on Data Management Technologies and Applications, pp. 78–85 (2014). ISBN 978-989-758-035-2
4. Palviainen, M., Kuusijarvi, J., Tuomisto, T., Ovaska, E.: Towards a data model of end-user programming of applications. In: Proceedings of 3rd International Conference on Data Management Technologies and Applications. Towards a Data Model of End-User Programming of Applications pp. 272–282 (2014) ISBN 978-989-758-035-2
5. Rossi, C., Guevara, A., Enciso, M.: A tool for user-guided database application development. In: Automatic Design of XML Models using CBD Proceedings of the Fifth International Conference on Software and Data Technologies, ICSOFT 2010, vol. 2, pp. 195–200
6. Rogozov, Y.I.: Methodology of creation of subject-oriented systems. J. Inf. Commun. 6–10 (2014)
7. Rogozov, Y.I., Sviridov, A.S.: The concept of building in-methodological information systems J. Inf. Commun. 11–14 (2014)
8. http://citforum.ru/database/classics/nfs_report/
9. Software Engineering Computer Systems. Paradigms, techniques and CASE-tools programming/EM Lavrishcheva, 283 p. K:. Science, Dumka (2013)
10. Ginige, A.: Meta-design paradigm based approach for iterative rapid development of enterprise WEB applications. In: Proceedings of the Fifth International Conference on Software and Data Technologies, ICSOFT, vol. 2 (2010)
11. Rogozov, Y., Sviridov, A., Belikov, A.: Approach to CASE-tool building for configurable information system development. In: Information and Communication Technology for Education (ICTE-2013), vol. 58, pp. 173–181. WIT Press, Southampton, Boston (2014)
12. Rogozov, Y.I., Degtyarev, A.A.: The method of configuring the software functionality of sonar information systems IzvSFU. Technical science N1. -S. pp. 13–18 (2014)
13. Rogozov, Y.I., Sviridov, A.S.: The approach to the construction of information systems based on a methodological approach. In: Collection of materials of the II International Conference Innovative technologies and didactics in teaching: a collection of articles of the international scientific-practical conference Tom 1. Taganrog, Ed of SFU, S. pp. 3–9 (2014)
14. Kucherov, S.A.: User-configurable information systems as a means to overcome the semantic gap. Mag. Inf. Commun. № 5, S. 135–137
15. Rogozov, Y.I.: The general approach to the organization of the system of definitions of concepts based on the principle of generating knowledge XII All-Russia meeting on VSPU-214 control problems, RAS, pp. 7822–7833 (2014)
16. Rogozov, Y., Sviridov, A.: The approach to the configurable system construction based on a methodological approach. In: International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM, vol. 1 no. 2, pp. 353–360 (2015)
17. Belikov, A., Rogozov, Y., Sviridov, A.: The approach to the information systems design based on the properties of the domain. In: International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM, vol. 1, no. 2, pp. 361–366 (2015)
18. McCreary, D., Kelly, A.: Making Sense of NoSQL: A Guide for Managers and the Rest of Us, 312 p. Manning Publications (2013). ISBN 978-1-61729-107-4
19. Fowler, M., Sadaladzh, J.P.: NoSQL:. develop a new methodology of non-relational databases=NoSQL Distilled, 192 p. M.: “Williams” (2013). ISBN 978-5-8459-1829-1
20. Kucherov, S., Sviridov, A., Belousova, S.A.: The formal model of structure-independent databases. In: Proceedings of 3rd International Conference on Data Management Technologies and Applications, Vienna, Austria, Scitepress, pp. 146–152. Science and Technology Publications. ISBN 978-989-758-035-2

21. Kucherov, S., Samoylov, A., Grishchenko, A.: Flexible database for configurable information systems. In: Proceedings of IEEE 8th International Conference on Application of Information and Communication Technologies-AICT2014, pp. 374–377. IEEE Catalog Number CFP1456H-PRT. ISBN 978-1-4799-4120-92
22. Kucherov, S., Rogozov, Y., Sviridov, A.: NoSQL storage for configurable information system: the conceptual model DATA2015. In: Proceedings of 4th International Conference on Data Management Technologies and Applications, pp. 279–287. ISBN 978-989758103-8

Part V
Fuzzy Graphs, Fuzzy Networks and Fuzzy Inference for Planning and Cognitive Modelling

Coloring Method of Fuzzy Temporal Graph with the Greatest Separation Degree

Alexander Bozhenyuk, Stanislav Belyakov and Igor Rozenberg

Abstract The concept of chromatic sets is introduced and discussed in this paper as invariant fuzzy temporal graph. Fuzzy temporal graph is a graph in which the degree of connectivity of the vertices is changed in discrete time. Chromatic set determines the greatest reparability degree of vertices of temporal fuzzy graph, when they color in a predetermined number of colors at any time. The example of finding the chromatic set of fuzzy temporal graph is considered too.

Keywords Fuzzy temporal graph · Invariant · Fuzzy subgraph · Graph coloring · Fuzzy chromatic set · Degree of reparability

1 Introduction

Graph models attract the great attention of specialists in the various fields of knowledge. They are used as models of various complex objects and phenomena with some defined structure. In addition, along with the applications of graph

This work has been supported by the Russian Foundation for Basic Research, Project № 15-07-00185a, and the Ministry of Education and Science of the Russian Federation under Project № 213.01-11/2014-48 (Base part, State task 2014/174).

A. Bozhenyuk (✉) · S. Belyakov
Southern Federal University, Nekrasovsky 44, 347922 Taganrog, Russia
e-mail: avb002@yandex.ru

S. Belyakov
e-mail: beliacov@yandex.ru

I. Rozenberg
Public Corporation "Research and Development Institute of Railway Engineers", Nizhegorodskaya 27/1, 109029 Moscow, Russia
e-mail: i.yarosh@vniias.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 450, DOI 10.1007/978-3-319-33609-1_30

models in the sciences such as chemistry, electrical engineering, physics, they are used in science, previously thought away from them—in economics, linguistics, and sociology.

Graph models can be used to define the relations between the structures of the different nature of their elements [1–4]. In this case, the relations between the elements (vertices of the graph) are considered as permanent and cannot be changed during the simulation. Such graphs in [5] were called static.

However, if the relationship between the elements of the structure under consideration change over time, the traditional graph models are not suitable to describe them. They cannot be used for modeling processes in time. In this case, consideration is relevant graph models, i.e., graphs in which connections between the vertices may change in a discrete (or continuous) time. These graphs were called temporal [6, 7]. When in the temporal graph, the links between vertices are partially undefined or fuzzy, then we arrive at the graphs that have been named fuzzy temporal [8–10].

When these graphs are used as models of complex systems actual is consideration their invariants, i.e. such characteristics of the graph that do not change when it is isomorphic transformation [11]. To crisp graphs such invariants are the numbers of internal and external stability, chromatic number, chromatic class et al. [3]. In this paper we introduce the notion of temporal chromatic set, which is an invariant of fuzzy temporal graph.

2 Temporal Chromatic Set

Let a fuzzy temporal graph $\tilde{G} = (X, \{\tilde{\Gamma}_t\}, T)$ be given [10, 12], where set X is a set of vertices $(|X| = n)$, set of natural numbers $t = \{1, 2, \ldots, T\}$ defines a discrete time and set $\{\tilde{\Gamma}_t\}$ defines a family correspondences, which display the vertices of X into itself at times $t = \overline{1, T}$.

We introduce the fuzzy subgraph $\tilde{G}_t = (X, \tilde{U}_t)$, in which the set of vertices X is the same as in the original temporal fuzzy graph $\tilde{G} = (X, \{\tilde{\Gamma}_t\}, T)$; set $\tilde{U}_t = \{\mu_t(x_i, x_j) \mid (x_i, x_j) \in X^2\}$ is a fuzzy set of edges at the discrete times $t = \overline{1, T}$, and a function μ_t is a membership function, which displays $X^2 \to$ [0, 1].

We color each vertex $x \in X$ of fuzzy subgraph $\tilde{G}_t = (X, \tilde{U}_t)$ by one of k colors $(1 < k < n)$ and consider a subgraph $\tilde{G}_i^{(t)} = (X_i, \tilde{U}_i^{(t)})$. Here X_i is a subset of vertices, which have i color. Then the value $\alpha_i = 1 - \vee_{x, y \in X_i} \mu_G(x, y)$ defines the degree of internal stable of subset X_i [13, 14].

Definition 1 Separation degree of fuzzy subgraph $\tilde{G}_t$ with k colors is called a value:

$$L = \underset{i = \overline{1, k}}{\&} \alpha_i = \underset{i = \overline{1, k}}{\&} (1 - \vee_{x, y \in X_i} \mu_G(x, y)).$$

Separation degree L of fuzzy subgraph $\tilde{G}_t$ depends on the number of colors k and concrete coloring of vertices.

We associate with fuzzy subgraph a family of fuzzy sets $\mathfrak{R}^{(t)} = \left\{ \tilde{A}_G^{(t)} \right\}$, $\tilde{A}_G^{(t)} = \{\langle L_A^{(t)}(k)/k\rangle | k = \overline{1,n}\}$. Here, the value $L_A^{(t)}(k)$ determines the separation degree of subgraph $\tilde{G}_t$ when coloring in particular k colors.

Definition 2 A fuzzy set $\tilde{\gamma}^{(t)} = \{\langle L_\gamma^{(t)}(k)/k\rangle \ k = \overline{1,n}\}$ is called a fuzzy chromatic set of subgraph $\tilde{G}_t$ if the condition $\tilde{A}_G \subseteq \tilde{\gamma}$ is performed for any set $\tilde{A}_G^{(t)} \in \mathfrak{R}^{(t)}$, or else:

$$(\forall \tilde{A}_G^{(t)} \in \mathfrak{R}^{(t)})\ (\forall k = \overline{1,n})\ [L_A^{(t)}(k) \le L_\gamma^{(t)}(k)].$$

Chromatic set of subgraph $\tilde{G}_t$ determines the highest degree of separability in the color of it vertices by $1, 2, \ldots, n$ colors in time t.

Definition 3 A fuzzy set $\tilde{\gamma} = \cap_{t=\overline{1,T}} \gamma^{(t)} = \{\langle a_1/1\rangle, \langle a_2/2\rangle, \ldots, \langle a_n/n\rangle\}$ is called a fuzzy chromatic set fuzzy temporal graph $\tilde{G}$.

Chromatic set of fuzzy temporal graph $\tilde{G}$ determines the highest degree of separability in the color of it vertices by $1, 2, \ldots, n$ colors at time $t \in T$.

3 The Method of Finding Temporal Chromatic Set

Definition 3 [13] We call the subset $\Psi' \subseteq X$ a subgraph $\tilde{G}_t$ as maximal internal stable set with the degree $\alpha(\Psi')$, if the condition $\alpha(\Psi'') < \alpha(\Psi')$ is true for any subset $\Psi' \subset \Psi''$.

The following proposition is true.

Proposition *A fuzzy set $\tilde{\gamma}^{(t)} = \{\langle L_\gamma^{(t)}(k)/k\rangle | k = \overline{1,n}\}$ is a fuzzy chromatic set if and only if it is not more then k maximal internal stable sets $\Psi_1, \Psi_2, \ldots, \Psi_{k'}$ with the degrees of internal stability $\alpha_1, \alpha_2, \ldots, \alpha_{k'}$, $(k' \le k)$ and moreover:*

(1) $\min\{\alpha_1, \alpha_2, \ldots, \alpha_{k'}\} = L_{\tilde{\gamma}}^{(t)}(k)$;
(2) $\cup_{j=\overline{1,k'}} \Psi_j = X$;
(3) *There is not another family* $\{\Psi'_1, \Psi'_2, \ldots, \Psi'_{k''}\}$, $k'' \le k$ *for which* $\min\{\alpha'_1, \alpha'_2, \ldots, \alpha'_{k''}\} > \min\{\alpha_1, \alpha_2, \ldots, \alpha_{k'}\}$ *and the condition 2 is true.*

The observed proposition proves the following algorithm of finding fuzzy chromatic set of subgraph $\tilde{G}_t$:

(1) To determine the family of maximal internal stable sets $T = \{\Psi_1, \Psi_2, \ldots, \Psi_p\}$ with the degrees of internal stability $\alpha_1, \alpha_2, \ldots, \alpha_p$ respectively.

(2) To determine the family $\{\Psi'_1, \Psi'_2, \ldots, \Psi'_k\} \subseteq T$ for which the condition (2) is true and $\min\{\alpha'_1, \alpha'_2, \ldots, \alpha'_k\}$ takes the biggest value. To determine $L_{\tilde{\gamma}}^{(t)}(k) = \min\{\alpha'_1, \alpha'_2, \ldots, \alpha'_k\}$.

(3) The step 2 is repeated for all $k = \overline{1, n-1}$.

We will consider a method for determination of all maximal internal stable sets with the highest degree of internal stable. This method is a generalisation of the Maghout's method for fuzzy graphs [13].

Let Ψ be a certain maximal internal stable set with the degree of internal stable $\alpha(\Psi)$. For arbitrary vertices $x_i, x_j \in X$, one of the following cases may be realised: (a) $x_i \notin \Psi$; (b) $x_j \notin \Psi$; (c) $x_i \in \Psi$ and $x_j \in \Psi$. In the last case the degree $\alpha(\Psi) \leq 1 - \mu_U(x_i, x_j)$.

In other words, the following expression is true:

$$(\forall x_i, x_j \in X)[x_i \notin \Psi \vee x_j \notin \Psi \vee (\alpha(\Psi) \leq 1 - \mu_U(x_i, x_j))]. \quad (1)$$

We connect a Boolean variable p_i taking 1 when $x_i \in \Psi$ and 0 when $x_i \notin \Psi$, with each vertex $x_i \in X$. We associate the expression $\alpha(\Psi) \leq 1 - \mu_U(x_i, x_j)$ with a fuzzy variable $\xi_{ij} = 1 - \mu_U(x_i, x_j)$.

Considering the expression (1) for all possible values i and j we obtain the truth of the following expression:

$$\Phi_\Psi = \mathop{\&}_{i} \mathop{\&}_{j \neq i} (\bar{p}_i \vee \bar{p}_j \vee \xi_{ij}) = 1. \quad (2)$$

We open the parentheses and reduce the similar terms using the following rule:

$$\xi' \,\&\, a \vee \xi'' \,\&\, a \,\&\, b = \xi' \,\&\, a, \text{ for } \xi' \geq \xi''. \quad (3)$$

Here, $a, b \in \{0, 1\}$ and $\xi', \xi'' \in [0, 1]$. Then for each disjunctive term, the totality of all vertices corresponding to the variables missing in the totality, gives a maximal internal stable set with the obtained degree of internal stable.

We construct the matrix $F = \|f_{ij}\|$, $i = \overline{1, n}$, $j = \overline{1, p}$ where the lines correspond to the vertices of subgraph $\tilde{G}_t$ and columns correspond to the maximal internal stable sets. If $x_i \in \Psi_j$, then the value f_{ij} has the value α_j, if $x_i \notin \Psi_j$, then the value f_{ij} has the value 0. So, the task of finding fuzzy chromatic set $\tilde{\gamma}^{(t)}$ of subgraph $\tilde{G}_t$ is the task of finding the covering of all lines by k columns ($k = \overline{1, n-1}$) with the maximum of the volume $\min\{\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_k}\}$.

Example Consider the example of the temporal fuzzy graph $\tilde{G} = (X, \{\tilde{\Gamma}_t\}, T)$ whose set of vertices $X = \{x1, x2, x4\}$, the time $T = \{1, 2, 3\}$, and multi-valued mapping $\{\tilde{\Gamma}_t\}$ has the form:

$$\tilde{\Gamma}_1(x_1) = \{\langle 0,2/x_2 \rangle\},\ \ \tilde{\Gamma}_2(x_1) = \{\langle 0,5/x_2 \rangle\},\ \ \tilde{\Gamma}_2(x_2) = \{\langle 0,4/x_3 \rangle\},$$
$$\tilde{\Gamma}_3(x_2) = \{\langle 0,6/x_3 \rangle\},\ \ \tilde{\Gamma}_1(x_3) = \{\langle 0,2/x_4 \rangle\},\ \ \tilde{\Gamma}_2(x_3) = \{\langle 0,3/x_4 \rangle\},$$
$$\tilde{\Gamma}_1(x_4) = \{\langle 0,2/x_1 \rangle\},\ \ \tilde{\Gamma}_2(x_4) = \{\langle 0,9/x_2 \rangle\},\ \ \tilde{\Gamma}_3(x_4) = \{\langle 0,1/x_1 \rangle, \langle 1/x_2 \rangle\}.$$

A graphical representation of this graph is shown on Fig. 1. Here the membership functions at the times $T = \{1,2,3\}$ are specified on the graph edges.

Temporal fuzzy graph can be represented as the union of T fuzzy subgraphs defined on the same set of vertices X. Since the graph $\tilde{G}$ is represented in the form $\tilde{G} = \bigcup_{t=1}^{3} \tilde{G}_t$, where fuzzy subgraphs $\tilde{G}_1$, $\tilde{G}_2$, and $\tilde{G}_3$ are shown on Figs. 2, 3 and 4.

For fuzzy subgraph $\tilde{G}_t$, presented in Fig. 2, we find fuzzy chromatic set.

Fig. 1 Fuzzy temporal graph $\tilde{G}$

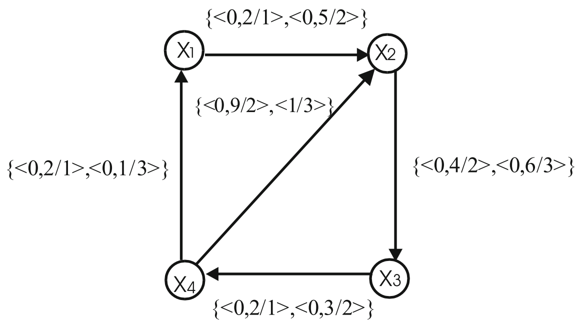


Fig. 2 Subgraph of subgraph $\tilde{G}_t$ at time t = 1

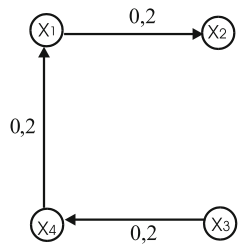
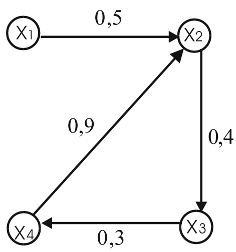


Fig. 3 Subgraph $\tilde{G}_2$ at time t = 2



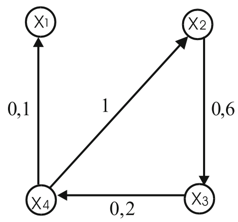


Fig. 4 Subgraph $\tilde{G}_3$ at time t = 3

The vertex matrix R_x for this subgraph has view:

$$R_x = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{pmatrix} 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2 \\ 0.2 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

The corresponding expression (2) for this subgraph has the following form:

$$\Phi_\Psi = (\bar{p}_1 \vee \bar{p}_2 \vee 0.8)\,\&\,(\bar{p}_3 \vee \bar{p}_4 \vee 0.8)\,\&\,(\bar{p}_4 \vee \bar{p}_1 \vee 0.8) = 1.$$

Multiplying parenthesis (1) and (2), and using rules (3) we obtain:

$$\Phi_\Psi = (\bar{p}_1\bar{p}_3 \vee \bar{p}_1\bar{p}_4 \vee \bar{p}_2\bar{p}_3 \vee \bar{p}_2\bar{p}_4 \vee 0.8)\,\&\,(\bar{p}_4 \vee \bar{p}_1 \vee 0.8) = 1.$$

Completing the transformations of the fuzzy logical variables, we finally have:

$$\Phi_\Psi = \bar{p}_1\bar{p}_3 \vee \bar{p}_1\bar{p}_4 \vee \bar{p}_2\bar{p}_4 \vee 0.8 = 1.$$

It follows from the last expression that the considered fuzzy subgraph $\tilde{G}$ has 4 maximal internal stable sets: $\Psi_1 = \{x_2, x_4\}$, $\Psi_2 = \{x_2, x_3\}$, $\Psi_3 = \{x_1, x_3\}$, with the degrees of internal stability $\alpha_1 = \alpha_2 = \alpha_3 = 1$, and $\Psi_4 = \{x_1, x_2, x_3, x_4\}$ with the degree of internal stability $\alpha_4 = 0.8$.

For the fuzzy subgraph $\tilde{G}$, presented in Fig. 2, the matrix F has view:

	Ψ_1	Ψ_2	Ψ_3	Ψ_4
x_1	0	0	1	0.8
x_2	1	1	0	0.8
x_3	0	1	1	0.8
x_4	1	0	0	0.8

If k = 2, then the covering defines the columns Ψ_1 and Ψ_3 with the degree L(2) = 1. If k = 1, then the covering defines the column Ψ_4 with the degree

L(1) = 0.8. So, for the fuzzy subgraph $\tilde{G}_1$ presented in Fig. 2, the fuzzy chromatic set is:

$$\gamma^{(1)} = \{\langle 0.8/1\rangle, \langle 1/2\rangle, \langle 1/3\rangle, \langle 1/4\rangle\}.$$

Similarly, we define fuzzy chromatic sets of subgraphs presented in Figs. 3 and 4:

$$\gamma^{(2)} = \{\langle 0.1/1\rangle, \langle 0.6/2\rangle, \langle 0.7/3\rangle, \langle 1/4\rangle\};$$
$$\gamma^{(3)} = \{\langle 0/1\rangle, \langle 1/2\rangle, \langle 1/3\rangle, \langle 1/4\rangle\}.$$

Hence the chromatic set of fuzzy temporal graph is defined as:

$$\tilde{\gamma} = \bigcap_{t=\overline{1,3}} \gamma^{(t)} = \{\langle 0/1\rangle, \langle 0.6/2\rangle, \langle 0.7/3\rangle, \langle 1/4\rangle\}.$$

Thus, fuzzy temporal graph presented in Fig. 1, at any given time can be colored by two colors with the degree of reparability not less than 0.6; by three colors with the degree of reparability at least 0.7 and four colors with a degree of reparability at least 1.

4 Conclusions

In this paper, we considered the concept of fuzzy chromatic set as an invariant of fuzzy temporal graph, i.e., such graph, in which the degree of connectivity of the vertices is changed in discrete time. Chromatic set determines the greatest degree of reparability vertices of temporal fuzzy graph, when they color in the specified number of colors at any time. It has been shown that the task of painting a fuzzy temporal graph is reduced to finding the fuzzy chromatic set of graph. The example of finding the chromatic set of fuzzy temporal graph is considered too.

References

1. Ore, O.: Theory of Graphs. American Mathematical Society Colloquium Publications, Providence (1962)
2. Minieka, E.: Optimization Algorithms for Networks and Graphs. Inc. New York (1978)
3. Bershtein, L., Bozhenyuk, A.: Graph Theory. UFY, Rostov-on-Don (2014)
4. Bershtein, L., Bozhenyuk, A.: Fuzzy graphs and fuzzy hypergraphs. In: Dopico, J., de la Calle, J., Sierra, A. (eds.) Encyclopedia of Artificial Intelligence, Information SCI, pp. 704–709. Hershey, New York (2008)
5. Kostakos, V.: Temporal graphs. In: Proceedings of Physica A: Statistical Mechanics and its Applications, vol. 388, Issue 6, pp. 1007–1023. Elsevier (2008)

6. Bershtein, L., Bozhenyuk A.: The using of temporal graphs as the models of complicity systems. Izvestiya UFY. Technicheskie nayuki. Taganrog: TTI UFY. №4 (105), 198–203 (2010)
7. Bozhenyuk, A., Gerasimenko, E., Rozenberg, I.: Determining the minimum cost flow in fuzzy dynamic network with GIS "ObjectLand". In: Proceedings of 9th International Conference on Application of Information and Communication Technologies, AICT 2015, pp. 294–298. UFY, Rostov on Don (2015)
8. Bershtein, L., Beliakov, S., Bozhenyuk, A.: Maghout method for definition of fuzzy base set of fuzzy temporal graph. Izvestiya UFY. Technicheskie nayuki. Taganrog: TTI UFY. № 1. 70–76 (2014)
9. Bershtein, L., Bozhenyuk, A., Rozenberg, I.: Definition method of strong connectivity of fuzzy temporal graphs. Vestnik RGUPS. Rostov-on-Don: RGUPS. №3 (43), 15–20 (2011)
10. Bershtein, L., Beliakov, S., Bozhenyuk, A.: Routing in the conditions of uncertainty with use fuzzy temporal external stable sets. Izvestiya UFY. Technicheskie nayuki. Taganrog: TTI UFY. №1 (126), 121–127 (2012)
11. Bozhenyuk, A., Ginis, L.: Algorithmic support of the research system links to the socio-economic system based on fuzzy graph models. Economica i menedzhment system ypravleniya. №1.1 (15), 115–122 (2015)
12. Bershtein, L., Bozhenyuk, A.: Fuzzy coloring and evaluation of the degree of izomorphizm of fuzzy graphs. J. Comput. Syst. Sci. Int. **41**(3), 447–453 (2002)
13. Bershtein, L., Bozhenuk, A.: Maghout method for determination of fuzzy independent, dominating vertex sets and fuzzy graph kernels. Int. J. Gen. Syst. **30**(1), 45–52 (2001)
14. Bozheniuk, V., Bozhenyuk, A., Belyakov, S.: Optimum allocation of centers in fuzzy transportation networks with the largest vitality degree. In Proceedings of the 2015 Conference of the International Fuzzy System Association and the European Society for Fuzzy Logic and Technology, pp. 1006–1011. Atlantis Press (2015)

Temporal Coloured Petri Nets as a Tool for Modelling of Complex Dynamic Systems

Alexander P. Eremeev and Yury I. Korolev

Abstract This paper considers the problem of using Petri nets to model and analyse the processes in complex dynamic systems such as intelligent decision support systems and real-time systems. A new coloured Petri net subclass is proposed—real-time coloured Petri net with Allen temporal logic support: formal definitions and an example of a simple model are given, analysis and verification methods for the subclass are considered.

Keywords Modified Petri nets · Intelligent systems · Dynamic systems · Temporal logic · Verification

1 Introduction

Complex technical or organizational systems (objects), especially real-time systems are rather difficult to formalize. Such systems have a number of atypical features of traditional management and control: uniqueness, lack of formalized (quantitative) objective function or optimality criteria, high dynamics, incomplete description of the object and subjectivity of the decision making persona (DMP) behaviour during the management of such systems. These features have to be considered during designing decision support systems (DSS) which help DMP to manage complex objects. Traditional purpose of information systems is having forwarded the requisite information to DMP, so the last one could make appropriate and effective decisions

The work was supported by the Ministry of Education and Science of the Russian Federation, RFBR and Foundation for Assistance to Small Innovative Enterprises.

A.P. Eremeev (✉) · Y.I. Korolev
National Research University "Moscow Power Engineering Institute",
Krasnokazarmennaya 14, 111250 Moscow, Russia
e-mail: eremeev@appmat.ru

Y.I. Korolev
e-mail: korolevyu@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 450, DOI 10.1007/978-3-319-33609-1_31

during control process. The management of today's complex systems and real-time processes requires more powerful intelligent DSS (IDSS) that not only provide information, but also carry out the preliminary analysis of the received data (from the sensors etc.). IDSS are able to give an advice or recommendation to DMP, to forecast subsequent developments and to select the most promising solutions.

IDSS intended to assist DMP in the management of complex objects and processes under tight time constraints and the availability of various types of uncertainties (in the information coming from the outside, in the knowledge embedded into the system), are called real-time IDSS (RT IDSS) [1]. High risk of errors in the design stage and extremely high price of manifestation of these errors during the operation phase are typical for such systems. Therefore there is a sustained interest in modeling and analysis of processes in RT IDSS.

For perspective RT IDSS construction and testing one should use advanced tools for modelling processes in complex systems that make possible to take into account temporal (time) dependence. The paper proposes to use tools based on colored Petri net subclass—so-called temporal coloured Petri nets [2].

2 Real-Time Coloured Petri Nets with Temporal Logic Support

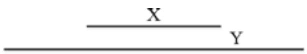
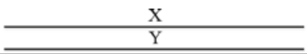
Today the use of mathematical models as a tool for process simulation is a relevant and promising research direction. This approach involves the use of well-known theoretical formalism which makes the modelling process significantly easier. The modelled system properties and the desired expressiveness of the model are the key criteria of formalism selection. It is also important to reflect not only regularities of functioning but also the structural characteristics. The visual expressiveness should be maximized for the detection of structural regularities. Therefore one of the graph oriented modelling tools—Petri nets—has been chosen as the base model.

Petri nets are a recognized tool of modelling and functional analysis of parallel and distributed computing systems and processes. One can enter a number of additional rules and conditions in the modelling algorithms to yield one or another Petri net subclass. Coloured Petri nets (CP-nets) [3] is a graph oriented language for design, specification, simulation and control of distributed and parallel systems: flow of the process is shown by graphic primitives, and the required data processing is simulated by a special language structures. For each CP-net a conventional Petri net can be built, and vice versa. Thus the use of CP-nets does not provide any theoretical advantages, but in practice such networks are more compact and user-friendly modelling language than conventional Petri nets.

Both classical Petri nets and CP-nets do not allow to take into account the time factor and it's a major shortcoming of these formalisms. Therefore, they cannot be used to effectively simulate real processes in which the analysed condition of the system depends on the current time. The importance of presenting time and temporal dependencies is discussed almost since the appearance of intelligent systems. This problem has become particularly acute because of the emergence and development of dynamic intelligent systems, typified by RT IDSS. Therefore the most important challenge in the advanced IDSS modelling is the problem of representing and operating the temporal dependencies.

Real-time coloured Petri nets (RTCP-nets) [4] is a functional CP-net subclass oriented to modelling and analysis of real-time systems. This formalism supports only quantitative temporal dependence, which significantly limits the scope of application and calls into question the "intelligence" of the approach. There are two types of temporal dependencies [5]: quantitative (metric)—when quantitative measures on the time axis are used to represent time (e.g. "Failure has occurred at 17.00"); and qualitative dependencies—when only the relative position of events or actions in time is used (e.g. "At first failure occurred in the supply system, and then outage occurs"). The expressiveness increases if the developer is able to operate both quantitative and qualitative temporal dependencies. A possible solution to this problem is to modify the RTCP-net formalism by adding the ability to use Allen's interval temporal logic [2]. Allen's logic [6] is characterized by sufficient expressiveness and the existence of polynomial output algorithms. Intervals are used as a time primitives, which is important because point-based temporal logics are usually not sensitive enough to detect semantic differences between verbal predicates. Temporal interval X is an ordered pair (X_{-}, X_{+}) where $X_{-} < X_{+}$ are time moments (e.g. points on the real line). Table 1 lists basic relations of Allen's logic between two intervals $B = \{b, bi, m, mi, o, oi, d, di, s, si, f, fi, e\}$. Atomic formula of Allen's logic is an expression XrY where X and Y are temporal intervals, r is one of the basic relations from B. Interval formula is an expression $\phi = X\{r_1, r_2, \ldots, r_n\}Y$.

Table 1 Basic relations of Allen's interval logic

Relation and it's inversion	Designation	Illustration
$X\ before\ Y/Y\ after\ X$	b/bi	
$X\ meets\ Y/Y\ met-by\ X$	m/mi	
$X\ overlaps\ Y/Y\ overlapped-by\ X$	o/oi	
$X\ during\ Y/Y\ includes\ X$	d/di	
$X\ starts\ Y/Y\ started-by\ X$	s/si	
$X\ finishes\ Y/Y\ finished-by\ X$	f/fi	
$X\ equals\ Y$	e	

2.1 Formal Definitions and Model Example

The formal definition of the proposed formalism—RTCP-nets that support Allens logic (AL RTCP-nets)—is based on the RTCP-nets definition [4]. Let us define some typed expression language L assuming the expressions are built from variables and values by means of multi-set summation, and the finite model of this language U consisting of distinguishable tokens. The set of all finite ordered subsets $U^{[n]} \cong \{[U[1], U[2], \dots, U[n]]|(\forall i \in 1 \dots n)U[i] \in U\}$ where $(\forall i', i'' \in \in 1 \dots n)(i' \neq i'' \supset U[i'] \neq U[i''])$ denoted as $U^{[\,]} \cong \bigcup_{n \in \mathbb{N}_0}$ and set of all possible multi-sets over U denoted as $U^{(\,)} \cong \{(u_1, u_2, \dots, u_n])|n \in \mathbb{N}_0, (\forall i \in 1 \dots n)u_i \in U\}$. The type of an element $x \in U$ denoted as $\xi(x)$, the type of an expression $\theta \in L$ as $\xi(\theta)$, the set of variables in an expression θ as $\nu(\theta)$.

A real-time CP-net that support Allens logic (AL RTCP-net) is a tuple $N = \langle \Sigma, P, T, F, \dot{\xi}, \gamma, \pi, \varepsilon_\Sigma, \varepsilon_\mathrm{T}, m_0 \rangle$ satisfying the requirements below:

- Σ is a finite set of types (colours), $\Sigma \subseteq U^{[n]}$;
- $P \equiv [p_1, p_2, \dots, p_{|P|}]$ is a finite set of places;
- T is a finite set of transitions, $T \cap P = \varnothing$;
- $F \subseteq (P \times T) \cup (T \times P)$ is a finite non-empty set of arcs;
- $\dot{\xi} : P \to \Sigma$ is a colour function, each token $u \in U$ on $p \in P$ belongs to the type $\dot{\xi}(p) \in \Sigma$;
- $\gamma : T \to Bool$ is a guard function maps each transition $t \in T$ to an expression of type Boolean, i.e. a predicate, such that $(\forall t \in T)\xi(\nu(\gamma(t))) \subseteq \Sigma$;
- $\pi : T \to \mathbb{R}_0$ is a priority function;
- $\varepsilon_\Sigma : F \to L$ is an arc expression function which maps each arc to some L-language expression and $(\forall p \in P)(\forall f \in F)((f = \langle t, p \rangle \vee f = \langle p, t \rangle) \supset \supset \xi(\varepsilon_\Sigma(f)) \subseteq \dot{\xi}(p))$;
- $\varepsilon_\mathrm{T} : F \to \mathbb{R}_0$ is a time expression function;
- $m_0 \in M$ is an initial marking where $M \cong \{m|m : P \to U^{(\,)} \times \mathbb{R}\}$ denotes the set of all possible markings. Marking $m \in M$ is a function which maps each place $p \in P$ to a pair $m(p) \equiv \langle \mu(p), \tau(p) \rangle$ where $\mu(p) \in U^{(\,)}$ is a multi-set of tokens, $(\forall u \in \mu(p))u \in \dot{\xi}(p)$, and $\tau(p) \in \mathbb{R}$ is a time stamp.

The terms "variables" and "expressions" are used for AL RTCP-net in the same way as for CP-net in general: as in typed lambda-calculus and functional programming languages. This means that variables are bound to values (instead of being assigned to). Let's denote the set of variables in functions ε_Σ and ε_T of transition $t \in T$ input and output arcs and in function $\gamma(t)$ as V_t. The binding of transition $t \in T$ is a function $\beta : V_t \to U^{(\,)}$ satisfying next requirement: $(\forall v \in V_t)\beta(v) \in \xi(v)$. The value obtained by evaluating some expression—$\gamma(t)$, $\varepsilon_\Sigma(f)$ or $\varepsilon_\mathrm{T}(f)$—in a binding β denoted by $\gamma(t)_\beta$, $\varepsilon_\Sigma(f)_\beta$ and $\varepsilon_\mathrm{T}(f)_\beta$ respectively, where $t \in T$ and $f \in F$. The set of all bindings of a transition $t \in T$ is denoted by $Bind_t$, the set of all bindings is denoted by $Bind$.

Formal definition of interval time model that used in AL RTCP-nets is given below. Time interval is a finite ordered subset $Int \cong [\tau_0, \tau_1, \dots, \tau_n]$, $\tau_i \in \mathrm{T}_D$, $i \in 0 \dots n$, $\mathrm{T}_D \cong \{\tau|\tau = k * \tau', k \in \mathbb{N}_0, \tau' \in \mathbb{R}_+\}$ where $(\forall i \in 1 \dots n)(\exists k \in \mathbb{N})(\tau_i =$

$= k * \tau' \supset \tau_{i-1} = (k-1) * \tau'$). The set of all such intervals (including empty interval $\varnothing$) for the set of time points T_D is denoted by Int_D. The function that determines the marking of the net at time point $\tau \in \mathrm{T}_D$ is denoted by $\rho : P \times \mathrm{T}_D \to U^{()} \times \mathbb{R}$, the set of various token multi-sets in $p \in P$ is denoted by $U_p \cong \{\dot{u} | \dot{u} \in U^{()}, (\forall u \in \dot{u}) u \in \dot{\xi}(p)\}$. Now a function $\zeta : P \times U_p \to \mathrm{T}_D$ can be defined:

$$\zeta(p, \dot{u}) \cong [\tau_0, \tau_1, \dots, \tau_n], (\forall i \in 0 \dots n)\tau_i \in \mathrm{T}_D, (\forall i, j \in 0 \dots n)(i > j \supset \tau_i > \tau_j),$$
$$(\forall i \in 0 \dots n)(\rho(p, \tau_i) = m(p) = \langle \mu(p), \tau(p) \rangle \supset \dot{u} \subseteq \mu(p) \wedge \tau(p) \leq 0)\ .$$

Each subset $\zeta(p, \dot{u})$ may be represented as a tuple of time intervals $\zeta(p, \dot{u}) = \langle Int_0, Int_1, \dots, Int_n \rangle$. Let us now define a function $\chi : \zeta \times \mathrm{T}_D \to Int_D$ that maps each time point $\tau \in \mathrm{T}_D$ to a time interval $Int_i = [\tau_{i_0}, \tau_{i_1}, \dots, \tau_{i_{\dot{n}}}]$ from a tuple $\zeta(p, \dot{u})$ using denotation $[q \to e' : e'']$: "if q then e' else e''":

$$\chi(\zeta(p, \dot{u}), \tau) \cong [(\exists Int_i = [\tau_{i_0}, \tau_{i_1}, \dots, \tau_{i_{\dot{n}}}] \in \zeta(p, \dot{u}))(\tau_{i_0} \leq \tau \wedge (\forall Int_j \in \zeta(p, \dot{u}),$$
$$Int_j = [\tau_{j_0}, \tau_{j_1}, \dots, \tau_{j_{\dot{n}}}])(j > i \supset \tau_{j_0} > \tau)) \to Int_i : \varnothing]\ .$$

Function χ allows to work with the current or last ended time interval at each moment of modelling. One of the most important AL RTCP-net characteristic is the possibility to use Allen's logic formula as an expression of guard function $\phi : Int_D \times Int_D \to Bool$:

$$\phi = \chi(\zeta(p_1, \dot{u}_1), \tau)\{r_1, r_2, \dots, r_n\}\chi(\zeta(p_2, \dot{u}_2), \tau),$$
$$p_1, p_2 \in P, \dot{u}_1 \in U_{p_1}, \dot{u}_2 \in U_{p_2}, r_i \in B, i \in 1 \dots |B|, \tau \in \mathrm{T}_D\ .$$

Let's consider the model of a simple automatic train stop (ATS) system [4] to illustrate the main aspects of the AL RTCP-net formalism (Fig. 1).

A light signal is turned on every 60 s to check whether the driver controls the train. If the driver fails to acknowledge the signal within 6 s, a sound signal is turned on. Then if the driver does not deactivate the signals within 3 s, the emergency brakes are applied automatically to stop the train. The set of places is specified as follows: $P = \{Timer1, LightSig, SoundSig, Brake, Timer2, Driver\}$. Allen's logic formulas are used in the model as guard functions of transitions *DisactLS* and *DisactSS* and indicate the timely response of the driver (DMP) to the light signal and to the sound signal respectively.

The main advantage of new formalism in this case is the possibility to set not a specific reaction time (the only option for RCTP-nets) but the time intervals that determine further behaviour of the model. Thus the support of interval temporal logic has made possible to reflect correctly the uncertainty inherent to the problem, which is one of the fundamental principles in the design of perspective RT IDSS.

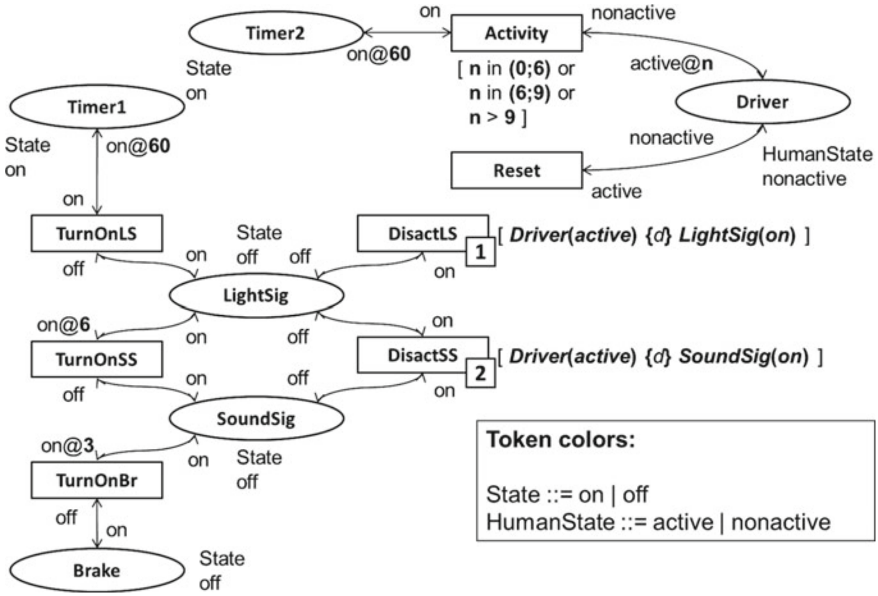


Fig. 1 Example of AL RTCP-net: ATS system model

2.2 Model Analysis and Verification

Using Petri nets implies a high level of parallelism. As well as the work with temporal dependencies it requires tools for analysis and verification of the models based on this formalism. AL RTCP-net is a visual programming language with a formally defined syntax. Models developed on the basis of this apparatus seem to be completely formalized. But semantically this is not true: full formal behaviour description is not derived directly from the model itself. Therefore one must use additional tools for the analysis and verification of AL RTCP-net models. There are three main groups of methods for Petri net analysis [7]: methods based on the graphs of the net states changes; matrix methods that use the equation of net and invariants; reduction methods. The latter two ones are rarely used for CP-nets due to the high complexity of formal definitions. Reachability and coverability graphs are usually considered as the primary tool for analysing [4].

Let formally define the key concepts to work with tools for analysis of AL RTCP-nets. Transition from one state to another may be caused either by firing transition $t \in T$ in binding β or by passage of time—a gradual decrease of each time stamp by a fixed value τ' until the transition that may fire appears. An event of firing has absolute priority. The passage of time is needed just to wait until the next transition fires. A state m' is reachable from state m if there is a finite sequence of transitions feasible from state m and leading to state m'. Let's denote by $R(m)$ the set of all states reachable from state m. Analysis of AL RTCP-net can be performed using

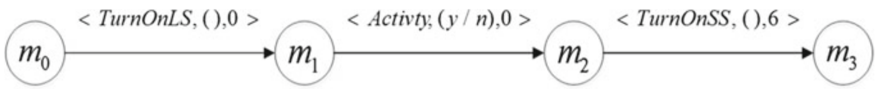


Fig. 2 Fragment of AL RTCP-net reachability graph

reachability graph (RG) which nodes are marked by elements of the set $R(m_0)$, and each arc indicates a change of state via firing of the transition t in binding β after passage of time $k * \tau' \geq 0$. Fragment of RG constructed according to these rules for the ATS system model (Fig. 1) is shown in Fig. 2.

Analysis of the net properties can be performed using labels of RG nodes and arcs. Each arc is marked with a triple of a transition, its binding and the time period before its firing. Thus the last parameter makes possible to define the time between two neighbour states. Using standard algorithms for finding the shortest or longest path between two nodes of the graph, one can find the minimum and maximum time of transition from one state to another.

RG is infinite due to the continuous reduction of the time stamps values. A similar situation arises during the analysis of almost every AL RTCP-net. Thus RG formalism is inconvenient for the net analysis. The introduction of additional conditions on the time stamps allows to transform an infinite graph into finite structure. Let say two states $m_1, m_2 \in M$ cover each other if next requirements are satisfied:

$$(m_1(p) = \langle \mu_1(p), \tau_1(p) \rangle \wedge m_2(p) = \langle \mu_2(p), \tau_2(p) \rangle \\ \wedge \mu_1(p) = \mu_2(p) \wedge (\tau_1(p) = \tau_2(p) \vee \tau_1(p) \leq -\delta_{max}(p) \wedge \tau_1(p) \leq -\delta_{max}(p))) \; .$$

Here $\delta_{max}(p) = max_{\langle p,t \rangle \in F}\{max_{\beta \in Bind_t} \varepsilon_{\mathrm{T}}(\langle p, t \rangle)_\beta\}$. The conditions of reflexivity, symmetry and transitivity are satisfied, hence the coverability relation is an equivalence relation ($\sim$) on the set $R(m_0)$. This relation is used to construct another net analysis tool—a coverability graph (CG) $CG = \langle \dot{P}, \ddot{\xi}, \dot{\varepsilon} \rangle$:

- $\dot{P}$ is an nonempty set of nodes;
- $\ddot{\xi} : \dot{P} \leftrightarrow R(m_0)/\sim$ is a bijective marking function;
- $\dot{\varepsilon} : \dot{P} \times \dot{P} \rightarrow T \times Bind \times \mathrm{T}_D$ is an arc function which maps each arc with a triple (Fig. 2).

Here $R(m_0)/\sim = \{\dot{R} | \dot{R} \subseteq R(m_0), (\forall m, m' \in \dot{R}) m \sim m'\}$ is a finite quotient set. Graphs RG and CG are constructed almost identically, the only difference is in the process of adding new nodes to the graph. Each node of CG is marked GA with an element of the set $R(m_0)/\sim$ wherein there are no two or more nodes that are marked with the same element and the number of nodes equal to the cardinality of the quotient set. CG for AL RTCP-net is always finite and provides the same capabilities of net properties analysis as RG.

To verify a technical system it's behaviour properties should be expressed formally with simple, concise and unambiguous propositional formulas. Logic statements with time-dependent values are used to specify such properties. Propositional logic is poorly suited to formulate the statements about the technical systems behaviour. Therefore temporal logics are commonly used in the verification of temporal structures. If there is no need to describe in detail the patterns of system behaviour and its objects interaction it's appropriate to apply not complex interval logic (like Allen's logic) but simple extensions of propositional logic like linear temporal logic (LTL) [8] or computational tree logic (CTL) [9].

A promising method of AL RTCP-nets verification is model checking (MC) [10]. There are highly effective verification MC algorithms that make possible to check the real industry hardware and software system. These algorithms check if the formula of LTL or STL that express some property of the dynamic system behaviour is true in a system model with a finite number of states. Kripke structure—a tuple $K = \langle C, C_0, H, AP, \lambda\rangle$—is normally used as such model. Kripke structure can be easily created on the basis of CG if the system is represented by AL RTCP-net:

- finite nonempty set of states C corresponds to the set of CG nodes $\dot{P}$;
- nonempty set of initial states $C_0 \subseteq C$ corresponds to the node of CG $\dot{p} \in \dot{P}, m_o \in \ddot{\xi}(p)$;
- set of arcs $H \subseteq C \times C$ corresponds to the set of CG arcs from the domain of function $\dot{\varepsilon}$;
- set of atomic propositions can be defined as $AP = \{\alpha_{\dot{u}}^{p} | \alpha_{\dot{u}}^{p} : M \rightarrow Bool\}$ where predicate $\alpha_{\dot{u}}^{p}$ for marking $m(p) = \langle \mu(p), \tau(p)\rangle \in M$, place $p \in P$ and multiset $\dot{u} \in U_p$ can be defined as follows:

$$\alpha_{\dot{u}}^{p}(m) = \begin{cases} true & if \quad \dot{u} \subseteq \mu(p), \tau(p) \leq -\delta_{max}(p); \\ false & otherwise; \end{cases}$$

- labeling function $\lambda : \dot{P} \rightarrow AP$ (since $C \equiv \dot{P}$) can be defined as follows: $(\forall \dot{p} \in \dot{P})(\lambda(\dot{p}) = \{\alpha_{\dot{u}}^{p} | \alpha_{\dot{u}}^{p} \in AP, (\forall m \in \ddot{\xi}(\dot{p}))\alpha_{\dot{u}}^{p}(m)\})$.

Thus AL RTCP-net verification using the MC algorithms is a natural extension of the initial net analysis using state graphs. A generalized diagram of the verification process is shown in Fig. 3.

3 AL RTCP-net Computer Simulation Tools Developing

Tools for computer modelling of processes in complex systems are relevant for both theoretical research and practical applications. Creating toolkit for solving such a problem is a serious challenge because the development environment should support the real-time concept and developed models should be flexible enough. However some systems like G2 complex (Gensym Corp., USA) allow the implementation of such projects.

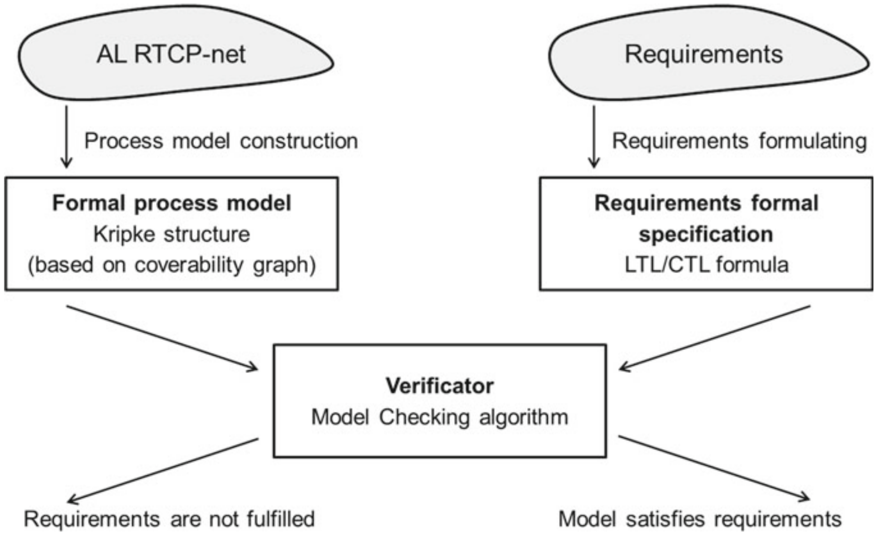


Fig. 3 AL RTCP-net verification process diagram based on Model Checking algorithm

G2 is an object-oriented IDE for developing and maintaining applications using the knowledge base. In contrast to the systems oriented to one methodology or a particular subject area G2 integrates many complementary techniques of artificial intelligence. This simplifies and speeds up development process and allows to make designed applications versatile enough. These features have allowed to develop in G2 IDE a prototype of a toolkit for modelling processes based on AL RTCP-nets [11].

The prototype development and the subsequent designing of process models with it confirmed the rationality of using modified CP-nets—AL RTCP-nets. However G2 environment significantly narrows the applicability of the developed software despite the undoubted ease of designing. Therefore toolkit full version is being developed in high level programming languages (C++, C Sharp) in Microsoft Visual Studio. Two tasks are performed in parallel: development of so-called "model" that includes the application functional logic and design of the graphical user interface (GUI) that allows user to create and tune process models. Modern graphics system for building the client applications Windows Presentation Foundation (WPF) is used in the GUI design which led to the decision to use Model-View-ViewModel pattern [12]. On the one hand ViewModel is a View abstraction and on the other hand it represents data of the Model. Classic Model-View-Controller pattern is not so suitable since WPF technology is based on the concept of data binding. Further it is planned to continue the work on the analysis and verification of complex dynamic system models on the basis of AL RTCP-net formalism using developed tools.

It is generally recognized that both testing and verification alone cannot guarantee a sufficient level of the developed systems correctness. Verification as well as testing has its advantages and disadvantages, however, these approaches can be considered mutually complementary. To improve the reliability of realizations in the development of control systems, software and hardware systems, both approaches should be applied.

4 Conclusions

In this paper the problem of using Petri nets to model and analyse the processes in complex dynamic systems is considered. The need of working with both types of temporal dependencies—quantitative and qualitative—is highlighted. A new subclass of Petri nets is proposed, real-time coloured Petri net with Allen temporal logic support - suitable for solving these problems. In the first part a formal definitions and an example of a simple system model developed on its basis are shown. In the second part authors gave a brief review of the model analysis and system verification necessity and proposed to use graphs of the net state changes to analyse and—as soon as they correspond to Kripke structure—to verify with Model Checking algorithms the models on the basis of described formalism. In the last part a developing computer tools for complex systems modelling was considered.

References

1. Vagin, V.N., Eremeev, A.P.: Some basic principles of construction of real-time intelligent decision support systems. J. Comput. Syst. Sci. Int. **6**, 953–961 (2001)
2. Eremeev, A.P., Korolev, Y.I.: Realization of real-time intelligent systems on the basis of Petri Nets with support of temporal dependencies (in Russian). J. Softw. Syst. **3**(103), 88–94 (2013)
3. Jensen, K.: Coloured Petri Nets. Basic Concepts, Analysis Methods and Practical Use, vol. 1–3. Springer, Heidelberg (1997)
4. Szpyrka, M., Szmuc, T.: Integrated approach to modeling and analysis using RTCPnets. IFIP Int. Fed. Inf. Process. **227**, 115–120 (2006)
5. Eremeev, A., Troickiy, V.: Temporal dependencies representation models in intelligent decision support systems (in Russian). In: Proceeding of the Russian Academy of Sciences. Theory and Control Systems, vol 5, pp 75–88 (2003)
6. Allen, J.F.: Maintaining knowledge about temporal intervals. Commun. ACM **11**(26), 832–843 (1983)
7. Murata, T.: Petri Nets. Properties, analysis and applications. In: Proceedings of the IEEE, vol. 77, No. 4, pp. 541–580 (1989)
8. Pnueli, A.: The temporal logic of program. In: Proceedings of the 18th Annual Symposium on Foundations of Computer Science (FOCS), pp. 46–57 (1977)
9. Clarke, E.M., Emerson, E.A., Sistla, A.P.: Automatic verification of finite-state concurrent systems using temporal logic specifications. ACM Trans. Program. Lang. Syst. **8**(2), 244–263 (1986)
10. Clarke, E.M., Emerson, E.A.: Design and Synthesis of Synchronization Skeletons Using Branching-Time Temporal Logic. In: LNCS, vol. 131, pp. 52–71. Springer (1982)
11. Eremeev, A.P., Korolev, Y.I.: Temporal models based on Petri Nets and intelligent systems development (in Russian). J. Inform. Models Anal. **2**(4), 336–344 (2013)
12. Gossman, J.: Introduction to Model/View/ViewModel pattern for building WPF apps. http://blogs.msdn.com/b/johngossman/archive/2005/10/08/478683.aspx

Fuzzy Graphs Clustering with Quality Relation Functionals in Cognitive Models

Alexander Tselykh, Vladislav Vasilev and Larisa Tselykh

Abstract In this study we present a new approach for developing input-output data set (antecedents, consequents) for fuzzy rules of expert system production based on the mechanism of fuzzy logic inference. Integration of the methods for cognitive modeling and of analysis with the expert system has been proposed. To generate the logical conclusion attributes on the basis of fuzzy graph models, the clustering procedure and the detection of system response are used. The approach, called Self-Constructing Attribute Generator, SCAG, consists in consecutive transformation of the initial matrix of the fuzzy graph model using two types of quality functionals. The first step is the initialization of the primary transformation matrix to upper-triangular sight using the square barrier penalty functions and "inverse" functions. At the second stage, the feedback on disturbances is generated in the form of a vector set. Further graph clustering is directly made based on the minimization of the potential energy functional.

Keywords Fuzzy clustering · Graph models · Quality functional · Cognitive modeling

---

This work was supported by the grant of Russian Foundation for Basic Research № 16-01-00098.

A. Tselykh (✉) · V. Vasilev
SFEDU, Taganrog, Russia
e-mail: ant@sfedu.ru

V. Vasilev
e-mail: vsvasilev@sfedu.ru

L. Tselykh
RSUE, Taganrog, Russia
e-mail: larisa@tgn.sfedu.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 450, DOI 10.1007/978-3-319-33609-1_32

1 Introduction

In the expert systems (ES) (including ES of the production type), based on mechanism of fuzzy logic inference, the process of knowledge acquisition is one of the main and most complicated stages in their structure [1, 2]. The goal of the modeling of the acquisition (extracting) process of domain experts' knowledge is one of the main bottlenecks of ES technology that determines their effectiveness [3–6].

The main methodological approaches used to extract the rules of logical inference, involving experts, are different groups of methods of questioning and interviewing. However, in some underdetermined, poorly formalized environments (management, social and economic system) there are situations, in which it is difficult, and sometimes impossible to directly (expertly) identify not only the attributes of the rules (consequents, antecedents), but also to highlight a specific problem task in the control system. Often the task is formulated only in general terms to the system, as a whole, without allocation of specific controlled parameters.

Automation of generation of input-output data set (antecedents, consequents) for fuzzy rules, especially in weakly formalized underdetermined environments, remains insufficiently studied. In such systems, the task of developing mathematical tools for processing the model of subject domain, generating the ES attributes, becomes apparent. The solution to this problem is proposed and based on the integration of cognitive modeling and analysis by an ES. One of the types of graph representations of the investigated system is the cognitive model presented in the form of Fuzzy Cognitive Maps. Fuzzy Cognitive Maps based on cause-and-effect relations (in the form of directed weighted edges) between elements (characteristics, concepts) of the controlled system are relatively easy to create. Nevertheless, this is time consuming and causes difficulties in the model validation (the adequacy, precision and accuracy of their settings).

A graphical representation of the model system behavior in the form of a fuzzy cognitive map allows to understand which system concepts affect other concepts and to what extent, however, the overall system behavior and selection of any key elements are difficult to understand [7]. This complication is due to the large graph dimensionality as well as the presence of numerous feedbacks (the complexity of concepts and relations between them). The size of modern graphs exceeds the capabilities of human perception, so the standard representation of the graph loses its informativity.

One of the possible solutions is multilevel representation of the initial graph via the union of the vertices in the group or community. It is important, however, to group vertices without changing the information content of a graph. The selection of communities allows to analyze and investigate the graph properties without losing information about its structure [8]. To generate the attributes of logical ES inference, based on the fuzzy graph model, using the procedures of clustering and detection of system response is a reasonable approach. Then, the problem of simultaneous selection of communities in the graph (i.e. vertices subsets) and causation (i.e. sequences of edges) becomes an urgent task.

In the present work, the problem of finding the order of index vertices, which achieves the optimal value of a quality functional, is solved. This allows clustering of fuzzy graph model, on the basis of which, the input-output data set of the expert system will be formed.

2 Related Work

In scientific works devoted to problems of method development for automatic generation of fuzzy logic rules, the following methods of intellectual data mining are presented: adaptive genetic algorithms; neural networks; fuzzy decision trees; subtractive clustering techniques, Takagi-Sugeno-Kang, TSK; fuzzy and hybrid methods; Mamdani algorithms, Multi-Agent algorithms; fuzzy formal concept analysis; an association rule extraction algorithm; DOC-BASED method; NextClosure algorithm; the parallel approach etc. In these investigations, the issues of optimization or establishment of the existing optimum set of fuzzy rules, on the basis of the attribute list received from the experts, are discussed.

In any case, solutions are offered based on the original input-output data set with defined relationships between them and with overlapping areas of interactions. The majority of the subject area tasks are developed for characteristics (attributes) that have possibilities to be numerically measured.

The theory of fuzzy graphs and hypergraphs is described systematically in the literature. There is a wide variety of graph clustering algorithms, depending on the problem description, on quality criteria laid down in the algorithms principles, and on the use for the processing of relevant data. Thus, the considered algorithms are based on:

- proximity measure or similarity, for example, the distance—Euclidean distance, Manhattan distance (city-block distance), the Chebyshev distance, the Mahalanobis distance, the exponential distance etc. [9];
- vertex colorings of graph with the different restrictions (the minimum number of the compact groups, specified number colors, given number of colors with a limit on the number of color-matching vertices etc.) [10];
- the use of fuzzy sets and the choice of transitive nearest connection (“greedy” optimization, multilevel functional optimization) [11];
- eigenvector of modularity matrix, including using some heuristics. [12, 13];
- computing of fluxes [14, 15]; computing the shortest route [16–18]; the maximum occurrence of labels among adjacent vertices [19]; random walks [20]; selection of stable subsets of graph [21] etc.

The above algorithms are designed to solve only one of the required tasks.

Unlike the mutually exclusive actions of the above mentioned algorithms, the solution of the task of minimization of square quality functional through the unified action is capable of identifying both the community and the causation. Review of the problem with this position can lead to the new views based on the fractional rational functions preserving not only the continuity of the initial functional, but also the smoothness, and possibly, the convexity (under certain conditions).

The order of numbering of the n-vertices is assigned as a point in the n-dimensional real space. Different orders of the vertices numbering can have a different value for the graph. Therefore, the relevant function n of real variables is introduced—the quality functional, reaching the optimum value at the permutations of the vertices that are the most appropriate for this particular task. It allows to involve the functions optimization of several real variables and can have additional effects, for example, on the minimization (optimization) using the gradient methods. Each variable receives a new value, i.e. the direction that changes the value of all variables is produced. It can also serve as the basis for the organization of focused enumeration in the set of permutations of the numbering of vertices.

3 Methods

Let us consider a finite graph $G = \langle V, E \rangle$, where $V = \{\nu_1, \nu_2, \ldots, \nu_n\}$—a finite set of vertices, n is the number of vertices, $E = \{\langle \nu_j, \nu_k \rangle | \nu_j \in V, \nu_k \in V\}$ is a finite set of edges. Graph G corresponds to the adjacency matrix $A = \|a_{jk}\|_{n \times n}$, where the weight a_{jk} of edge $\langle \nu_j, \nu_k \rangle$ can express the existence of edge (a Boolean value), the multiplicity of edge, weight of edge, fuzzy measure of the adjacency of vertices ν_j and ν_k etc.

Graph can have internal structure. This will be expressed in the structuredness of the adjacency matrix. The adjacency matrix portrait depends on the sequence of vertex numbering. When changing the order $\langle 1, 2, \ldots, n \rangle$ of vertices numbering to $\langle i_1, i_2, \ldots, i_n \rangle$, the adjacency matrix portrait will be changed as well. Identification of structures in the adjacency matrix portrait is an identification of structures in the original graph. If to run an uninterrupted continuation of set of permutations $\langle i_1, i_2, \ldots, i_n \rangle$ of vertex indices in the set of real numbers $(x_1, x_2, \ldots, x_n) \in R_n$, then for the search of the best order, the methods of optimization of functions of several real variables $\Phi(x_1, x_2, \ldots, x_n)$ can be applied.

In this general case, the objective function of the optimization task considers not only the original data about graph, but also information about reached on the next iteration relative position of vertices.

It should be kept in mind that initially the goal is to find the best ordering of vertices in the graph, so the optimization process in the space of real numbers will be mapped to the set of permutation of vertex indices. Required mapping of real-valued solutions $(x_1, x_2, \ldots, x_n)$ to the set of permutation $\langle i_1, i_2, \ldots, i_n \rangle$ will manifest itself in the “shocks” that can change the ratio of the algorithm productivities. However, the aim is to keep the resulting algorithms within the $O(n^3)$ cost boundaries.

Generalization of the presented approach will be its application not only to the quadratic programming tasks, but also to the tasks of minimizing of a rather general type functionals. The existence of the solution of the minimization problem will be provided by finite cardinality of permutation set of the vertex indices in graph G, and not by limits below the functional values. In other words, problems of conditional optimization will be solved. A desirable requirement for the uniqueness of the considered problem solutions and for the independence of the process results optimization from the choice of initial approximations, is at least not a strict convexity of minimized functionals.

Consider a large dimension graph of causation, created as a result of the expert opinion survey that expresses influences in the control system. For the effective control it is proposed to simplify the graph by combining vertices with similar vectors of response to perturbations outgoing from the impact on the individual vertices of the graph into communities. To obtain the responses of these vectors, using the *back propagation method* is possible. But for the algorithm work there is a need for control system stability, expressed by a directed graph. Thus, it is necessary that the feedbacks that take place because of the existence of cycles in the graph must be not only negative, but the transmission coefficients of feedbacks should be within the boundaries of the allowable areas. This problem does not occur if the cycles in the graph are simply missing.

On the other hand, in the created ES there should be no cycles in the graph as well. If the graph has a cycle, then the nonzero elements of the adjacency matrix corresponding to graph edges will be located both above and below the diagonal of the adjacency matrix. The opposite is true for the reverse. If the adjacency matrix of a directed graph has no nonzero elements above (or below) the diagonal, then cycles cannot exist in the finite graph. To exclude cycles (or to correct the transmission coefficients of the feedbacks) it is requested to give the adjacency matrix (as possible) to the upper (or lower) triangular form. Thus, if there are nonzero elements below (above) the diagonal, then the experts are asked to reconsider the value (or influence) of graph edges corresponding to these elements.

After that, there will be no problem to obtain response vectors for each vertex of the graph, and likewise the problem of community isolation is solved. Thus, an approach of consistent transformation of the initial matrix of the fuzzy graph model, using two types of quality functionals can be proposed, which essentially allows to generate the attributes of a production ES, called the Self-Constructing Attribute Generator, SCAG.

The complete SCAG algorithm is summarized as follows:

I. a. The selection of quality functional for the primary transformation matrix.
   b. The initialization of the primary transformation matrix M0 to upper-triangular form of the matrix M1.
   c. Examination of non-zero elements below (above) the diagonal matrix.
   d. Obtaining upper-triangular matrix M1.

II. a. The feedback response to perturbations in the form of a set of response vectors (matrix M2).
   b. Generation of the distance matrix M3.

III. a. The selection of the quality functional for the second transformation matrix—clustering.
b. Conducting of matrix M3 clustering with the formation of clusters $\{C_i\}$.
IV. Evaluation of the obtained results of clustering.

4 Results

Let us consider the procedure of bringing of the adjacency matrix M0 to upper-triangular form M1 on the basis of the quality functional choice.

Among square functionals consider the following:

$$\Phi_1(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk} \left(x_j - x_k\right)^2 \to \min. \tag{1}$$

The functional (1) can be interpreted as the potential energy of a mechanical spring system, in which a_{jk} is stiffness of spring, and $\left|x_j - x_k\right|$—tension of spring. According to one of the basic physics principles, a mechanical system tends to minimize its potential energy. As a result of minimization of functional (1), vertices connected by edges with a greater "spring stiffness" should receive neighboring indices in the index order, if possible. Thus, clusters or causation (or both, depending on the task to be solved) will be revealed. From the functional, the components of which are odd functions,

$$\Phi_2(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk} \left(x_j - x_k\right)^3 \to \min \tag{2}$$

we can expect that the indices i_k of vertices ν_k, for which quantity of incoming edges (that is increasingly claiming the role of consequence vertices) is predominant as a result of minimizing, will receive larger values $i_k > i_j$ than the indices i_k of the vertices ν_k, for which quantity of outgoing edges is predominant (that is increasingly claiming the role of cause vertices).

The difference between parity and odd of summands of functionals (1) and (2) manifests the opposite of priorities in the tasks of selection of communities adjacent to each other graph vertices and causation in the graph. But along with this, to the solution of tasks with the opposite priorities, a unified considered approach is applicable. Addends of functional:

$$\Phi_3(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk} \left(x_j - x_k\right) \to \min \tag{3}$$

are also odd functions, but the minimization by gradient methods, in contrast to the functional (2), will take into account only information regarding the graph without data on reached order of vertices.

If the selectivity of the functional (2) and (3) is insufficient, it is proposed to be increased by the use of functional

$$\Phi_4(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk} \mathrm{f}_\varepsilon(x_k - x_j) \to \min, \quad (4)$$

where

$$\mathrm{f}_\varepsilon(x) = \begin{cases} 1/x, & x > \varepsilon > 0, \\ (1/\varepsilon)\left(1 + (1 - x/\varepsilon) + (1 - x/\varepsilon)^2\right), & x \le \varepsilon. \end{cases} \quad (5)$$

Function $f_\varepsilon(x)$ can be considered as a barrier function or a penalty function, quickly increasing in $x_k - x_j \le 0 \le \varepsilon$, that is, if the vertex ν_k, for which the quantity of incoming edges (vertex-consequence) is predominant, receives a smaller value of the index $i_k < i_j$ than the index i_j of the vertex ν_j, for which the quantity of outgoing edges (cause vertices) is predominant. The expression $f_2(x) = (1/\varepsilon)\left(1 + (1 - x/\varepsilon) + (1 - x/\varepsilon)^2\right)$ is the decomposition in a Taylor's expansion of a function f $f_1(x) = 1/x$ in a neighbourhood of a point $x = \varepsilon$, and $\varepsilon > 0$ is the penalty parameter.

Additions of functionals (1)–(4) are "inverse" functionals:

$$\Phi_5(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk}\left((x_j - l)^2 + (u - x_k)^2\right) \to \min, \quad (6)$$

$$\Phi_6(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk}\left((x_j - l)^3 + (u - x_k)^3\right) \to \min, \quad (7)$$

$$\Phi_3^{'}(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk}\left((x_j - l) + (u - x_k)\right) \to \min, \quad (8)$$

$$\Phi_7(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk}\left(\mathrm{f}_\varepsilon(l - x_j) + \mathrm{f}_\varepsilon(x_k - u)\right) \to \min. \quad (9)$$

In these functionals the indices i_j of vertices ν_j, for which quantity of outgoing edges (vertices-causes) is predominant, as a result of minimization, tend to lower value of the indices l. The indices i_k of vertices ν_k, for which the quantity of incoming edges (vertices-consequences), as a result of minimization, tend to the upper value of the indices u. Possible values are the following pairs: ($l = 0$, $u = n - 1$), ($l = 1$, $u = n$), ($l = 0$, $u = n$), or other. Linear functionals (8) and (3) from the point of view of the minimization task are equivalent.

As an example, consider a graph with the following adjacency matrix:

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
v_1	0	0	0	0.15	0	0.75	0	0
v_2	0	0	0	0	0	0	0.90	0
v_3	0.60	0.10	0	0	0	0	0	0.25
v_4	0	0	0	0	0	0	0	0
v_5	0	0	0	0.55	0	0	0.40	0
v_6	0	0	0	0.45	0	0	0	0
v_7	0	0	0	0	0	0	0	0
v_8	0	0.85	0	0	0.30	0.70	0	0

The functional (1) is designed not to bring the adjacency matrix to upper-triangular view, but to identify communities, and therefore, it is not considered here.

As a result of minimization of functional (2) by changing the order of the vertices the following four edges remain below the diagonal: $\langle v_3, v_8 \rangle$; $\langle v_3, v_1 \rangle$; $\langle v_2, v_7 \rangle$; $\langle v_6, v_4 \rangle$:

Φ_2:	v_8	v_1	v_3	v_5	v_7	v_4	v_2	v_6
v_8	0	0	0	0.30	0	0	0.85	0.70
v_1	0	0	0	0	0	0.15	0	0.75
v_3	0.25	0.60	0	0	0	0	0.10	0
v_5	0	0	0	0	0.40	0.55	0	0
v_7	0	0	0	0	0	0	0	0
v_4	0	0	0	0	0	0	0	0
v_2	0	0	0	0	0.90	0	0	0
v_6	0	0	0	0	0	0.45	0	0

As a result of minimization of functional (3) the one edge remains below the diagonal $\langle v_6, v_4 \rangle$:

Φ_3:	v_8	v_3	v_5	v_1	v_2	v_6	v_4	v_7
v_8	0	0	0.30	0	0.85	0.70	0	0
v_3	0.25	0	0	0.60	0.10	0	0	0
v_5	0	0	0	0	0	0	0.55	0.40
v_1	0	0	0	0	0	0.75	0.15	0
v_2	0	0	0	0	0	0	0	0.90
v_6	0	0	0	0	0	0	0.45	0
v_4	0	0	0	0	0	0	0	0
v_7	0	0	0	0	0	0	0	0

Functionals (4), (6), (7) and (9) due to minimization lead to the solution, meanwhile, the results of the minimization of the functionals (4) and (9) and the results of the minimization of the functionals (6) and (7) match each other. The difference between these pairs of functionals is a valid permutation of some vertices:

$\Phi_{4,7}$:	v_3	v_8	v_1	v_5	v_2	v_6	v_7	v_4
v_3	0	0.25	0.60	0	0.10	0	0	0
v_8	0	0	0	0.30	0.85	0.70	0	0
v_1	0	0	0	0	0	0.75	0	0.15
v_5	0	0	0	0	0	0	0.40	0.55
v_2	0	0	0	0	0	0	0.90	0
v_6	0	0	0	0	0	0	0	0.45
v_7	0	0	0	0	0	0	0	0
v_4	0	0	0	0	0	0	0	0

In another example, the matrix 75×75, which has 411 edges with a total weight of 216.5, is considered. 219 edges with the sum of the weights of 109.3 were located initially below the diagonal. Obviously the graph contained cycles. Remaining after minimization of functionals the quantity of edges below the diagonal and their total weight is given in Table 1.

It is interesting to note that "worst" result of minimization of "inverse" functionals (6), (7) and (9) turns out to be a better than the "best" result of the minimization of the functionals (2)–(4). Also note that one iteration of minimization of the functionals (4) and (9) takes more time due to the computation of values of the penalty function (5). Therefore, for graphs of large dimension the choice in favor of the functional (6) showing the results coinciding with the results of the minimization of the functional (9) can be crucial. Please note also that the functional (6) is square, which confirms the above mentioned positive feedback in favor of square programming tasks. Subsequent clustering of the resulting matrix showed six clusters, and two of them have their own internal structure.

The obtained data can serve as a basis for the formation of ES attributes, because not only the reduction of analyzed structures dimension of the original graph model

Table 1 The result of minimization of functional in the matrix 75×75

Functional	Φ_2	Φ_3	Φ_4	Φ_5	Φ_6	Φ_7
Edges number	135	105	139	98	103	98
Edges weight	68.6	51.7	69.5	47.2	49.6	47.2

is reached, but also the relevant solution for highlighting the key system elements and respective constituent communities is achieved.

5 Discussion

5.1 Evaluation of SCAG Approach

There are five different criteria for the evaluation of our approach as follows:

1. *Applicability for the subject area.*

Studies have shown the applicability of this approach using quality functionals for the analysis of cognitive models, representing the directed graphs of causation.

2 *The feasibility.*

Functionals (1), (4), (6) and (9) are familiar-defined (i.e. restricted below by zero values) and convex, so the minimization even in the formulation of the problem of unconstrained optimization has a unique solution, achievable by any methods of the 0th and 1st order. Methods of the 2nd order will lose their efficiency because of the need to display a real-valued solution to integer permutation. Functionals (2), (3), (7) and (8) are unlimited, however, as a constant display of real-valued solutions on integer permutations takes place, the problem also has a solution, achievable with the same methods.

3. *Cluster Format.*

As a result of SCAG procedures, there is a selection of clusters that represents local communities of vertices, corresponding to the key concepts of the system. SCAG has no predefined number of clusters, and respectively, number of attributes. The algorithm performs clustering to obtain clusters "as they are". The size of the clusters is reduced to a reasonable quantity of vertices and edges, which allows to achieve an acceptable informative level of the obtained graph model for the expert (decision-maker).

4. *Functional Complexity of the approach.*

Regardless the used computer processor, the time to solve the task for a matrix 75×75 takes no more than a split second. Minimization of all functionals occurs monotonically and does not require user intervention. The user functions are switching from one functional to another. Therefore, SCAG is computationally efficient, corresponding to $O(n^3)$ costs.

5. *Quality of the selected clusters.*

The clustering quality is expressed in the degree of trust for the obtained input-output data sets (attributes). The obtained clusters can be evaluated for accuracy, fidelity, and comprehensibility of key attributes included in the selected clusters. As a criterion, defining the border of the cluster, the procedures of changing the sign for the gradient projection of the minimized functional are used. Due to the fact that the indices of vertices have taken integer values, the pair of vertices on the borders of neighbouring clusters will tend each to their cluster, whence the change of projection gradient sign comes from. Then, the input (antecedents) output (consequent) data sets is received in the process of further analysis of local clusters. These sets form a fuzzy rough model of the key system factors, subjecting to further analysis (adaptation) for a given system.

6 Conclusion

The conducted experiment showed the efficiency for the matrices of $10^3 \times 10^3$ size, however, for the matrices of very large dimension (big data) the solution of the problem of accumulation of rounding errors during floating-point calculations will be required. The integer calculations, with the actual conversion of fuzzy measures into the appropriate integer levels, can serve as a solution to this problem. Nevertheless, the issue of a significant narrowing of the range of computer representable figures will appear.

References

1. Gaines, B.R., Shaw, M.L.G.: eliciting knowledge and transferring it effectively to a knowledge-based system. In: IEEE Transactions on Knowledge and Data Engineering, vol. 5, no. 1, pp. 4–14. IEEE Press, New York (1993)
2. Nasuti, F.W.: Knowledge acquisition using multiple domain experts in the design and development of an expert system for disaster recovery planning. Doctoral Thesis Proposal. Nova Southeastern University (2000)
3. Feigenbaum, E.A.: The art of artificial intelligence: themes and case studies of knowledge engineering. In: Proceedings Fifth International Joint Conference on Artificial Intelligence, pp. 1014–1029. Morgan Kaufmann, Los Altos, CA (1977)
4. Hayes-Roth, F., Waterman, D.A.D., Lenat, B.: Building Expert Systems. Addison-Wesley, London (1983)
5. Bing, W., Chenyan, Z.: Dynamics of knowledge acquisition via e-learning community. JCIT **8**, 168–175 (2013)
6. Kadhim, M.A., Alam, M.A., Kaur, H.: A multi-intelligent agent architecture for knowledge extraction: novel approaches for automatic production rules extraction. J. IJMUE **9**(2), 95–114 (2014)
7. Tselykh, L.A., Panfilova, E., Prichina, O.C., Karasheva, A.G., Karanashev, A.K.: Methods of fuzzy set theory in the purpose of expansion of the value chains based on the main factors of corporate culture. J. MCSER **6**(5), S3, 249–258 (2015)
8. Henry, N., Fekete, J.-D.: Matrix explorer: a dual representation system to explore social networks. IEEE Trans. Vis. Comput. Graph. **12**(5), 687–684 (2006)
9. Kormen, T.: Algorithms, Construction and Analysis. Williams, Moscow (2005). Кормен, Т

10. Gorbatov, V.A., Smirnov, M.I., Khlytchiyev, I.S.: Logical Control of Distributed Systems. Energoatomizdat, Moscow (1991)
11. Kofman, A.: Introduction to the fuzzy sets theory. Radio and communications, Moscow (1982)
12. Newman, M.E.J.: Finding community structure in networks using eigenvectors of matrices. Phys. Rev. E **74**(3), 036104 (2006)
13. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech.: Theory Exp. **10** 10008 (2008)
14. Pons, P., Latapy, M.: Computing communities in large networks using random walks. JGAA **10**(2), 191–218 (2006)
15. Ye, Y., Jiang, Z., Diao, X., Du, G.: Extended event-condition-action rules and fuzzy Petri nets based exception handling for workflow management. Expert Syst. Appl. **38**(9), 10847–10861 (2011)
16. Girvan, V., Newman, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA **99**(12), 7821–7826 (2002)
17. Zhang, L., Du, W., Song, J.-J., Xiang, H.-Y.: Traffic assignment method based on dynamic fuzzy control and fuzzy decision making. Jiaotong Yunshu Gongcheng Xuebao. J. Traffic Transp. Eng. **10**(3), 110–117 (2010)
18. Kristianto, Y., Gunasekaran, A., Helo, P., Hao, Y.: A model of resilient supply chain network design: a two-stage programming with fuzzy shortest path. Expert Syst. Appl. **41**(1), 39–49 (2014)
19. Radhavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Phys. Rev. E **76**(3), 036106 (2007)
20. Goyal, S., Grover, S.: A fuzzy multi attribute decision making approach for evaluating effectiveness of advanced manufacturing technology—in Indian context. IJPQM **11**(2), 150–178 (2013)
21. Hipel, K.W., Kilgour, D.M., Bashar, A.M.: Fuzzy preferences in multiple participant decision making. J. Sci. Iranica **18**(3), 627–638 (2011)

Context-Based Trip Planning in Infomobility System for Public Transport

Alexander Smirnov, Nikolay Teslya, Nikolay Shilov and Alexey Kashevnik

Abstract During the urbanization process, many cities face the problem of comfortable and safe transportation for their inhabitants. One of solutions of this problem is an intelligent transportation systems (ITS) development: from simple personal navigation services to automated traffic management systems. A further development of ITS in connection with context management leads to the appearance of infomobility systems that provide access to personalized information, routes, and services based on the user's geographic location. The paper describes an approach to trip planning using public transport network. The approach applies a multigraph with dynamic edges' weighting based on the information about routes and context that describes current situation in the network. Context management and dynamic edges' weighting allows to find a route that satisfies user's preferences and to adjust the route in real time in case of current situation changes.

The presented results are part of the research carried out within the projects funded by grants # 14-07-00345, # 15-07-08092, # 15-07-08391, # 16-07-00462 of the Russian Foundation for Basic Research, and programs # I.5 and # III.3 of the Russian Academy of Sciences. The work has been partially financially supported by Government of Russian Federation, Grant 074-U01.

A. Smirnov · N. Teslya (✉) · N. Shilov · A. Kashevnik
SPIIRAS, 39, 14 Line, St. Petersburg, Russia
e-mail: teslya@iias.spb.su

A. Smirnov
e-mail: smir@iias.spb.su

N. Shilov
e-mail: nick@iias.spb.su

A. Kashevnik
e-mail: alexey@iias.spb.su

A. Smirnov · N. Teslya · N. Shilov · A. Kashevnik
ITMO University, 49, Kronverksky Pr., St. 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 450, DOI 10.1007/978-3-319-33609-1_33

Keywords Infomobility · Graph · Route planning · Context management · Public transport

1 Introduction

The population of modern cities increasingly refuses personal car ownership and advocates the usage of public transport, thus reducing the traffic on the road and contributing to the development of public transport. The development includes not only increasing the number of routes and vehicles, but also the developing of intelligent transportation systems (ITS) for internal control and services for users, that help them to plan their trips.

ITSs are aimed to provide innovative services to users (inhabitants of a city). These services are connected to different types of transport and traffic control systems and provide information about the current traffic conditions to the users. ITSs allow users to increase efficiency of using the existing transport networks as well as to make trips safer. ITS integrate telecommunication, electronic and information technology for planning, developing, supporting and managing city transport system. Using information and communication technologies in the field of road transport with its interconnection with other transport types allows making a significant contribution into increasing the transport usage efficiency, road transportation safety, mobility of passengers and freights, and reduction of the environmental impact [1].

Infomobility systems are a type of ITS. The "infomobility" term means that ITS provides access to personalized information and services based on the user's geographic location [2–4]. The main purpose of such systems is to increase user's mobility through information assistance at the stage of trip planning as well as during the trip. Therefore, infomobility systems provide wide range of services, from finding points of interests (POI), such as museums, attractions or restaurants along the user's route, to real-time notifications about road events such as accidents, timetable changes, etc. The latter type of the services is important for the multimodal transport network. In the multimodal route, a user can change several types of transport during one trip. For example, the user can pass a part of the trip on foot, reach nearest metro station by bus and use the metro for the rest of the trip.

One of the tasks that should be solved during the infomobility system development is a trip planning using public transportation network. The peculiarity of the problem is that the trip planning must be integrated with the context management to provide trip that would meet the user's needs in the current situation.

2 Related Work

Many infomobility systems have been developed to provide information support for travellers. Systems and studies in the field of infomobility can be divided into three main categories: (i) personal navigation systems: services for navigating the pedestrians, vehicle drivers and public transport passengers (Personal Travel Companion [5], Smart Travel Information Service (STIS) [6], Navitime [7]); (ii) environments for integration of infomobility resources (iTransit [6], Highway Traveler Information System in the Jiangsu Province of China [8], Arktrans [9]); (iii) context and location-based platforms using communications over next-generation networks (LoL@ [10], MapWeb [11], OnRoute [12], Im@gine IT [13]). The most interesting systems are described below.

Personal Travel Companion includes personalized multimodal trip planning, as well as management of multimodal trips. The prototype provides a multimodal trip planning "from door to door" with desktop or mobile device application. The trip can integrate pedestrian and road network routes as well as public transport usage. User settings (for example, the preferred modes of transport, mobility requirements, possibilities and time limits) are stored in the system for calculating personalized trips. The planned trip can be saved in personal history of trips.

Multimodal extensions in car navigation systems or personal assistants on the user mobile devices carry out information support during the trip. The continuous integral support is provided by mobile tools for multimodal trip planning on a variety of devices. The response in real-time is not guaranteed, and its dependency on the context is considered only in terms of services content adaptation to the user's start location, his/her preferences and available vehicles.

Smart Travel Information Service (STIS) offers travelers a multimodal travel planning service, aimed at bridging the gap in the coordination of existing transportation systems. Travel plan is created based on the preferences expressed by end-users, and integrates static and dynamic information about traffic jams and public transport.

STIS uses data provided by the iTransit environment for infomobility integration, to provide users personalized travel plans. The service is available on a personal computer or mobile phone. The context includes information about traffic and public transport, while the use of user context (e.g., his/her location) is not a major feature in the service. Besides, the service is focused on a trip planning, rather than traveller support during the trip. The information in accordance to the mobility conditions can be obtained in real time by integration of traffic and transport monitoring through the iTransit environment. STIS does not support dynamic trip adjustment based on real-time information, such as public transport delays and traffic jams.

Im@gine IT system serves to mobile users usually with help of mobile phone or onboard vehicle information system via wireless networks. It utilizes information from GPS, GIS, routing and mapping systems, and various kinds of distributed content, such as public transport timetable, traffic information, etc. In this system, the user may, for example, to find a route from one location to another, view it on a map and add it to a bookmark.

The system provides a platform for combining services, allowing to provide multimodal trip planning and to assess the current needs of the user, providing him/her support in different ways in different situations. For example, when the user is in the car, alerting is available in audio format instead of icons on the screen.

Presented systems allow the integration of multiple intelligent transportation systems that form a heterogeneous environment, which makes it possible to build multimodal trips, taking into account the current traffic situation. However, these systems require pre-processing of the route network, which imposes restrictions on using the context, and adjusting the route dynamically. In this paper an approach that allows trip planning using available information such as routes and timetables of public transport, taking into account the user's current situation and his/her preferences.

3 Contextual Information in Infomobility Systems

To support the user during the trip, the system has to constantly monitor the environment of the user and react to its changes. The environment is described by the context, which is formed by the user him/herself, and by various objects of the environment that are independent from the user. Contextual information that is needed for infomobility service often consists of the following parts:

- User type in the system (a tourist, a passenger, a driver, a fellow traveler, etc.). Since the infomobility system is aimed at multimodal trip support, the user type can dynamically change during a trip, and therefore, the system has to adjust to each type (use different notification types, showing or hiding different types of points of interest, etc.);
- User profile, which includes user personal preferences. Information should be delivered based on the user's preferences: audio of vireo notifications, museums of certain style, stores of certain brands and chains, etc.
- User current coordinates. All information is selected based on the current location of the user and addressed to support him/her in the current point.
- Current time and time-zone. The availability of specific vehicles, availability for visit the museums, shops, etc. depends on time.

In order to provide timely support to the user the context information collection should be carried out in real time, without the need for additional actions from the user. This makes it possible to reduce the results waiting time and increase the responsiveness of the service. In addition to responses to the context changes, services should track the development of the situation and gather additional information to offer services that are most likely to be useful in the current situation. This behavior is called proactive and the support of proactive mode is currently included in many information systems [14].

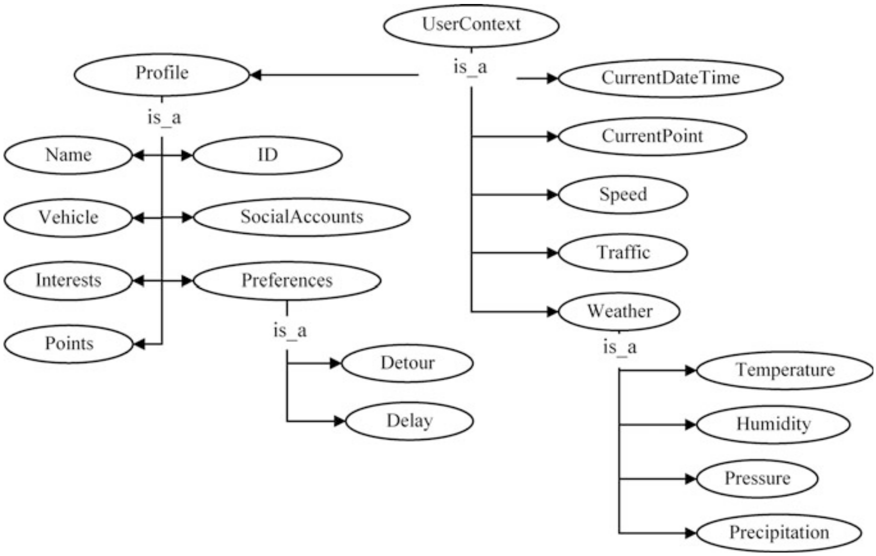


Fig. 1 User context structure

The context information is collected by using sensors installed on user devices. This allows gathering the physical features of the environment around the user. With the help of various GISs the system can obtain information about different objects on the map, routes that can be planned using different modes of transport, the weather forecast at specified locations, etc. Different social networks and information services allow the rating of the various objects (Fig. 1).

4 Dynamic Multigraph for Public Transport Network

The main transport type for a multimodal route planning is public transport. All other types of transport are often used to move between public transport stops or if the use of other type of transport is preferable for a given trip optimality criteria. Multimodal routing allows creating a greater interest of the travellers to use public transport. Due to that, the number of vehicles in the streets can be reduced, and so do the frequency and density of traffic jams.

The timetable of public transport and the current situation on the roads should be taken into account while planning a trip using public transport. Accounting for timetables allows estimating the waiting time for the vehicle in the case of fixed and strictly abided timetable and make an initial trip planning. Accounting for the current situation refines the basic plan, allowing estimation of the waiting and transfer times at public transport stops.

Information about routes and timetables of public transport in cities is usually provided using the GTFS format (General Transit Feed Specification) [15] by the

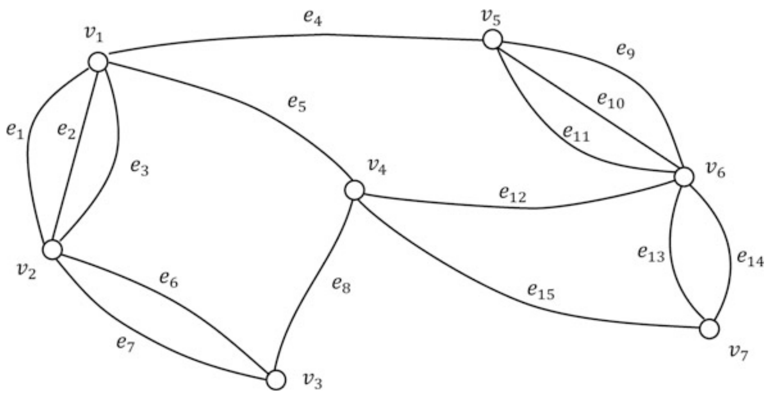


Fig. 2 Example of the multigraph that represents public transport network

relevant local authorities. Currently, this format is not the only way for the submission of the timetable data, but the ease of use and its extensibility allows suggesting that in the nearest future it will replace the majority of third-party formats.

The timetable of public transport network in GTFS format can be represented as a directed weighted multigraph $G(V, E, L)$, in which the vertices $V = \{v_1,\ldots, v_n\}$, n $\in$ N correspond to stops, and the edges $E = \{e_1,\ldots, e_m\}$, m $\in$ N—are routes connecting these stops, $L = \{l_1,\ldots, l_k\}$, k $\in$ N—weights of edges E (Fig. 2). The number of edges connecting vertices corresponds to the number of routes between the respective stops.

In case of transfer (public transport route changing), nodes are connected with dummy edges, reflecting the pedestrian routes. For edges' weighing two parameters are used: time (l_t) of moving between the stops of the edge e_i and cost (l_c)—ticket price for moving along the edge e_i. In both cases, the weights are dynamic that allows reflecting the current situation in the public transport network.

While planning a multimodal route the criteria related to the minimum distance is only important in the case of transfers that require the user to walk between transfer stops or ride a bicycle. In other cases, the criteria for selecting the optimal trip can be a minimal amount of travel time, minimal cost, or minimal number of transfers. However, it should be taken into account that a trip that meets one criterion may be inappropriate according to other criteria. Hence, it is needed to find a trip based on several criteria, solving the problem of multi-criteria optimization.

5 Dynamic Trip Planning Based on Multigraph and Context Management

It is rather hard to estimate how optimal a trip is according to a criterion during the trip planning using the multigraph, due to the dynamic nature of the edges' weights. Therefore, it is impossible to use search algorithm for shortest path that uses heuristics, like A* algorithm [16]. One of the algorithms that can guarantee obtaining the result in the given set of conditions is Dijkstra's algorithm [17]. To find the route through the multigraph with dynamic weights based on the criteria defined above the modification of the Dijkstra's algorithm has been proposed: the choice of vertices at each step is carried out taking into account the type of vehicle, its route and timetable (see Listing 1).

The algorithm modification allows taking into account features of formation of multigraph edges weights. It consists in changing the principle of vertices tagging in the graph. Each node is labeled with a timestamp at which the user expected to be in the node and the number of stops required to reach the node. Transfer is considered to take place if an edge e_i is chosen, and $e_i.rt \neq e_{i-1}.rt$, where $e_i.rt$—number (ID) of the new public transport route, $e_{i-1}.rt$—number (ID) of the public transport route used to pass the edge e_{i-1}.

Algorithm parameters are:

- Graph (V, E, L)—multigraph for public transport routes;
- *rt(Num, Type, Cost)* $\in$ *RT*—public transport route. Includes route number, transport type and transportation cost;
- *u(RT** $\in$ *RT, T(RT*)), v(RT** $\in$ *RT, T(RT*))* $\in$ *V*—graph vertices (public transport stops including routes and its timetables);
- *T(rt, v)*—arrival time of route *rt* to stop *v*;
- *e(u, v, rt)* $\in$ *E*—graph edge that represents route *rt* between nodes *u* and *v*;
- *time(e)* $\in$ *L*, *cost(e)* $\in$ *L*—weights of edge *e(u, v, rt)* that represents travel time and cost.

Program listing 1. Dijkstra's algorithm for multimodal route planning

```
while u != end do  // do while end point is not reached
  u = minT(rt)(u)     // node choosing by the minimal time
                                                  criteria
  for each neighbor v of u do // u,v - search through
                                        neighbourhood of u
    for each e(u,v,rt) do  // check al routes between u, v

      if rt != u.rt do     // route is not matches to the
                              minimal in node u
        alt_time = u.time + time(e) + (T(u.rt,u) -
                                               T(rt,u))
        alt_cost = u.cost + cost(e)// travel cost
      else do                   // using the same route
        alt_time = u.time + time(e)
        alt_cost = u.cost
        if alt_time < v.time do  // new route is faster
          v.time = alt_time
          v.prev_time_stop = u
          v.cost = alt_cost
          v.rt = rt
end
```

A transfer can be done in the current node or can require travelling between nodes on foot or by bicycle. In the former case, the time required to achieve the next node is calculated based on available timetable and the current traffic situation using the following equation:

$$T(v,\ e_j) = T(u,\ e_i) + time(e_j) + (T(u,\ e_i) - T(u,\ e_k)), \tag{1}$$

where $T(u,\ e_i)$—user's arrival time to the node u by the edge e_i; $time(e_j)$—the estimated time of moving along the edge e_j; $T(u,\ e_i) - T(u,\ e_k)$—waiting time for the next route vehicle arrival.

In the latter case, a different type of transport (walking or bicycle riding) is used to move to the next node. That requires querying an appropriate service. Thus, if the transfer requires a pedestrian route between stops the number of transfers is incremented by 1, and when using a different type of transport—by 2. In this case, there is no waiting time of the new vehicle in the node, but time of reaching the next vertex is calculated taking into account the route received from another service (*time(e')*):

$$T(v,\ e') = T(u,\ e_i) + time(e') \tag{2}$$

The computational complexity of the modified algorithm includes a computational complexity of the Dijkstra's algorithm, which increased by the additional edges between vertices: $O(|V| log|V| \cdot |RT|)$, where $|V|$ is the number of vertices of the multigraph, and $|RT|$ is the number of routes between nodes. The nodes in the array of route nodes *prev_time[v]* determine the types of vehicles, route numbers and their point of intersection. The transfer stops are searched in the area of the routes intersection points.

6 Use Cases

Information about routes and timetables of public transport is distributed in public GTFS format. This format contains files that simulate the structure of the required database tables. Full description of the format is presented in [15]. Due to its structure the data in GTFS format can be imported into a database to allow quick search of routes and additional information about them.

The city portal of public transport of St. Petersburg [18] provides information about public transport routes in GTFS format. In addition to the timetable the portal provides the current position of the vehicles, which can be used to predict the time of arrival of a vehicle to the selected stop. For intercity trips planning, the Yandex. Schedule service is used. It provides the timetable for public transport between the cities through the REST API.

After the import of public transport routes and timetable data for St. Petersburg to a PostgreSQL database and creating the multigraph, the analysis of data has been

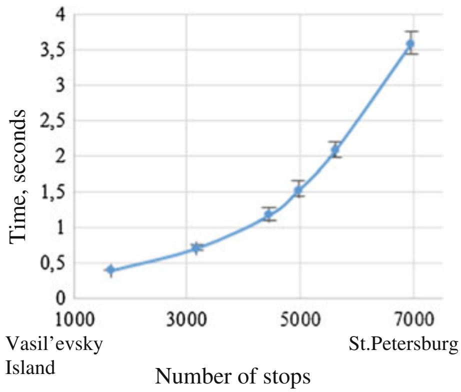


Fig. 3 Public transport routing time estimation using the modified Dijkstra's algorithm

carried out. There are 6962 stops in the city, between which 965 routes run. Number of multigraph edges used for displaying routes is 19773.

The performance of the modified algorithm has been estimated by searching a fixed route with a gradual increase of the size of the area from Vasilyevsky Island, where 1656 stops are located to entire St. Petersburg, where, as mentioned before, there are 6962 stop. At each of the stages, there have been 100 measurements carried out and average value and error range have been calculated. The analysis has been carried out based on the measurements defined by the 90 percentile (Fig. 3).

7 Conclusions

The use of information and communication technologies in the field of road transport and its relationship with other types of transport makes a significant contribution to more efficient use of transport and the environmental impact of a city, increase of road safety, as well as significant increase of the mobility of passengers and goods.

User support before the trip and during the trip is very important. Intelligent transportation system, which can provide for such a support, is called infomobile. Such systems should collect and analyze user context. Gathering context information from user device and from other sources allows to define user requirements in a particular situation, which contributes to the high quality information support.

Public transport has been selected as the main mode of transport for providing infomobility. This choice can be explained by extensive passenger capacity provided by all modes of public transport, as well as its efficiency compared to private cars.

Network of public transport routes has been modelled with a multigraph for user trip planing. In the multigraph, the vertices correspond to stops, and the edges—to the routes between stops. Each edge has a dynamic weight, based on the available context values. Weight may represent the time of movement along the edge, and a trip cost.

To plan a route by the multigraph the modified Dijkstra's algorithm has been used. The proposed modification allows to dynamically recalculate the weights of edges and to take into account the presence of multiple routes between nodes. The carried put tests carried out has shown that the proposed algorithm modification has acceptable efficiency for a megapolis like St. Petersburg.

References

1. Directive 2010/40/EU of the European parliament and of the council of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:207:0001:0013:EN:PDF (Accessed 08.02.2016)
2. Ambrosino, G., Boero, M., Nelson, J.D., Romanazzo, M.: Introduction. Infomobility Systems and Sustainable Transport Services. ENEA, Chapter 1, 340p. (2010)
3. Moraitis, P., Petraki, E., Spanoudakis, N.I.: Providing advanced, personalised infomobility services using agent technology. Appl. Innov. Intell. Syst. XI **2004**, 35–48 (2004)
4. Canali, C., Lancellotti, R.: A distributed architecture to support infomobility services. In: Proceedings of the 2nd International Workshop on Advanced Architectures and Algorithms for Internet Delivery and Applications, (Pisa, Italy), ACM International Conference Proceeding Series, vol. 198 (2006)
5. Rehrl, K., Bruntsch, S., Mentz, H.J.: Assisting multimodal travelers: design and prototypical implementation of a personal travel companion. IEEE Trans. Intell. Transp. Syst. **8**(1), 31–42 (2007)
6. Brennan, S., Meier, R.: STIS: smart travel planning across multiple modes of transportation. In: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, pp. 666–671 (2007)
7. Arikawa, M., Konomi, S., Ohnishi, K.: Navitime: supporting pedestrian navigation in the real world. Pervasive Comput. **6**(3), 21–29 (2007)
8. Xiang, Q.J., Ma, Y.F., Lu, J., Xie, J.P., Sha, H.Y.: Framework design of highway traveller information system of Jiangsu Province of China. Intell. Transp. Syst. IET **1**(2), 110–116 (2007)
9. Natvig, M.K., Westerheim, H.: National multimodal travel information—a strategy based on stakeholder involvement and intelligent transportation system architecture. Intell. Transp. Syst. IET **1**(2), 102–109 (2007)
10. Umlauft, M., Pospischil, G., Niklfeld, G., Michlmayr, E.: LoL@, a mobile tourist guide for UMTS. Inf. Technol. Tourism **5**(3), 151–164 (2003)
11. Huang, D., Liu, F., Shi, X., Yang, G., Zheng, L., Zhou, Z.: MapWeb: a location-based converged communications platform. Bell Labs Tech. J. **11**(1), 159–171 (2006)
12. García, C.R., Pérez, R., Lorenzo, A., Quesada-Arencibia, A., Alayón, F., Padrón, G.: Architecture of a framework for providing information services for public transport. Sens. (Switz.) **12**(5), 5290–5309 (2012)
13. Moraitis, P., Petraki, E., Spanoudakis, N.I.: An agent-based system for infomobility services. In: 3rd European Workshop on Multi-agent Systems (EUMAS2005), pp. 224–235 (2005)
14. Korzun, D., Galov, I., Balandin, S.: Proactive personalized mobile mutliblogging service on SmartM3. J. Comput. Inf. Technol. **20**(3), 175–182 (2012)
15. General transit feed specification (GTFS). Google Developers. https://developers.google.com/transit/gtfs/ (Accessed: 08.02.2016)

16. Hart, P.E.; Nilsson, N.J.; Raphael, B.: A formal basis for the heuristic determination of minimum cost paths. IEEE Trans. Syst. Sci. Cybern. **4**(2), 100–107 (1968)
17. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numer. Math. **1**(1), 269–271 (1959)
18. St. Petersburg Public Transport Portal http://transport.orgp.spb.ru/Portal/transport/main?lang=en (Accessed: 08.02.2016)

Processing of Qualitative Data

Vadim L. Stefanuk

Abstract To overcome some problems with deep understanding of fuzzy values, certain learning finite automaton was put into a fuzzy environment. Previously such a device has been studied in the probabilistic environment, where the classic technique of standard Markov chains was applicable. The new study became possible due to several previous results by the present author, namely the axiomatic of fuzzy evidences accumulation and the theory of generalized Markov chains. The mathematical results, obtained in the paper, prove that the learning automaton has the property of asymptotic optimality. This property is proposed to be used for measuring membership. The obtained results might lead to a fuzzy measurement procedure resembling statistics developed in probability area.

Keywords Fuzzy environment • Probabilistic environment • Finite automata with learning • Asymptotic optimality • Generalized markov chain • Fuzzy singletons

---

This work was partially supported by the Russian Fund for Basic Research (RFBR), Grant # 12-07-00209a, and by the Presidium of Russian Academy of Science, Program П15 and Program 1.5П.

V.L. Stefanuk (✉)
Institute for Information Transmission Problems RAS,
Bolshoi Karetny Per., 19, 101447 Moscow, Russia
e-mail: stefanuk@iitp.ru

V.L. Stefanuk
Peoples' Friendship University of Russia, Moscow, 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 450, DOI 10.1007/978-3-319-33609-1_34

1 Introduction

The exchange with quantitative information plays an important role in the interaction among people and technical devices. The temperature at home or in the street, the cost of goods or services—without those precise figures our life would be meaningless. In the theory of man-machine systems [1] the transmission of the quantitative information in the form of dial or menu readings for various crisp values does not create any problems.

However, people commonly use for communication the *qualitative information*, speaking on the temperature in the room in terms of *high*, *low*, *comfortable*, *suitable,* or reasoning on the cost of goods and services as *expensive*, *cheap*, *and reasonably priced* and etc. To support such qualitative information various schemes have been proposed, including Fuzzy Sets [1], Grey Numbers, Probability Theory [2] and more.

These schemes are used also in the flexible interfaces intended to establish a man-machine contact, which satisfies "both sides". Among various schemes the axiomatic of Fuzzy Sets Theory is the most developed and the most popular in applications. Yet, the formal analysis shows that many problems are still not completely resolved.

For instance, in Fuzzy Sets theory by Zadeh [1], which is extensively used in the area of Expert Systems [3, 4], the meaning of the fuzzy membership does not reach the level of transparency comparable to that of probability schemes. The Fuzzy Sets theory does not provide anything similar to the statistics.

Another problem is the lack of well understood mechanisms for defining fuzzy membership values as well as the mechanisms for direct understanding of obtained fuzzy values by a person. The use of Natural Language (NL) description does not resolve the problem. It is this problem that probably led Prof. L.A. Zadeh to formulation of the well established axiomatic theory for Fuzzy Sets [1].

The present paper provides an attempt to find some solution to above problems, using the mathematical model of a learning finite automaton, provided that it is being put into the Fuzzy Environment [5].

Our formal analysis is based on the two previous results. The first is our axiomatic for the problem of evidence accumulation [6]. This axiomatic has been used in our system SEISMO for prognosis of seismic phenomena. In the Fig. 1 the results are demonstrated that have been obtained with system SEISMO in case of middle term prediction of an earthquake.

The second important result is our generalization of the concept of Markov chains [7]. The classic Markov chain describes probabilistic events that changed in accordance with the Markov chain logic [8]. There are a number of various generalizations in the literature for the classic Markov chains, however the presence of some probabilistic phenomena is always assumed. Yet, generalization proposed in [7] is suitable for any chains with the Markov property and it does not necessary assume the probabilistic character of the values involved. To avoid

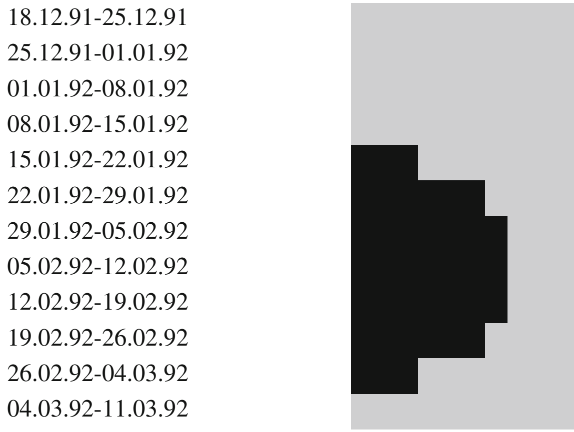


Fig. 1 SEISMO: Middle-term forecast (the week of the event)

misunderstanding in this paper our generalization is referred to as the *Markov-Stefanuk chain.*

The present paper is organized in the following way. Using the approach described in [6] in the first chapter an expression is derived for summary fuzzy effect of two fuzzy values: current automaton state and the penalty/reward obtained by the automaton. The mathematical expression has the form different from the one proposed by Zadeh [1], but in fact agrees with his logic.

Second chapter considers the behavior Linear Tactic automaton, which was introduced and studied by M.L. Tsetlin in a probabilistic environment, which penalize or rewarding the automaton with fixed probabilities [9]. However, in the second chapter it is being put into a fuzzy environment, which issues the rewards and penalties with some fuzzy membership functions. It is demonstrated how from assumed ergodicity of the generalized Markov chain [7] it is possible to obtain formulas, relating fuzzy values of inner states of automaton with its actions performed in the external fuzzy media.

In the third chapter the obtained formulas are used for inference some final assertions on behavior of the Linear Tactics Automaton (LTA). It is shown that in accordance with our expectation, the character of LTA behavior does not essentially changes due to transfer of LTA from probabilistic environment to the Fuzzy one.

In the fourth chapter some discussion is provided for the membership function used and why it is possible to work with *fuzzy singletons*. The fuzzy singletons differ from the common singletons known from literature [11] as they open some new possibility for defining values of unknown membership functions in certain classes.

In the fifth chapter it is shown that the obtained property of asymptotic optimality opens, in principal, the possibility for measuring fuzzy singletons, i.e. singletons with arbitrary membership values.

In Conclusion it is discussed that the experimental verification of obtained expressions meets some serious difficulty as unlike probability theory, where statistics gives certain grounds, the fuzzy set theory does not have theoretically confirmed experimental base. For this reason the results obtained and the steps used in this process are to day the only evidences of correctness of our analysis. From the other side, the obtained results provide us with the hope for building some analog of statistics, but oriented to the area of fuzzy systems.

For Markov-Stefanuk chain it is possible to consider the ergodicity property and to find final values in case this chain has one ergodicity class [7].

2 Collecting Evidences with Axiomatic Tools

Following the approach described in [7] an expression is derived for total fuzzy effect of two fuzzy values: current automaton state and the penalty/reward obtained by the automaton. The mathematical expression has the more "smooth" form, which is different from the maximum expression proposed by L. A. Zadeh, but in fact follows his theory logic. Our expression simplifies the equations, describing the behavior of fuzzy system in a fuzzy environment.

Let us consider first very simple learning machine shown in Fig. 2. It is a finite automaton, which has only two inner *states* (1, 2) and two external *actions* (**1**, **2**). This automaton is being put into some environment giving to it a *feedback*, which consists from *rewards* or *penalties,* depending on the actions performed by the automaton.

The upper part of this figure shows transition between its states, when it obtained a penalty when it was in the state 1 (left state in this figure), performing action **1**, or obtained the penalty, when it was in the state 1 (right in the figure), performing the action **2**.

The bottom graph shows transition between its states, when it obtained a reward when it was in the state 1, performing action **1**, or obtained the reward, when it was in the state 1, performing the action **2**.

2.1 Fuzzy Environment

When this automaton in a fuzzy environment it comes to its inner states in accordance with some fuzzy scheme, when $\mu_1^{(1)}(t)$ denotes the membership function to the state 1, corresponding to the action **1** in the moment $t = 1, 2, \ldots$. The same

Fig. 2 The simplest learning automaton with two actions

way one may define $\mu_1^{(2)}(t)$—the membership function to the state 1, that corresponds to the action **1** in the moments t = 1, 2,….

Let $\lambda^{(1)}$ is a fuzzy singleton meaning the penalty for the action **1** performed in the state 1, and $\lambda^{(2)}$ be a fuzzy singleton meaning the penalty for the action **2**, performed in the state 2.

As the automaton does not change its state, when the feedback is neutral, it is natural to consider that the singleton value $(1-\lambda^{(1)})$ corresponds to the reward for the action **1** performed in the state 1, and $(1-\lambda^{(2)})$ is the reward for the action **2**.

As the penalty/reward at the moment t defines *the next time* fuzzy states $\mu_1^{(i)}(t+1),\ i=1,2,$ one may be sure that $\mu_1^{(1)}(t)$ and $\mu_1^{(2)}(t)$ *do not depend* on the feedback at the moment t. Hence, following Prof. L.A. Zadeh “the total fuzzy value” is defined with the following classic expression:

$$f(x,y)=\min\{x,\ y\},\ \text{where x }=\mu_1^{(i)}(t);\ y=\lambda^{(i)}, i=1,2. \tag{1}$$

However this expression is difficult for analytic study of our automaton behavior. Hence taking the axiomatic of evidence summarization from our publication [6], and two new axioms are added, concerning the total result, namely:

$$f(0,0)=f(1,0)=f(0,1)=0 \tag{2a}$$

$$f(1,\alpha)=f(\alpha,1)=\alpha \tag{2b}$$

It is easy to see that these axioms are true for the above formula of L. A. Zadeh. However, it important for us that they are also valid for the next expression for the total sum:

$$f(x,y)=x\times y, \tag{3}$$

where x $=\mu_1^{(i)}(t);\ y=\lambda^{(i)}, i=1,2.$ Fig. 3 shows graphically the difference (1), (3).

2.2 *Ergodicity*

It easy to guess that the behavior of our simplest automaton is described with an ergodic Markov-Stefanuk chain. The parameters of this generalized Markov-Stefanuk chain are not known, and the property of ergodicity will be proved later, if it would be possible to find stationary values satisfying the property of equality of flows [7].

Due to the expected ergodicity the *final* singleton meaning transition from the left state to right state should be equal to the final singleton meaning transition from right state of automaton to the left its state:

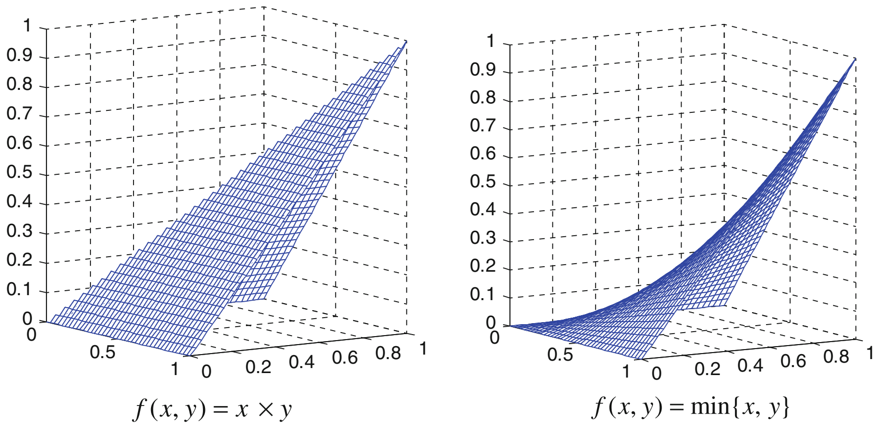


Fig. 3 The comparison of expressions (3) and (1)

$$\mu_1^{(1)}\lambda^{(1)} = \mu_1^{(2)}\lambda^{(2)} \tag{4}$$

Hence, if, $\lambda^{(1)} \geq \lambda^{(2)}$ then $\mu_1^{(1)} \leq \mu_2^{(2)}$. It means that our simplest automaton is able to learn to reduce the number of punishments!

Using same technique one may study the automaton with linear tactics, proposed by Tsetlin [9]. The linear tactics automaton is an extension of our simplest automaton considered above.

3 Behavior of Linear Tactics Automaton

In all the states belonging to the left side in Fig. 1 the automaton performs the first action (**1**) and obtains the penalty (in the upper graph) and the reward (in the bottom graph). Similar situation is valid for the action (**2**).

The linear tactic automaton was studied in [9] in a *random environment*, which issues the penalties and rewards with some fixed probabilities.

In the present paper the situation is different from that studied by M.L. Tsetlin as now this automaton is put into a fuzzy environment and performs in correspondence with the fuzzy membership functions similarly to the simplest automaton described above.

Again it may be shown that the performance of the linear automaton is controlled with a Generalized Markov Chain [7], i.e. Markov-Stefanuk chain. It turned out possible to obtain the final expressions, describing the behavior of linear automata, taking into account also some our results from [6].

Let memory depth of the automaton in Fig. 4 is equal to n that means that the linear tactic automaton has exactly n inner states for each of its actions (**1** and **2**).

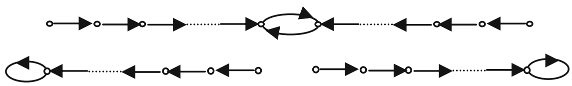


Fig. 4 State transitions for the LTA

Assuming ergodicity property one may obtain the following equations. (The states are numbered in the way that the "deepest state" in Fig. 4 has number *n*.)

$$\lambda^{(1)}\mu_i^{(1)} = (1-\lambda^{(1)})\,\mu_{i-1}^{(1)},\ \ i=1,\ldots,n, \tag{5}$$

where $(1-\lambda^{(1)})$—the reward for the action **1**, a $\mu_i^{(1)}$—is the membership function for the state *i* on the left line of Fig. 4, corresponding to the first action, i.e. action **1**.

Let us temporally fix values $\mu_1^{(1)}$ and $\mu_1^{(2)}$. Then from (4) and (5) one has

$$\mu_k^{(1)} = \mu_1^{(1)}\left(\frac{1-\lambda^{(1)}}{\lambda^{(1)}}\right)^{k-1} \text{ и } \mu_k^{(2)} = \mu_1^{(2)}\left(\frac{1-\lambda^{(2)}}{\lambda^{(2)}}\right)^{k-1}, k=1,\ldots,n \tag{6}$$

As the $\mu_1^{(1)}$ and $\mu_1^{(2)}$ are related with (4), one has

$$\mu_k^{(1)} = \mu_1^{(1)}\left(\frac{1-\lambda^{(1)}}{\lambda^{(1)}}\right)^{k-1} \text{ и } \mu_k^{(2)} = \mu_1^{(1)}\frac{\lambda^{(1)}}{\lambda^{(2)}}\left(\frac{1-\lambda^{(2)}}{\lambda^{(2)}}\right)^{k-1}, k=1,\ldots,n \tag{7}$$

If $M^{(1)}$—is the membership function (singleton) for the first action and $M^{(2)}$ is the membership function for the second action, then it may be shown [5] that the following expressions are valid for the final values (singletons):

$$M^{(1)} = 1-\prod_{k=1}^{n}(1-\mu_k^{(1)}) = 1-\prod_{k=1}^{n}(1-\mu_1^{(1)}\left(\frac{1-\lambda^{(1)}}{\lambda^{(1)}}\right)^{k-1}) \tag{8a}$$

$$M^{(2)} = 1-\prod_{k=1}^{n}(1-\mu_k^{(2)}) = 1-\prod_{k=1}^{n}(1-\mu_1^{(1)}\frac{\lambda^{(1)}}{\lambda^{(2)}}\left(\frac{1-\lambda^{(2)}}{\lambda^{(2)}}\right)^{k-1}) \tag{8b}$$

3.1 Asymptotic Optimality

From the last two expressions one has

$$\frac{M^{(1)}}{M^{(2)}} = \frac{1 - \left\{ \prod_{k=1}^{n} (1 - \mu_1^{(1)} \left(\frac{1-\lambda^{(1)}}{\lambda^{(1)}} \right)^{k-1}) \right\}}{1 - \left\{ \prod_{k=1}^{n} (1 - \mu_1^{(1)} \frac{\lambda^{(1)}}{\lambda^{(2)}} \left(\frac{1-\lambda^{(2)}}{\lambda^{(2)}} \right)^{k-1}) \right\}} \tag{9}$$

where $\mu_1^{(1)}$ is some positive value, satisfying $\mu_1^{(1)} \leq 1$ (details are in [5]).

As in the case of simplest automaton, one may observe the learning property of this linear tactics automaton. Indeed, if the reward for the first action is greater than the reward for the second action, i.e. $\lambda^{(1)} > \lambda^{(2)}$, then one obtains $M^{(2)} > M^{(2)}$.

Moreover, if in addition the following inequality holds

$$\frac{1 - \lambda^{(2)}}{\lambda^{(2)}} > 1 \tag{10}$$

then from (8a, 8b) under $n \to \infty$ the following is valid:

$$M^{(1)} \to 0, \; M^{(2)} \to 1,$$

i.e. Tsetlin's automaton with linear tactics being put into a fuzzy environment has the property of *asymptotic optimality* shown previously in [9] for the probabilistic environments.

4 Comments on Singletons

The membership functions used in this paper usually were referred to as *fuzzy points*, i.e. single element fuzzy sets. They may be found in many publications as *singletons* [10]. However usually for singletons it is assumed that membership function in this point is equal to 1.

Thus the paper [10] says that in realizations, where a fuzzy system is being built from some set of elementary functional elements, non-singleton fuzzification usually create problems. The author of [10] shows that application of de-fuzzification of the type DCOG may be modeled with the use of singleton architecture.

In the present paper a bit different concept of a singleton is used. It is referred to here as a fuzzy singleton. The fuzzy singleton may take any value from the interval [0–1]. Similar values have been used in Expert System MYCIN, where a simple heuristic rule was applied to combine evidences in favor of some phenomenon [3].

The use of fuzzy singletons corresponds to the membership functions from the general Fuzzy Set theory of Prof. Zadeh, presenting a convenient tool for numerical calculations. Indeed, the value of *feedback* represents the traditional fuzzy value λ, with the help of couple of "delta-functions λ_1, λ_0, provided that $\lambda_o + \lambda_1 = 1$ is true".

As it is assumed that then in our mathematical analysis it was natural to restrict with a single value, namely *fuzzy singleton for punishment* λ_0, and to use *fuzzy singleton* $(1 - \lambda_0)$ *for reward*. In the result the fuzzy singleton $\lambda_0^{(1)}$ is used for punishment feedback for the action **1** and the fuzzy singleton $\lambda_0^{(2)}$ is used for punishment feedback for the action **2**, or simply $\lambda^{(1)}$ and $\lambda^{(2)}$.

5 The Possibility to Measure the Level of Fuzziness

It was the use of fuzzy singletons that let us obtain the above expressions, describing final behavior of the linear tactic automaton designed by M.L. Tsetlin [7] when it was put into a fuzzy environment [5]. It follows that this finite automaton in our fuzzy environment has the important property of *asymptotic optimality* [9]. In other words, this automaton lets one to establish, which of the following relations is true $\lambda^{(1)} < \lambda^{(2)}$ or $\lambda^{(1)} \geq \lambda^{(2)}$.

First of all it is important to stress that the values, which define the penalties and rewards obtained by the finite automaton, *are not observable values*.

If the penalties or rewards are issued by a person, then these values are defined with a fuzzy considerations in his/her brain. In technical systems and for theoretical analysis these values may be logically deduced from some other non-observable factors.

From the other side, the values $\lambda^{(1)}$ and $\lambda^{(2)}$ are not observable in the same sense as $p(A)$—the probability of a certain event A is not observable either. Yet, in the probability theory one has an indirect way of approximate calculation of $p(A)$ by collecting statistics for the event.

Is there some possibility to make a fuzzy value known with some precision?

The positive answer to this question follows from important property of asymptotic optimality of automaton. The latter means this automaton allows to learn whether it is true that $\lambda_1 > \lambda_2$ or one has an opposite $\lambda_1 \leq \lambda_2$ under $n \to \infty$.

Hence the learning automaton described above lets one collect some "statistics". Indeed, such automaton lets one to define the position of unknown λ within the set of ordered values, $\left(\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(s)}\right)$. The ordering may be established by the several applications of the linear tactics automaton. The reliability of the ordering increases with the value of n in accordance with the property of its asymptotic optimality.

This ordering allows establishing the membership value with a prescribed precision. For instance, for this purpose the following ruler of "reference membership values" may be used such as (0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9).

Obviously this ruler and the described automata allow measuring some fuzzy value with 10 % precision. It was mentioned in our paper [12], where some other procedure was discussed as well ("Boston device").

Actually the automaton of Fig. 3 due to the relation (10) forces us to restrict with the following ruler of functions (0.5 0.6 0.7 0.8 0.9). This difficulty will be discussed elsewhere.

Presently it is not clear how to physically generate the penalties if the value λ is given. The procedures of fuzzifizction and de-fuzzification were aimed to it. However, this problem is removed if the penalties are created by a person, who is operating with the fuzzy values in the way he/she understands it [10]. The same person in exactly same way should then to define fuzzy values which are sent to the input of machine. It is important as otherwise the result obtained from some technical system may be incorrectly interpreted by a person.

Please note that in [10] a number of different ways of *understanding fuzziness* are demonstrated. Some problems may be avoided if there will be organized a preliminary tutoring of a group of users, involved in Man-Machine communication [13]. In the process of the tutoring they will develop common understanding of what is the fuzzy membership function.

Thus, it is understandable that a person may formulate a fuzzy value for some fact and say something like: "I believe that $\lambda^{(1)} = 0.4$" in order to reward the machine for its action **1**. Yet, it is not clear presently how the person comes to this decision.

The difficulty is the lack of the clear understanding how the value $\lambda^{(1)} = 0.4$ is used for feedback.

6 Conclusion

The expressions obtained in the paper show that the behavior of Tsetlin's automaton is asymptotically optimal. It means that the character of the behavior do not change very much due to transition from probabilistic environment to the fuzzy one. The theoretical results correspond to expectation and hence are rather reliable.

Unlike to Probability Theory which finds serious support in statistics, the Fuzzy Theory does not have yet experimental base, that is theoretically justified. That is why the obtained mathematical results and correctness of all our approach presently are the most convincing evidence correctness of the analysis demonstrated in this paper and correctness of our results.

From the other side in a real situation of exchange with technical systems the probabilistic scheme is not 100 % justified either. When one pushes the buttons on our learning machine [11] the person acts using some intelligent considerations, that are closer to some fuzzy considerations, not probabilistic ones: it difficult to imagine that a person keeps in his head a precise idea of a certain probability value, say 0.4.

Fuzzy sets theory has many questions that should be answered. And one may expect that the theoretical developments of the present paper possibly would make the Fuzzy Theory handier, more justified and more suitable for real life applications. Very important hope present author has with respect to the possibility of designing some analogy for the procedure of collecting statistics, but this time for the area of fuzzy systems.

References

1. Zadeh, L.A.: Fuzzy sets. Inf. Control, № 8, pp. 338–348 (1965) (USA)
2. Kolmogorov, A.N.: Zur Theorie der Markoffschen Ketten. Math. Ann. **101**, 126–136 (1929)
3. Shortliffe, E.H.: Computer-Based Medical Consultations: MYCIN. Elsevier, North Holland, New York (1976)
4. Stefanuk, V.L.: Dynamic expert systems. KYBERNETES Int. J. Syst. Cybern. **29**(5/6), 702–709 (2000) (MCB University Press, U.K.)
5. Stefanuk, V.L.: Behavior of Tsetlin's learning automata in a fuzzy environment. In: Second World Conference on Soft Computing (WConSC), pp. 511–513. Letterpress, Azerbaijan, Baku (2012)
6. Stefanuk, V.L.: Should one trust evidences? In: Proceedings of the All-Country AI Conference, vol. 1, pp. 406–410. Moscow (1988)
7. Stefanuk, V.L., Deterministic Markovian chains. Inf. Process. **11**(4), 702–709 (2011) (Moscow)
8. Romanovskii, V.I.: Discrete Markov Chains, p. 436. Гостехиздат Gostechizdat, Moscow (1949)
9. Tsetlin, M.L.: Some problems of finite automata behaviour. Doklady USSR Acad. Sci. **139**(4) (1961) (Moscow)
10. Munakata, T.: Fundamentals of the New Artificial Intelligence. Neural, Evolutionary, Fuzzy and More. Springer, USA (2008)
11. Stefanuk, V.L.: An example of collective behaviour of two automata. Autom. Remote Control **24**(6), 781–784 (1963) (Moscow)
12. Stefanuk, V.L.: Discovery of values of membership functions. In: VII International science and practice conference "Integral Models and Soft Computing in Artificial Intelligence". Fizmatlit, Kolomna, T.3, c. 1338–1343, Russia (2013)
13. Stefanuk, V.L.: On Man-Machine Interaction with Qualitative Data. In: Proceedings of 12th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, August 11–15 2013, Las Vegas, USA

An Approach to the Fuzzy Inference in Logical-Type Systems with Many Inputs

Vasiliy G. Sinuk and Maxim V. Panchenko

Abstract This paper describes an approach to the fuzzy inference in MISO-type systems in case when logical type system is used. It is shown that complex rules can be broken down into simple via represented implications when used max-min composition. It also shows that using of generalized modus ponens provides an efficient mechanism of inference with polynomial computational complexity. It is proposed to use this approach to create a neuro-fuzzy system solving the problem of diagnosis of rotary clinker burning kiln.

Keywords Logical-type reasoning · Generalized modus ponens · Max-min composition

1 Introduction

Over the last several decades fuzzy sets and fuzzy logic introduced in 1965 by Lotfi Zadeh [1] have been used in a wide range of problem domains including process control, image processing, pattern recognition and classification, management, economics and decision making. Specific applications include washing-machine automation, camcorder focusing, TV colour tuning, automobile transmissions and subway operations [2]. We have also been witnessing a rapid development in the area of neural networks. Both fuzzy systems and neural networks, along with probabilistic methods, evolutionary algorithms, rough sets and uncertain variables, constitute a consortium of soft computing techniques [3]. These techniques are often used in combination.

V.G. Sinuk (✉) · M.V. Panchenko
BSTU named after V.G. Shukhov, Belgorod, Russia
e-mail: vgsinuk@mail.ru

M.V. Panchenko
e-mail: panchenko.maks@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 450, DOI 10.1007/978-3-319-33609-1_35

The most popular designs of neuro-fuzzy structures fall into one of the following categories:

- Takagi-Sugeno method;
- Mamdani-type method;
- Logical-type reasoning method.

Mamdani method, which is currently popular is not free from shortcomings. However, it was emphasized by Yager [4] that “no formal reason exists for the preponderant use of the Mamdani method in fuzzy logic control as opposed to the logical method other than inertia.” Moreover, Yager said [5] that “as a matter of fact the Mamdani approach has some disadvantages: its inability to distinguish more specific information in the face of vague information and the requirement of having the antecedents of the rules span the whole input space.”

Let’s consider an approach to the fuzzy inference with polynomial computational complexity, based on compositional rule of inference for logical systems [6] with many inputs.

2 The Complexity of the Fuzzy Production with N-Inputs

The relation between inputs and outputs of the fuzzy production is described as follows:

$$\text{If } \beta_1 \text{ is } A_1 \text{ and } \beta_2 \text{ is } A_2 \text{ and} \ldots \text{and } \beta_n \text{ is } A_n \text{ then } \gamma \text{ is } B \tag{1}$$

where: A_i—term of input linguistic variable β_i, which is formalized by a fuzzy variable $\langle A_i, X_i, \tilde{A}_i \rangle$; X_i—range of definition of the fuzzy variable, $\tilde{A}_i = \int\limits_{X_i} \frac{\mu_{\tilde{A}_i}(x_i)}{x_i}, x_i \in X_i$ fuzzy set that describes a restriction for the value of the fuzzy variable A_i, $i = \overline{1,n}$. B—term of an output linguistic variable γ which is formalized by fuzzy variable $\langle B, Y, \tilde{B} \rangle$, Y—respectively the range of definition of the fuzzy variable, $\tilde{B} = \int\limits_{Y} \frac{\mu_{\tilde{B}}(y)}{y}, y \in Y$—the fuzzy set, where $\mu_{\tilde{B}}(y)$ is the possibility that the variable y takes «B» as the variable name.

The fuzzy production is formalized by (n + 1)-ary fuzzy relation $\tilde{R}$:

$$\tilde{A}_1 \times \tilde{A}_2 \times \cdots \tilde{A}_n \times Y \to X_1 \times X_2 \times \cdots \times X_n \times \tilde{B} \triangleq \int\limits_{\substack{\times X_i \times Y \\ i=\overline{1,n}}} \frac{\mu_{\tilde{R}}(x_1, x_2, \ldots, x_n, y)}{(x_1, x_2, \ldots, x_n, y)} \tag{2}$$

where $\to$ denotes one of the implications of multi-valued logic.

As follows from (2) complexity of (n + 1)-ary fuzzy relation $\tilde{R}$—$O(|X|^n * |Y|)$.

3 Statement

In the proof of this statement the max-min basis is used. If the membership function of a binary relation $\mu_{\tilde{R}}\left(\mu_{\tilde{A}_i}(x_i), \mu_{\tilde{B}}(y)\right), i = \overline{1, n}$ obtained by formalizing implication $\tilde{A}_i \rightarrow \tilde{B}$ does not increase by the argument $\mu_{\tilde{A}_i}(x_i)$, then the membership function of (n + 1)-ary fuzzy relation, which characterizes the implication:

(a) $\tilde{A}_i \, and \, \tilde{A}_2 \, and \ldots and \, \tilde{A}_n \rightarrow \tilde{B}$ has a property:

$$\mu_{\tilde{R}}(x_1, x_2, \ldots, x_n, y) = \bigvee_{i = \overline{1, n}} \mu_{\tilde{R}_i}\left(\mu_{\tilde{A}_i}(x_i), \mu_{\tilde{B}}(y)\right) \tag{3}$$

(b) $\tilde{A}_1 \, or \, \tilde{A}_2 \, or \ldots or \, \tilde{A}_n \rightarrow \tilde{B}$ has a property:

$$\mu_{\tilde{R}}(x_1, x_2, \ldots, x_n, y) = \bigwedge_{i = \overline{1, n}} \mu_{\tilde{R}_i}\left(\mu_{\tilde{A}_i}(x_i), \mu_{\tilde{B}}(y)\right) \tag{4}$$

The proof of the property (a)

Definition 1 Function $\mu_{\tilde{R}}\left(\mu_{\tilde{A}_i}(x_i), \mu_{\tilde{B}}(y)\right)$ is a non-increasing by the argument $\mu_{\tilde{B}}(y)$, if $\forall \mu_{\tilde{B}}(y) \in [0, 1]$, $\mu_{\tilde{A}}(x) \in [0, 1]$ from the condition $\mu_{\tilde{A}_i}(x_i) > \mu_{\tilde{A}_i}(x_j)$ follows the inequality $\mu_{\tilde{R}}\left(\mu_{\tilde{A}_i}(x_i), \mu_{\tilde{B}}(y)\right) \leq \mu_{\tilde{R}}\left(\mu_{\tilde{A}_i}(x_j), \mu_{\tilde{B}}(y)\right)$

Proof let's denote an arbitrary value

$$\mu_{\tilde{B}}(y) = b \in [0, 1], \tag{5}$$

Then the right part of the expression (3) takes the form:

$$\mu_{\tilde{R}}\left(\bigwedge_{i = \overline{1, n}} \mu_{\tilde{A}_i}(x_i), b\right) = \mu_{\tilde{R}_i}\left(\min_{i = \overline{1, n}} \mu_{\tilde{A}_i}(x_i), b\right)$$

Let's choose an arbitrary set of arguments $(x_1, x_2, \ldots, x_n) \in X_1 \times X_2 \times \cdots \times X_n$ and denote: $\mu_{\tilde{A}_k}(x_k) = \min_{i = \overline{1, n}} \mu_{\tilde{A}_i}(x_i)$, from this follows:

$$\mu_{\tilde{A}_k}(x_k) \leq \mu_{\tilde{A}_i}(x_i) \; \forall \, i = \overline{1, n}.$$

According to the Definition 1:

$$\mu_{\tilde{R}_k}\left(\mu_{\tilde{A}_k}(x_k), b\right) \geq \mu_{\tilde{R}_i}\left(\mu_{\tilde{A}_i}(x_i), b\right) \; \forall \, i = \overline{1, n}$$

That is the value of $\mu_{\tilde{R}_k}\left(\mu_{\tilde{A}_k}(x_k), b\right)$ is not less than any of the $\mu_{\tilde{R}_i}\left(\mu_{\tilde{A}_i}(x_i), b\right)$.

Consequently, it is not less than the fuzzy disjunctions (i.e. max) of these values. Since the right part of the inequality also contains $\mu_{\tilde{R}}\left(\mu_{\tilde{A}_k}(x_k), b\right)$, the inequality is transformed into equality

$$\mu_{\tilde{R}_k}\left(\mu_{\tilde{A}_k}(x_k), b\right) = \bigvee_{i=\overline{1,n}} \mu_{\tilde{R}_i}(\mu_{\tilde{A}_i}(x_i), b)$$

or

$$\mu_{\tilde{R}_k}\left(\bigwedge_{i=\overline{1,n}} \mu_{\tilde{A}_i}(x_i), \mu_{\tilde{B}}(y)\right) = \bigvee_{i=\overline{1,n}} \mu_{\tilde{R}_i}(\mu_{\tilde{A}_i}(x_i), \mu_{\tilde{B}}(y))$$

The first part of statement is proved.

The proof of property (b)

In view of (5) the right side of expression (4) takes the form:

$$\mu_{\tilde{R}}\left(\bigvee_{i=\overline{1,n}} \mu_{\tilde{A}_i}(x_i), b\right) = \mu_{\tilde{R}_i}(\max_{i=\overline{1,n}} \mu_{\tilde{A}_i}(x_i), b)$$

For an arbitrary set of arguments $(x_1, x_2, \ldots, x_n) \in X_1 \times X_2 \times \cdots \times X_n$

Let's denote: $\mu_{\tilde{A}_k}(x_k) = \max_{i=\overline{1,n}} \mu_{\tilde{A}_i}(x_i)$, from this follows: $\mu_{\tilde{A}_k}(x_k) \geq \mu_{\tilde{A}_i}(x_i)$ $\forall i = \overline{1,n}$.

According to the Definition 1:

$$\mu_{\tilde{R}_k}\left(\mu_{\tilde{A}_k}(x_k), b\right) \leq \mu_{\tilde{R}_i}\left(\mu_{\tilde{A}_i}\ (x_i, b)\right) \quad \forall i = \overline{1,n} \tag{6}$$

Thus, the value $\mu_{\tilde{R}_k}\left(\mu_{\tilde{A}_k}(x_k), b\right)$ is not greater than any of the $\mu_{\tilde{R}_i}\left(\mu_{\tilde{A}_i}(x_i), b\right)$ values, and it is not greater than the conjunction of all these values. Since the right part of the inequality (6) also contains $\mu_{\tilde{R}_k}\left(\mu_{\tilde{A}_k}(x_k), b\right)$, the inequality is transformed into equality, i.e.

$$\mu_{\tilde{R}_k}\left(\mu_{\tilde{A}_k}(x_k), b\right) = \bigwedge_{i=\overline{1,n}} \mu_{\tilde{R}_i}\left(\mu_{\tilde{A}_i}(x_i)\right), b)$$

or

$$\mu_{\tilde{R}_k}(\bigvee_{i=\overline{1,n}} \mu_{\tilde{A}_i}(x_i), \mu_{\tilde{B}}(y)) = \bigwedge_{i=\overline{1,n}} \mu_{\tilde{R}_i}\left(\mu_{\tilde{A}_i}(x_i), \mu_{\tilde{B}}(y)\right)$$

The second part of the statement is proved.

Exploring the membership function of the fuzzy binary relation

$$\mu_{\tilde{R}}\left(\mu_{\tilde{A}}(x), \mu_B(y)\right)$$

the following implications that satisfy to the Definition 1 have been identified:
Zadeh implication

$$\mu_{\tilde{A}\to\tilde{B}}(x,y) = \max[\min[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(y)], 1 - \mu_{\tilde{A}}(x)]$$

Lukasiewicz implication

$$\mu_{\tilde{A}\to\tilde{B}}(x,y) = \min[1, 1 - \mu_{\tilde{A}}(x) + \mu_{\tilde{B}}(y)]$$

Rescher implication

$$\mu_{\tilde{A}\to\tilde{B}}(x,y) = \begin{cases} 1, & \text{if } \mu_{\tilde{A}}(x) \le \mu_{\tilde{B}}(y) \\ 0, & \text{if } \mu_{\tilde{A}}(x) > \mu_{\tilde{B}}(y) \end{cases}$$

Gödel implication

$$\mu_{\tilde{A}\to\tilde{B}}(x,y) = \begin{cases} 1, & \text{if } \mu_{\tilde{A}}(x) \le \mu_{\tilde{B}}(y) \\ \mu_{\tilde{B}}(y), & \text{if } \mu_{\tilde{A}}(x) > \mu_{\tilde{B}}(y) \end{cases}$$

Kleene-Dienesimplication

$$\mu_{\tilde{A}\to\tilde{B}}(x,y) = \max[1 - \mu_{\tilde{A}}(x), \mu_{\tilde{B}}(\mathrm{y})]$$

Reichenbach implication

$$\mu_{\tilde{A}\to\tilde{B}}(x,y) = \min[1, 1 - \mu_{\tilde{A}}(x) + \mu_{\tilde{A}}(x) * \mu_{\tilde{B}}(y)]$$

Aliev implications [7]:
ALI2-logic

$$\mu_{\tilde{A}\to\tilde{B}}(x,y) = \begin{cases} 1, & \text{if } \mu_{\tilde{A}}(x) \le \mu_{\tilde{B}}(y) \\ \min[1 - \mu_{\tilde{A}}(x), \mu_{\tilde{B}}(y)], & \text{if } \mu_{\tilde{A}}(x) > \mu_{\tilde{B}}(y) \end{cases}$$

ALI3-logic

$$\mu_{\tilde{A}\to\tilde{B}}(x,y) = \begin{cases} 1, & \text{if } \mu_{\tilde{A}}(x) \le \mu_{\tilde{B}}(y) \\ \frac{\mu_{\tilde{B}}(y)}{((1 - \mu_{\tilde{B}}(y)) + \mu_{\tilde{A}}(x))}, & \text{if } \mu_{\tilde{A}}(x) > \mu_{\tilde{B}}(y) \end{cases}$$

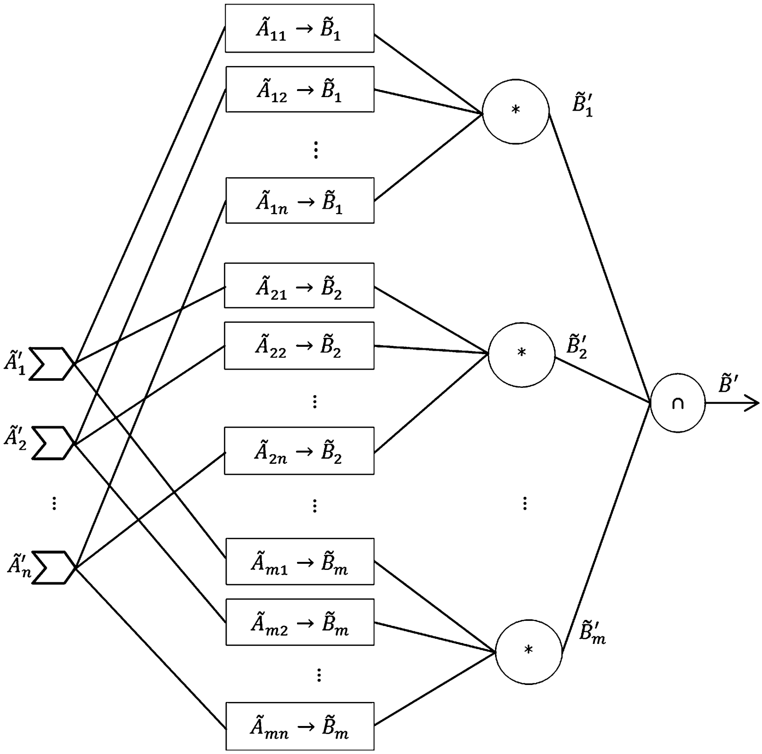


Fig. 1 Network structure for the block of rules

Based on the above proved it is possible to formulate a generalized modus ponens rule [8] for fuzzy production with n inputs (Fig. 1).

Definition 2 Let's suppose that fuzzy unary relations $R(x_1), R(x_2), \ldots, R(x_n)$ and $R(y)$, and also (n + 1)-ary fuzzy relation $R(x_1, x_2, \ldots, x_n, y)$ are fuzzy sets on the universes of discourse $X_1, X_2, \ldots, X_n, Y$ and $X_1 \times X_2 \times \cdots \times X_n \times Y$ and fuzzy sets $\tilde{A}'_1, \tilde{A}'_2, \ldots, \tilde{A}'_n$ on the base sets $X_1, X_2, \ldots, X_n$ and (n + 1)-ary relation, that formalizes an implication $\tilde{A}_1 * \tilde{A}_2 * \cdots * \tilde{A}_n \to \tilde{B}$ are given. The membership function of a binary relation corresponding to the implication $\tilde{A}_i \to \tilde{B}, i = \overline{1, n}$ satisfies the statement. Then compositional rule of inference confirms that the solution of assignment equations system with respect to $R(y)$ is:

$R(x_1, x_2, \ldots, x_n, y) \triangleq \tilde{A}_1 * \tilde{A}_2 * \cdots * \tilde{A}_n \to \tilde{B}$ knowledge

$$\left.\begin{array}{l} \tilde{R}(x_1) \triangleq \tilde{A}_1^{'}; \\ \tilde{R}(x_2) = \tilde{A}_2^{'}; \\ \quad \cdots \\ \tilde{R}(x_n) = \tilde{A}_n^{'}; \end{array}\right\} \text{fact}$$

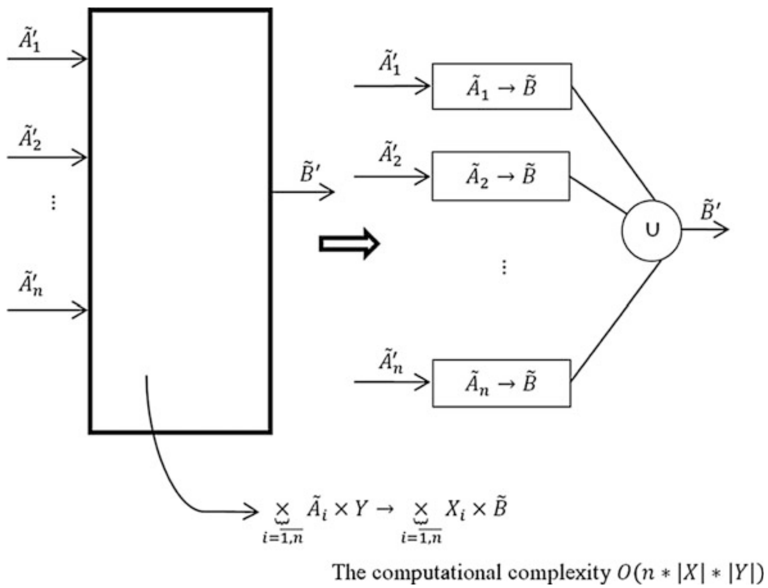
The inference looks like

$$\tilde{R}(y) = \tilde{B}^{'} = \bigcup_{i=\overline{1,n}} \tilde{A}_i^{'} \circ (\tilde{A}_i \to \tilde{B})]$$

If the option (b) is considered then

$$\tilde{R}(y) = \tilde{B}^{'} = \bigcap_{i=\overline{1,n}} [\tilde{A}_i^{'} \circ (\tilde{A}_i \to \tilde{B})$$

The graphic interpretation of the Definition 2 using the statement:



4 Calculation of Output for the Block of Rules

The rule block is a set of implications concerning terms of the output linguistic variable B, and is represented as

$$\begin{aligned} R^{(1)}&: \ \textit{If}\ \tilde{A}_{11} * \tilde{A}_{12} * \cdots * \tilde{A}_{1n}\ \textit{then}\ \tilde{B}_1 \\ R^{(2)}&: \ \textit{If}\ \tilde{A}_{21} * \tilde{A}_{22} * \cdots * \tilde{A}_{2n}\ \textit{then}\ \tilde{B}_2 \\ &\quad \cdots \\ R^{(n)}&: \ \textit{If}\ \tilde{A}_{m1} * \tilde{A}_{m2} * \cdots * \tilde{A}_{mn}\ \textit{then}\ \tilde{B}_m \end{aligned}$$

where m—is a number of terms of the output linguistic variable.

Considering the fact that the above described system refers to the logical type, the overall result of this conclusion is determined as [9]:

$$\tilde{B}' = \bigcap_{j=\overline{1,m}} \tilde{B}'_j$$

5 Conclusion

This paper considers an approach to fuzzy inference for logic-type fuzzy systems based on the modified compositional rule of inference. The input data can generally be either definite values or be represented by the fuzzy sets. The polynomial computational complexity of the proposed fuzzy inference allows the efficient using of it for solving different problems of modeling, diagnosis, forecasting in systems with many inputs.

This approach has been used in the development of a neuro-fuzzy system for identifying abnormal operating modes of the rotating cement kiln [10, 11]. Under the project there has been developed a neuro-fuzzy network based on the proposed modified compositional rule of inference. The setting of membership functions was carried out by a genetic algorithm [12]. There was also created a method for programming in Fuzzy Control Language (FCL) [13], which consists of a translator of tuned and trained network to the text in the standardized Fuzzy Control language, and also allows performing inverse transformation [14]. The corresponding software, which allows solving different problems of fuzzy modeling in a variety of problem-oriented fields, was developed [15].

References

1. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)
2. Hirota, K.: Industrial Applications of Fuzzy Technology. Springer Verlag, Tokyo (1993)
3. Aliev, R.A., Aliev, R.R.: Soft Computing and its Applications. World Scientific Publishing, Singapore-New Jersey-London-Hong Kong (2001)
4. Yager, R.R.: Fuzzy logic controller structures. Proc. SPIE Symp. Laser Sci. Optics Appl. 368–378, 1990
5. Yager, R.R.: A general approach to rule aggregation in fuzzy logic control. Appl. Intelligence **2**, 333–351 (1992)
6. Rutkowski, L., Cpałku, K.: Flexible Neuro-Fuzzy Systems. IEEE Trans. Neural Networks **14** (3), 554–574 (2003)
7. Aliev, R.A., Krivosheev, V.P., Liberzon, M.I.: Optimal decision coordination in hierarchical systems, News of Academy of Sciences of USSR. Tech. Cybern. **2**, 72–79 (1982)
8. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning. Inform. Sci. Part I, **8**, 199–249, Part II, **8**, 301–357, Part III, **9**, pp. 43–80 (1975)
9. L. Rutkowski Methods and techniques of artificial intelligence, Rutkowski, L.: [trans. from Pol. I. D. Rudinski], M.: Hot line Telecom, 520 p (2010)
10. Sinuk, V.G.: Algorithms and software tools to create intelligent problem-oriented systems based on fuzzy logic. Sinuk, V.G., Polyakov, V.M., Panchenko, M.V. Bulletin of BSTU named after V.G. Shukhov No 3 pp. 159–161 (2013)
11. Sinuk, V.G.: Diagnosis of abnormal operating conditions of the rotary clinker kiln based on neuro-fuzzy network. Sinuk, V.G., Polyakov, V.M., Panchenko, M.V. Devices and systems. Management, monitoring, diagnostics. No 9, pp. 42–48 (2014). http://elibrary.ru/contents.asp?issueid=1356763
12. Rutkowski, D., Pilinski, M., Rutkowski, L.: Neural networks, genetic algorithms and fuzzy systems: trans. from Pol. I. D. Rudinski. M.: Hotline Telecom, 452 p (2006)
13. International electrotechnical commission (IEC), technical committee no. 65: industrial process measurement and control sub-committee 65 b: devices IEC 1131 PROGRAMMABLE CONTROLLERS. Part 7—Fuzzy Control Programming
14. Sinuk,V.G.: Software for fuzzy modeling language using FCL. Sinuk, V.G., Polyakov, V.M., Panchenko, M.V. Bulletin of RSUR, No 3, pp. 117–120 (2011)
15. The certificate number 2015613935 of Russian Federation on the state registration of computer program. Neuro-fuzzy diagnostic system of abnormal operating conditions for the rotary clinker kiln. Panchenko, M.V., Sinuk, V.G., Polyakov, V.M., Buchanov, D.G.: the applicant and the right holder FSBEOHPE “Belgorod State Technological University named after V.G. Shukhov.” № 2015610657, req. 10.02.2015; pub. 31.03.2015

New Fuzzy Truth Value Based Inference Methods for Non-singleton MISO Rule-Based Systems

Vasily G. Sinuk, Vladimir M. Polyakov and Dmitry A. Kutsenko

Abstract At fuzzy modelling the inputs of fuzzy systems being modeled can receive both crisp and fuzzy information. The methods of fuzzy logical inference known for fuzzy input values have either low calculation efficiency, or do not allow the use the whole variety of fuzzy logical operations. The paper describes a new method of logical inference based on fuzzy truth value for Multi-Input Single-Output (MISO) systems, which receive non-singleton input values. It gives a comparison of the method with the original method of Zadeh and popular method of Mamdani, and shows high computational efficiency of the method proposed. The method is generalized to systems with rule blocks.

Keywords Fuzzy rule-based system · Fuzzy truth value · Fuzzy inference method

1 Introduction

Rule-based models, also known as a models, based on the If-Then rules, are widely used since the late 1960s as a basis for constructing expert systems. In mid 1970s they were extended from classic to fuzzy logic, that allowed to solve a number of problems that cannot be solved by conventional mathematical methods, and resulted in the creation of artificial intelligence new areas, such as fuzzy expert systems and fuzzy control.

For the past decades various methods of fuzzy logic conclusion have been proposed, allowing to get output of the system based on the fuzzy model knowledge. According to the classification given in [1], all the methods can be divided into three

V.G. Sinuk (✉) · V.M. Polyakov · D.A. Kutsenko
BSTU named after V.G. Shukhov, Belgorod, Russia
e-mail: vgsinuk@mail.ru; dimonster@list.ru

V.M. Polyakov
e-mail: p_v_m@mail.ru

D.A. Kutsenko
e-mail: dakutsenko@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 450, DOI 10.1007/978-3-319-33609-1_36

types: logical, Mamdani, and Takagi-Sugeno. The second type is linked with the work of Mamdani [2], which proposed an effective method of output for the case, where the inputs receive crisp scalar values. This method uses the Cartesian product instead of classic implication for modeling logical conclusion for each of the rules and taking the maximum instead of a minimum to combine the results obtained using different rules, that, in general, a significant deviation from the traditions, adopted in classical logical systems, but allows to simplify the output value calculation process. In addition, while using distinct scalar input values, the disadvantages of the method are not revealed, so it became the most popular and has had a significant impact on the further development of fuzzy logic (see [3–6]).

In recent years it was repeatedly pointed out (see, e.g., [7]), that the Mamdani method may be irrelevant to solve the problem, when the input values represent by non-singleton fuzzy sets. Such problems arise in case where the original data is not crisp either by its nature (e.g., linguistic values), or have other NON-factors [8], such as inaccuracy, uncertainty, and underdeterminess that can be converted to fizziness.

The article describes output methods with a polynomial computational complexity for logic type fuzzy rule-based systems with n non-singleton inputs.

2 Problem Statement

The problem that is solved by using fuzzy rule-based system is formulated as follows. Let us consider a system with n inputs $x_1, \dots, x_n$ and one output y, represented by linguistic variables with same names. Let X_i be the base set of values of i-th input of the system, $x_i \in X_i$, $i = \overline{1, n}$; Y is base set of output values, $y \in Y$. Let us introduce the following notations. Let Fuzzy(Z) be the set consisting of all fuzzy subsets of set Z, $\mu_{\tilde{F}} : Z \to [0, 1]$ be membership function of a fuzzy set $\tilde{F} \in \text{Fuzzy}(Z)$. The interconnection of inputs and outputs in the system is described by N fuzzy rules of the following form:

$$\tilde{R}_k : \text{If } \left\langle x_1 \text{ is } \tilde{A}_{1k} \right\rangle \text{ and } \dots \text{ and } \left\langle x_n \text{ is } \tilde{A}_{nk} \right\rangle \text{, then } \left\langle y \text{ is } \tilde{B}_k \right\rangle, \quad k = \overline{1, N}, \tag{1}$$

where $\tilde{A}_{1k} \in \text{Fuzzy}(X_1), \dots, \tilde{A}_{nk} \in \text{Fuzzy}(X_n)$ are terms from the term set of input linguistic variables $x_1, \dots, x_n$ respectively, of which an antecedent of k-th rule is formed; $\tilde{B}_k \in \text{Fuzzy}(Y)$ is the term of the linguistic output variable term-set used to describe the consequent of k-th rule.

Let assume non-singleton fuzzy sets $\tilde{A}'_i \in \text{Fuzzy}(X_i)$, $i = \overline{1, n}$ that represent real values of inputs $x_1, \dots, x_n$. The task is to determine the output of the system $\tilde{B}' \in \text{Fuzzy}(Y)$. According to the classification, given in [1] logical type systems are special due to the fact that the rule (1) is formalized with the use of fuzzy implication as a fuzzy relation $\tilde{R}_k \in \text{Fuzzy}(X_1 \times \dots \times X_n \times Y)$ in the following way:

$$\tilde{R}_k = \tilde{A}_{1k} \times \dots \times \tilde{A}_{nk} \times Y \Rightarrow X_1 \times \dots \times X_n \times \tilde{B}_k, \quad k = \overline{1, N},$$

where "⇒" is fuzzy implication, reflecting cause-and-effect connection between antecedent "$\langle x_1 \text{ is } \tilde{A}_{1k} \rangle$ and … and $\langle x_n \text{ is } \tilde{A}_{nk} \rangle$" and consequent $\langle y \text{ is } \tilde{B}_k \rangle$ of k-th rule.

3 Zadeh Inference Method

In theory of fuzzy production systems construction [9] is considered to be a fundamental work, in which L.A. Zadeh suggested the so-called compositional rule of inference (CRI), that allows to describe the relationship between the input and output of the system, specified in the rule as the composition of fuzzy relations.

Let us consider a system with one input $x \in X$, one output $y \in Y$ and a rule "If $\langle x \text{ is } \tilde{A} \rangle$, then $\langle y \text{ is } \tilde{B} \rangle$" based on the fuzzy extension of the classical deductive modus ponens (generalized modus ponens), that is a special case of the CRI. It is represented as follows:

$$\begin{array}{rl} \text{If } \langle x \text{ is } \tilde{A} \rangle, \text{then } \langle y \text{ is } \tilde{B} \rangle. & \text{(Premise 1, rule)} \\ \langle x \text{ is } \tilde{A}' \rangle. & \text{(Premise 2, input value)} \\ \hline \langle y \text{ is } \tilde{B}' \rangle. & \text{(Conclusion, output value)} \end{array}$$

In Zadeh method the rule "If $\langle x \text{ is } \tilde{A} \rangle$, then $\langle y \text{ is } \tilde{B} \rangle$" is expressed with the help of fizzy binary relation $\tilde{R} \in \text{Fuzzy}(X \times Y)$ with membership function $\mu_{\tilde{R}}(x, y) = \mathrm{I}\left(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(y)\right)$, where $\mathrm{I} : [0, 1] \times [0, 1] \to [0, 1]$ is operation of fuzzy implication. The output value $\tilde{B}'$ according to the method is defined by the equation

$$\tilde{B}' = \tilde{A}' \circ \tilde{R}, \tag{2}$$

where "∘" is operation of composition of fuzzy relationships. In terms of membership functions using sup-t-composition (2) will be rewritten as follows:

$$\mu_{\tilde{B}'}(y) = \sup_{x \in X} \left\{ \mu_{\tilde{A}'}(x) \overset{\mathrm{T}}{*} \mathrm{I}\left(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(y)\right) \right\}, \tag{3}$$

where "$\overset{\mathrm{T}}{*}$" may be any t-norm.

The main disadvantage of Zadeh's method, that makes practical implementation difficult when using fuzzy input values $\tilde{A}'_1, \ldots, \tilde{A}'_n$, generally is exponentially growing computation complexity of the output with increasing input number. For instance, for system with n independent inputs $x_1, \ldots, x_n$, on which fuzzy values are set $\tilde{A}'_1, \ldots, \tilde{A}'_n$ respectively, the conclusion of output fuzzy value $\tilde{B}'$ will look as follows:

$$\frac{\begin{array}{l}\text{If } \langle x_1 \text{ is } \tilde{A}_1 \rangle \text{ and } \dots \text{and } \langle x_n \text{ is } \tilde{A}_n \rangle, \text{then } \langle y \text{ is } \tilde{B} \rangle. \\ \quad \langle x_1 \text{ is } \tilde{A}'_1 \rangle \text{ and } \dots \text{and } \langle x_n \text{ is } \tilde{A}'_n \rangle.\end{array}}{\langle y \text{ is } \tilde{B}' \rangle,}$$

with

$$\tilde{B}' = \tilde{A}'_1 \times \cdots \times \tilde{A}'_n \circ (\tilde{A}_1 \times \cdots \times \tilde{A}_n \times Y \Rightarrow X_1 \times \cdots \times X_n \times \tilde{B}),$$

and if membership functions are used:

$$\mu_{\tilde{B}'}(y) = \sup_{(x_1,\dots,x_n)\in\prod_{i=1}^{n} X_i} \left\{ \underset{i=\overline{1,n}}{\mathrm{T}} \left\{ \mu_{\tilde{A}'_i}(x_i) \right\} \overset{\mathrm{T}}{*} \mathrm{I} \left(\underset{i=\overline{1,n}}{\mathrm{T}} \left\{ \mu_{\tilde{A}_i}(x_i) \right\},\ \mu_{\tilde{B}}(y) \right) \right\}, \tag{4}$$

where supermum has to be taken at all n-s $(x_1, \dots, x_n) \in X_1 \times \cdots \times X_n$, which defines the order of computation complexity $O(|X|^n)$.

4 The Inference Method Based on the Fuzzy Truth Values

Let us consider the relationship Using the rule of true modification [10, 11], it is possible to write down

$$\mu_{\tilde{A}'}(x) = \mu_{\mathrm{CP}(\tilde{A},\tilde{A}')}\big(\mu_{\tilde{A}}(x)\big),$$

where $\mu_{\mathrm{CP}(\tilde{A},\tilde{A}')}(\tau)$ is membership function of fuzzy truth value $\mathrm{CP}(\tilde{A}, \tilde{A}')$ of fuzzy set $\tilde{A}$ relatively to $\tilde{A}'$, representing compatibility of term $\tilde{A}$ relative to input value $\tilde{A}'$ [12] according to the definition:

$$\mu_{\mathrm{CP}(\tilde{A},\tilde{A}')}(\tau) = \sup_{\substack{x\in X \\ \mu_{\tilde{A}}(x)=\tau}} \mu_{\tilde{A}'}(x).$$

Computation complexity of this expression is $O(|X|^2)$. It should be pointed out that at piecewise linear functions $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{A}'}(x)$ function $\mu_{\mathrm{CP}(\tilde{A},\tilde{A}')}(\tau)$ is also piecewise linear, that is proved with the help of the method, offered in [13].

Let us pass from a variable x to variable τ, denoting $\tau = \mu_{\tilde{A}}(x)$. We will get

$$\mu_{\tilde{A}'}(x) = \mu_{\mathrm{CP}(\tilde{A},\tilde{A}')}\big(\mu_{\tilde{A}}(x)\big) = \mu_{\mathrm{CP}(\tilde{A},\tilde{A}')}(\tau). \tag{5}$$

Then the membership function of binary fuzzy relationship $\tilde{R} \in \mathrm{Fuzzy}(X \times Y)$ may be represented as

$$\mu_{\tilde{R}}(x, y) = \mathrm{I}\big(\mu_{\tilde{A}}(x),\ \mu_{\tilde{B}}(y)\big) = \mathrm{I}\big(\tau,\ \mu_{\tilde{B}}(y)\big),$$

where $\mathrm{I} : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is operation of fuzzy implication. Thus, (3) with the account of the above correlations is expressed with the following formula:

$$\mu_{\tilde{B}'}(y) = \sup_{\tau \in [0, 1]} \left\{ \mu_{\mathrm{CP}(\tilde{A}, \tilde{A}')}(\tau) \overset{\mathrm{T}}{*} \mathrm{I}\big(\tau, \mu_{\tilde{B}}(y)\big) \right\}. \tag{6}$$

For generalization (6) to systems with n independent inputs it is necessary to prove the following proposition.

Proposition 1 *If $\tilde{C}$ and $\tilde{D}$ are fuzzy statements, accordingly represented as*

$$\tilde{C} = \big\langle (x_1, \ldots, x_n) \text{ is } (\tilde{A}'_1 \text{ and } \ldots \text{ and } \tilde{A}'_n) \big\rangle,$$
$$\tilde{D} = \big\langle (x_1, \ldots, x_n) \text{ is } (\tilde{A}_1 \text{ and } \ldots \text{ and } \tilde{A}_n) \big\rangle,$$

where $\tilde{A}'_i, \tilde{A}_i \in \mathrm{Fuzzy}(X_i)$, $i = \overline{1, n}$, then the truth value of a fuzzy statement $\tilde{D}$ relatively to $\tilde{C}$ is represented in the following way:

$$\mathrm{CP}(\tilde{D}, \tilde{C}) = \underset{i=\overline{1,n}}{\tilde{\mathrm{T}}} \mathrm{CP}(\tilde{A}_i, \tilde{A}'_i), \tag{7}$$

and this expression has a polynomial computational complexity. Here $\mathrm{CP}(\tilde{A}_i, \tilde{A}'_i)$ is the value of truth $\tilde{A}_i$ relatively to $\tilde{A}'_i$, and $\tilde{\mathrm{T}}$ is extended by the generalization principle n-fold t-norm [14]:

$$\mu_{\underset{i=\overline{1,n}}{\tilde{\mathrm{T}}} \mathrm{CP}(\tilde{A}_i, \tilde{A}'_i)}(\tau) = \sup_{\substack{(\tau_1, \ldots, \tau_n) \in [0, 1]^n \\ \underset{i=\overline{1,n}}{\mathrm{T}} \tau_i = \tau}} \left\{ \underset{i=\overline{1,n}}{\mathrm{T}} \mu_{\mathrm{CP}(\tilde{A}_i, \tilde{A}'_i)}(\tau_i) \right\}.$$

Proof According to the definition of the truth value

$$\mu_{\mathrm{CP}(\tilde{D}, \tilde{C})}(\tau) = \sup_{\substack{(x_1, \ldots, x_n) \in \prod_{i=1}^{n} X_i \\ \mu_{\tilde{D}}(x_1, \ldots, x_n) = \tau}} \mu_{\tilde{C}}(x_1, \ldots, x_n),$$

which is also characterized by computational complexity $O(|X|^n)$. Let us denote

$$\mu_{\tilde{D}}(x_1, \ldots, x_n) = \underset{i=\overline{1,n}}{\mathrm{T}} \mu_{\tilde{A}_i}(x_i);$$
$$\mu_{\tilde{C}}(x_1, \ldots, x_n) = \underset{i=\overline{1,n}}{\mathrm{T}} \mu_{\tilde{A}'_i}(x_i).$$

Then

$$\mu_{\text{CP}(\tilde{D},\tilde{C})}(\tau) = \sup_{\substack{(x_1,\ldots,x_n)\in\prod_{i=1}^{n} X_i \\ \underset{i=\overline{1,n}}{\text{T}} \mu_{\tilde{A}}(x_i) = \tau}} \left\{ \underset{i=\overline{1,n}}{\text{T}} \mu_{\tilde{A}'_i}(x_i) \right\}.$$

This equality does not change if it be written as follows:

$$\mu_{\text{CP}(\tilde{D},\tilde{C})}(\tau) = \sup_{\substack{(\tau_1,\ldots,\tau_n)\in[0,1]^n \\ \underset{i=\overline{1,n}}{\text{T}} \tau_i = \tau}} \left\{ \underset{i=\overline{1,n}}{\text{T}} \mu_{\text{CP}(\tilde{A}_i,\tilde{A}'_i)}(\tau_i) \right\}, \tag{8}$$

as following (5),

$$\mu_{\tilde{A}'_i}(x_i) = \mu_{\text{CP}(\tilde{A}_i,\tilde{A}'_i)}(\tau_i),$$
$$\mu_{\tilde{A}}(x_i) = \tau_i.$$

Equation (8) corresponds to the extended by generalization principle n-fold t-norm (7), that may be rewritten in accordance with t-norm characteristics in the following way:

$$\text{CP}(\tilde{D},\tilde{C}) = \tilde{\text{T}}\left(\underset{i=\overline{1,n-1}}{\tilde{\text{T}}} \text{CP}(\tilde{A}_i,\tilde{A}'_i), \text{CP}(\tilde{A}_n,\tilde{A}'_n) \right). \tag{9}$$

Two-fold extended t-norm

$$\mu_{\tilde{\text{T}}(\text{CP}_1,\text{CP}_2)}(\tau) = \sup_{\substack{(\tau_1,\tau_2)\in[0,1]^2 \\ \text{T}(\tau_1,\tau_2)=\tau}} \text{T}\left(\mu_{\text{CP}_1}(\tau_1),\ \mu_{\text{CP}_2}(\tau_2)\right)$$

has computation complexity $O(|X|^2)$, where $|X|^2$ corresponds to the number of pairs $(\tau_1,\ \tau_2)$, taken for supermum calculation. Therefore expression (9), and consequently, (7), are characterized by polynominal computation complexity in $O(n\,|X|^2)$ range. The statement is proved. □

With the account of the above proof the conclusion of output fuzzy value $\tilde{B}'$ based on fuzzy truth value will look as

$$\mu_{\tilde{B}'}(y) = \sup_{\tau\in[0,1]} \left\{ \mu_{\underset{i=\overline{1,n}}{\tilde{\text{T}}} \text{CP}(\tilde{A}_i,\tilde{A}'_i)}(\tau) \overset{\text{T}}{*} \text{I}\{\tau, \mu_{\tilde{B}}(y)\} \right\}. \tag{10}$$

5 Calculation of Fuzzy Output Value for the Rule Block

To determine the output of N rules (1), it is necessary to state what method will be used to solve the problem. Let us consider a method of Mamdani type [2]. In this case the aggregation of output values $\tilde{B}'_1, \ldots, \tilde{B}'_N$ by each rule is made with the help of fuzzy sets union operation:

$$\tilde{B}' = \bigcup_{k=1}^{N} \tilde{B}'_k.$$

The membership function $\tilde{B}'$ is calculated by using t-conorms, that is

$$\mu_{\tilde{B}'}(y) = \underset{k=\overline{1,N}}{\mathrm{S}} \mu_{\tilde{B}'_k}(y). \tag{11}$$

If a logical model is being used, then the aggregation of outputs by every rule is made as follows:

$$\tilde{B}' = \bigcap_{k=1}^{N} \tilde{B}'_k.$$

Accordingly, the membership function $\tilde{B}'$ is calculated by using t-norm, i.e.

$$\mu_{\tilde{B}'}(y) = \underset{k=\overline{1,N}}{\mathrm{T}} \mu_{\tilde{B}'_k}(y). \tag{12}$$

Then the reflection of fuzzy set $\tilde{B}'$ into an accurate scalar value $\bar{y}$ is accomplished. When using a gravity centre method in its discrete variation we will get the following:

$$\bar{y} = \frac{\sum_{l=1}^{N} \bar{y}_l \cdot \mu_{\tilde{B}_k}(\bar{y}_l)}{\sum_{l=1}^{N} \mu_{\tilde{B}_k}(\bar{y}_l)}, \tag{13}$$

where $\bar{y}_l,\ l = \overline{1,N}$ are values, where membership function centres are located $\mu_{\tilde{B}_k}$ [15].

When considering Mamdani-type methods $\mathrm{I}(\cdot)$ functions as a t-norm, that is

$$\mathrm{I}\left(\tau,\ \mu_{\tilde{B}_k}(y)\right) = \mathrm{T}\left(\tau,\ \mu_{\tilde{B}_k}(y)\right). \tag{14}$$

Fuzzy set $\tilde{B}'$, that determines cumulative output of the block with the account of (10), (11) and (14), and is set as follows:

$$\mu_{\tilde{B}'}(\bar{y}_l) = \underset{k=\overline{1,N}}{\mathrm{S}} \left\{ \sup_{\tau\in[0,1]} \left\{ \mu_{\mathrm{CP}_k}(\tau) \overset{\mathrm{T}}{*} \mathrm{T}\left(\tau,\ \mu_{\tilde{B}_k}(\bar{y}_l)\right) \right\} \right\},$$

where

$$\mu_{\mathrm{CP}_k}(\tau) = \mu_{\underset{i=\overline{1,n}}{\tilde{\mathrm{T}}} \mathrm{CP}(\tilde{A}_{ik}, \tilde{A}'_{ik})}(\tau).$$

It follows that relationship (13) looks like:

$$\bar{y} = \frac{\sum_{l=1}^{N} \bar{y}_l \cdot \underset{k=\overline{1,N}}{\mathrm{S}} \left\{ \sup_{\tau \in [0,1]} \left\{ \mu_{\mathrm{CP}_k}(\tau) \overset{\mathrm{T}}{*} \mathrm{T}\left(\tau,\ \mu_{\tilde{B}_k}(\bar{y}_l)\right) \right\} \right\}}{\sum_{l=1}^{N} \underset{k=\overline{1,N}}{\mathrm{S}} \left\{ \sup_{\tau \in [0,1]} \left\{ \mu_{\mathrm{CP}_k}(\tau) \overset{\mathrm{T}}{*} \mathrm{T}\left(\tau,\ \mu_{\tilde{B}_k}(\bar{y}_l)\right) \right\} \right\}}. \tag{15}$$

In the original Mamdani method [2] a particular case of t-norm $\mathrm{T} = \min$ and t-conorm $\mathrm{S} = \max$ is used. However, when using other t-norms, as, for example,

$$\mathrm{T}\left(\tau,\ \mu_{\tilde{B}_k}(\bar{y}_l)\right) = \max\{\tau + \mu_{\tilde{B}_k}(\bar{y}_l) - 1,\ 0\}, \tag{16}$$

Mamdani algorithm at fuzzy inputs cannot be realized with polynominal computation complexity. The need to use t-norms, different from min, arises at adjusting fuzzy systems (see, for example, [16]). Figure 1 illustrates a supermum definition in (15), if t-norm is set in accordance with (16).

Figure 2 shows relationship (15) in the form of network structure of the system. It is primarily denoted:

$$F_{kl}\left(\mu_{\mathrm{CP}_k}(\tau),\ \bar{y}_l\right) = \sup_{\tau \in [0,1]} \left\{ \mu_{\mathrm{CP}_k}(\tau) \overset{\mathrm{T}}{*} \mathrm{T}\left(\tau,\ \mu_{\tilde{B}_k}(\bar{y}_l)\right) \right\}.$$

When considering logical type system $\mathrm{I}(\cdot)$ functions as a fuzzy implication. Later we make use of S-implication [17]. In this case

$$\mathrm{I}\left(\tau,\ \mu_{\tilde{B}_k}(y)\right) = \mathrm{S}\left(1 - \tau,\ \mu_{\tilde{B}_k}(y)\right). \tag{17}$$

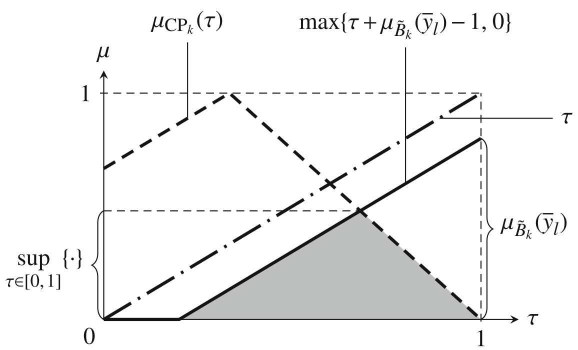


Fig. 1 Graphic interpretation to determine a supermum for correlation (15) at t-norm, set in accordance with (16)

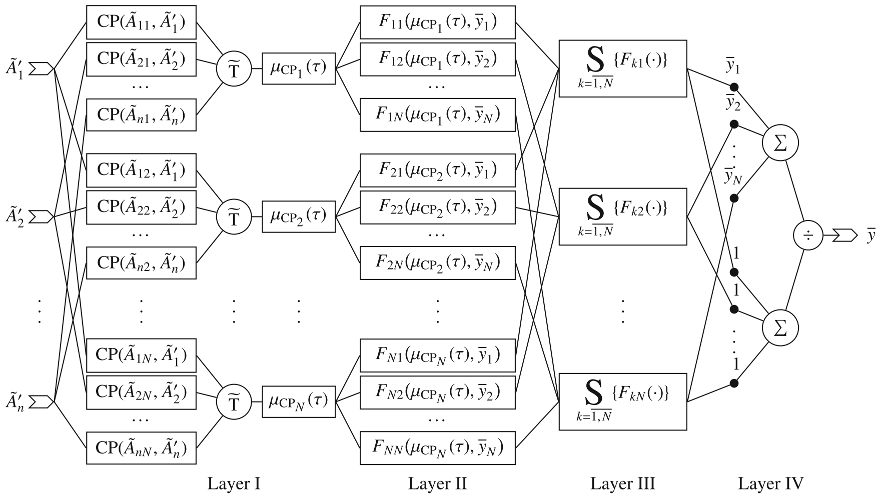


Fig. 2 Network structure, corresponding to correlation (15)

Following (10), (12), and (17), fuzzy set $\tilde{B}'$ will be determined as

$$\mu_{\tilde{B}'}(y) = \underset{k=\overline{1,N}}{\mathrm{T}} \left\{ \sup_{\tau \in [0,1]} \left\{ \mu_{\mathrm{CP}_k}(\tau) \overset{\mathrm{T}}{*} \mathrm{S}\big(1-\tau,\ \mu_{\tilde{B}_k}(y)\big) \right\} \right\}.$$

Hence the expression (13) looks like:

$$\bar{y} = \frac{\sum_{l=1}^{N} \bar{y}_l \cdot \underset{k=\overline{1,N}}{\mathrm{T}} \left\{ \sup_{\tau \in [0,1]} \left\{ \mu_{\mathrm{CP}_k}(\tau) \overset{\mathrm{T}}{*} \mathrm{S}\big(1-\tau,\ \mu_{\tilde{B}_k}(\bar{y}_l)\big) \right\} \right\}}{\sum_{l=1}^{N} \underset{k=\overline{1,N}}{\mathrm{T}} \left\{ \sup_{\tau \in [0,1]} \left\{ \mu_{\mathrm{CP}_k}(\tau) \overset{\mathrm{T}}{*} \mathrm{S}\big(1-\tau,\ \mu_{\tilde{B}_k}(\bar{y}_l)\big) \right\} \right\}}. \tag{18}$$

Let us illustrate graphic definition of supermum in Fig. 3 (18), if Łukasevich implication is applied, that is

$$\mathrm{S}\big(1-\tau,\ \mu_{\tilde{B}_k}(\bar{y}_l)\big) = \min\big\{1,\ 1-\tau+\mu_{\tilde{B}_k}(\bar{y}_l)\big\}.$$

The network structure, conforming to expression (18), will compile with the structure in Fig. 2, but the content at leyers II and III will be changed as follows:

$$F_{kl}\big(\mu_{\mathrm{CP}_k}(\tau), \bar{y}_l\big) = \sup_{\tau \in [0,1]} \left\{ \mu_{\mathrm{CP}_k}(\tau) \overset{\mathrm{T}}{*} \mathrm{S}\big(1-\tau,\ \mu_{\tilde{B}_k}(\bar{y}_l)\big) \right\},$$

$$\underset{k=\overline{1,N}}{\mathrm{S}}\{F_{kl}(\cdot)\} \text{ replaced by } \underset{k=\overline{1,N}}{\mathrm{T}}\{F_{kl}(\cdot)\}.$$

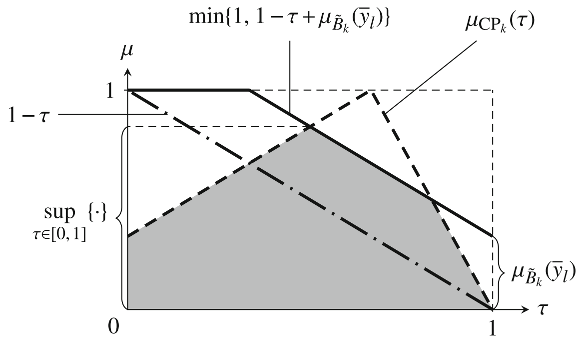


Fig. 3 Graphic interpretation of the supermum definition for expression (18)

6 Conclusion

The paper offers a method for fuzzy conclusion for logical type systems on the inputs of which fuzzy values are set. It provides a graphical illustration for the use of some implication functions.

The method presented has polynominal computational complexity, which allows to use it for solving modeling problems for systems with big amount of fuzzy inputs, such as diagnostics, forecast and control (see, for example [18]). The further objective of the investigation is training algorithms developing, necessary to transform network structures, obtained on the basis of fuzzy input value expressions for the block of rules into neuro-fuzzy systems.

References

1. Rutkowski, L.: Computational Intelligence: Methods and Techniques. Springer (2008)
2. Mamdani, E.H.: Applications of fuzzy algorithm for control a simple dynamic plant. Proc. IEEE. **121**(12), 1585–1588 (1974)
3. Kruglov, V.V., Dli, M.I., Golunov, R.U.: Nechetkaja logika i iskusstvennye nejronnye seti. Fizmatlit (2001)
4. Leonenkov, A.V.: Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH. BHV-Peterburg (2005)
5. Jarushkina, N.G.: Osnovy teorii nechetkih i gibridnyh sistem. Finansy i statistika, (2004)
6. Fedunov, B.E.: Mehanizmy vyvoda v baze znanij bortovyh operativno sovetujushhih ekspertnyh sistem. Izv. RAN. TiSU. 4 (2002)
7. Dubois, D., Prade, H.: Abe Mamdani: A Pioneer of Soft Artificial Intelligence. Combining Experimentation and Theory, pp. 49–60. Springer (2012)
8. Dushkin, R.V.: Metody poluchenija, predstavlenija i obrabotki znanij s NE-faktorami. Vebov i Knigin (2011)
9. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. Syst. Man Cybern. SMC-**3**(1), 28–44 (1973)
10. Borisov, A.N. et al.: Modeli prinjatija reshenij na osnove lingvisticheskoj peremennoj. Zinatne (1982)
11. Sinjuk, V.G., Kucenko, D.A.: Kosvennyj metod nechetkogo vyvoda dlja produkcionnyh sistem so mnogimi vhodami. Programmnye produkty i sistemy. **1**, 45–47 (2008)

12. Zadeh, L.A.: PRUF–a meaning representation language for natural language. Int. J. Man-Mach. Stud. **10**, 395–460 (1978)
13. Kucenko, D.A., Sinjuk, V.G.: Algoritmy nahozhdenija SR pri kusochno-linejnom predstavlenii funkcij prinadlezhnosti. Nechetkie sistemy i mjagkie vychislenija (NSMV-2008). UlGTU. **1**, 87–92 (2008)
14. Dubois, D., Prade, H.: PossibilIty Theory (1986)
15. Pegat, A.: Fuzzy Modeling and Control. BINOM, Laboratory of Knowledge (2009)
16. Rutkowska, D., Rutkowski, L., Nowicki, R.: Neuro-Fuzzy System with Inference Based on Bounded Product. Advances in Networks and Applications, pp. 104–109 (2001)
17. Baczyński, M., Jayaram, B.: Fuzzy Implications. Springer, Heidelberg (2008)
18. Kudłacik, P.: Advantages of an approximate reasoning based on a fuzzy truth value. J. Med. Inform. Technol. **16**, 125–132 (2010)

Part VI
Evolutionary Modeling, Bionic Algorithms and Computational Intelligence

Feature Selection Using a Genetic Algorithm for Solar Power Prediction

Sebastián Basterrech and Václav Snášel

Abstract We study an automatic procedure for selecting the most useful external variables for solar power forecasting. We use *Genetic Algorithm (GA)* as combinatorial optimisation tool of these feature variables. As forecasting model we use a particular case of Neural Network named *Echo State Networks (ESN)*, which has been successfully used in the community for solving temporal learning problems. We study more than 20 weather variables that can impact on the solar power, and we compare the obtained results by GAs with the Spearman's rank correlation coefficient. Our approach is evaluated on a well-known public dataset, and we obtain promising results.

Keywords Solar irradiance · Genetic algorithms · Echo state networks · Time-series problems · Forecasting

1 Introduction

There is still a large part of global population without access to electricity, for instance more than 2 billions of persons lacked access to it during 2011 [1]. Solar energy is an alternative of renewable resource that can be used for covering a part of the growing future demand. Although, it has the following technical drawbacks: the solar energy is very sensitivity to local weather conditions, it can have a large intra-hour variability, and it is produced only during a range of hours by day. These limitations affect operational costs in the grid. Therefore, the prediction of the amount of solar power and its variability can mitigate its negative impact [2]. Good predictions

S. Basterrech (✉) · V. Snášel
Faculty of Electrical Engineering and Computer Science,
Department of Computer Science, VŠB-Technical University of Ostrava,
Ostrava, Czech Republic
e-mail: Sebastian.Basterrech.Tiscordio@vsb.cz; sebabaster@gmail.com

V. Snášel
e-mail: Vaclav.Snasel@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 450, DOI 10.1007/978-3-319-33609-1_37

can help in the identification of grid congestions, it can enhance planning and security, as well as it can improve the management of the energy distribution. Last years, solar power forecasting has received significant attention. Some approaches have been based on statistical methods [2], other ones on Neural Networks [2–4], as well as other learning techniques have been studied [5–8].

In this article, we study a procedure for selecting the useful external variables that affect the solar power forecasting. We use the well-known *Genetic Algorithm (GA)* as combinatorial optimisation tool for feature selecting the variables. For forecasting the time series, we use a particular type of Neural Network named *Echo State Networks (ESN)*. Our main goal is defining a subset of meteorological variables useful for solving forecasting this problem, and we analyse GAs as feature selection tool on a well-known public dataset [9], which is composed by many meteorological variables (more than 20 s features).

The article is organised as follows. In the next section we define the problem of forecasting a time-series and we present a background about GAs and ESNs. We present the procedure of using GAs as feature selection technique in Sect. 3. Then, Sect. 4 is divided in two parts. First part describes the dataset, and second part presents the experimental results. The article ends with an outlook and conclusions.

2 Background

The section starts by formalising the problem of forecasting a time-series. Next, Genetic Algorithms, and Echo State Networks.

2.1 Formalisation of the Forecasting Problem

Forecasting a time-series consists in predicting future events or trends on some arbitrary period ahead using information from the past. Given a real time-series $y(1), y(2), \ldots, y(t)$, the problem consists in computing a learning tool $\varphi(\cdot, \mathbf{w})$ with parameters $\mathbf{w}$ for predicting values $y(t+\tau)$ with $\tau > 0$, such that an arbitrary distance between predicted values $\varphi(\cdot, \mathbf{w})$ and real values $y(t)$ is minimised in a given range of time. In this article there is considered a quadratic distance in the range $[t+1, t+T]$ that is defined as follows:

$$E = \frac{1}{T}\sum_{k=1}^{T}(\hat{y}(t+k) - y(t+k))^2, \tag{1}$$

where $\hat{y}(t+k)$ is the output of the model at time $t+k$. In many situations, we have a set of exogenous variables $\mathbf{a}(t)$ in some multidimensional space $\mathbb{R}^{N_a}$ that can be used for forecasting an endogenous real variable $y(t)$. In this case, the learning tool $\varphi(\cdot, \mathbf{w})$ has as inputs the lagged explanatory variable $y(t-1), y(t-2), \ldots, y(t-\Delta_t)$ and the exogenous variables $\mathbf{a}(t), \mathbf{a}(t-1), \ldots, \mathbf{a}(t-\Delta_t)$ [10]. In the following, we denote by $\mathbf{z} \in \mathbb{R}^{N_z}$ a vector that collects the input information of the learning tool, then $\mathbf{z}(t)$ is composed by both the Δ_t-lagged endogenous variable $y(t)$ and the exogenous variable $\mathbf{a}(t)$.

2.2 Genetic Algorithm

Probably, the most recognised *Evolutionary Algorithm (EA)* is the *Genetic Algorithm (GA)*, which was introduced at the beginning of the 1960s [11]. The algorithm is motivated by the natural selection processes [12]. A GA is an iterative process in which, at each iteration many feasible solutions for a given optimisation problem are analysed, and they are combined according to some evolutive rules. The aim of that this combination is producing a new collection of feature solutions with higher quality. The method is inspired by the selection on the breeding of the plants and animals, which consists in prioritising the life of live beings which have more adaptation abilities. Following this analogy, the collection of possible solutions is named *population*, which is composed by individuals called *genotype*. The solutions that encode the genotypes are named *phenotypes* [13]. We can interpret the phenotypes as points in a multidimensional space wherein each coordinate represents a *gene*. In each iteration, the phenotypes are evaluated by a fitness function that is in general a parametric distance. The phenotypes with a higher fitness value have a larger probability of surviving from one generation to the next one [13]. The transition from one generation to the next one is consequence of the following operations over the population:

- Selection: is the operation for selecting a subgroup of genotypes, which can reproduce themselves in order to breed a new generation where parents may survive from the current generation to the next one. There are several selection schemas presented in the literature, in our experimental results we present the selection following the *Baker's stochastic universal selection*.
- Crossover: is a function that takes two genotypes of the current generation and creates two new ones. The operation replaces some genes of one parent by the corresponding genes of another one. In general, the process for selecting those genes to be replaced, is a random mechanism. In our experimental set-up, we follow the criteria of *one-point crossover*, this means that given two parents A and B, this operation consists in random selecting a cutting-point and generating two new genotypes. One of them in its first part (until the cutting point) has genes from parent A and the second part (from the cutting point until the end) has genes from the another parent B, while the other one has in its first part the genes from B and in its second part has the genes from A.
- Mutation: the operation consists in selecting a subgroup of genes and changing their values. In our problem we consider binary genes, therefore a gene mutation consists in the binary complement operation of a random group of selected genes.

2.3 Echo State Networks

A Recurrent Neural Network (RNN) is a bio-inspired computational model that can be applied for solving supervised learning problems. The recurrences on the network allow to learn temporal dynamics about the input data. As a consequence of that, the model has been applied for modelling sequential data and solving temporal learning problems. An *Echo State Networks (ESN)* is a RNN with a particular topology and training schema [14]. The ESN model consists of two structures, one is a RNN used for memorising the input sequence, another one is a simple linear regression for learning input patterns and generates the model outputs. The network has three layers connected in a forward schema. The first layer processes the input patterns, the second one contains recursive connections, and the last one produces the outputs. The role of the second structure, named *reservoir*, is to memorise the serial order of the inputs and to map the input space into a larger dimensional space. The third layer generates a linear combination from the mapped data to the outputs. The ESN has circuits only in the second layer. In addition, the weights of the second layer are fixed a priori, only the weights from the second layer until the third one are adjusted during the learning process. As a consequence, the learning algorithm is fast and robust, because it consists in training the parameters of a linear regression. The model has proven to obtain good performances in many real applications [15, 16].

The model is defined as follow: let N_{a}, N_{x} and $\mathrm{N_o}$ be the number of input units, reservoir and output units, respectively. The weights are collected in the following matrices: let $\mathbf{w}^{\mathrm{in}}$ be a $N_{\mathrm{x}} \times N_{\mathrm{a}}$ matrix collecting input-reservoir weights, let $\mathbf{w}^{\mathrm{r}}$ be a $N_{\mathrm{x}} \times N_{\mathrm{x}}$ matrix collecting hidden-hidden weights, and let $\mathbf{w}^{\mathrm{out}}$ a $N_{\mathrm{o}} \times (N_{\mathrm{a}} + N_{\mathrm{x}})$ matrix with the parameters from the input and the projected space to the output space. The hidden-hidden weights generates the dynamics of the reservoir, which are represented in a high dimensional state by:

$$\mathbf{x}(t) = f_1(\mathbf{w}^{\mathrm{in}}\mathbf{a}(t) + \mathbf{w}^{\mathrm{r}}\mathbf{x}(t-1)), \tag{2}$$

and

$$\hat{\mathbf{y}}(t) = f_2(\mathbf{w}^{\mathrm{out}}\mathbf{x}(t)), \tag{3}$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are two predefined coordinate-wise functions. For the sake of notation simplicity, we omit the bias term which is included in the weight matrices. The functions $f_1(\cdot)$ and $f_2(\cdot)$ in general are non-linear ones. Instead of using the expression (2), we are using a generalisation of the canonical ESN model that computes the reservoir state using a leaky parameter as follows:

$$\mathbf{x}(t) = (1-\alpha)\mathbf{x}'(t) + \alpha\mathbf{x}(t-1), \tag{4}$$

where $\mathbf{x}'(t)$ is computed using the expression (2), and α is called leaky rate.

The ESN model has the following global parameters that impact in the model performance as follows [16, 17]:

- Size of the reservoir is given by the number of reservoir neurons, and it impacts on the linear separability of the data.
- Input scaling factor weights the input patterns.
- Spectral radius of the reservoir matrix controls the dynamics of the reservoir. It is set for controlling the memory capability of the model [15]. The stability is controlled when the spectral radius of $\mathbf{w}^{\mathrm{r}}$ satisfies $\rho(\mathbf{w}^{\mathrm{r}}) < 1$, although the stability can happen even when $\rho(\mathbf{w}^{\mathrm{r}}) >= 1$ [16]. For more details about the reservoir stability see the *Echo State Property* [14]. In practice, the reservoir is random initialised and scaled as follows: $\mathbf{w}^{\mathrm{r}} \leftarrow (\beta/\rho(\mathbf{w}^{\mathrm{r}}))\mathbf{w}^{\mathrm{r}}$, where β is a constant in (0, 1].
- Density of the weight matrix of the reservoir, it is suggested to use around a 20 % of non-zero values on the reservoir matrix.

In our experiments, the training data is normalised, we consider the input scaling factor equal to 1 (therefore all the input patterns have equal relevance). The sparsity of the reservoir matrix is often set on 20 % non-zero values.

3 Feature Selection for Solar Power Prediction

We assume that the solar power is affected by several meteorological external variables, such as: air temperature, humidity, wind characteristics, etc. We apply GAs for automatically selecting a set of meteorological variables that we use for predicting the solar global irradiance. We are considering many variables (more than 20 features), therefore the selection can not be done using a brute-force strategy or a greedy method due to the large dimension of the earthing space. The proposed procedure is as follows. We define without loss of generality a set of input features $\{1, \ldots, N\}$, where N is the number of meteorological variables. We have a combinatorial optimisation problem with a searching space $\{0, 1\}^N$, where the points solutions have the form $\mathbf{s} = [s_1, s_2, \ldots, s_N]$ where $s_i = 0$ represents that the input variable i is omitted for generating the model, and $s_i = 1$ represents that the variable i is an input of the model. The objective is to find the *best* combination $\mathbf{s} \in \{0, 1\}^N$ for using as inputs of the ESN model.

We begin by defining a *good* forecasting model for our purposes. Therefore, we start for finding the "best" global parameters of the ESN. To fit the parameters of an ESN can be a hard task. We evaluate the model over three predicted days for parameters defined in a regular spaced-grid points in the following intervals: $\alpha \in [0.5, \ldots, 0.9]$, $N_{\mathrm{x}} \in [30, 35, \ldots, 120, 125]$ and $\rho(\mathbf{w}^{\mathrm{r}}) \in [0.1, 0.15, \ldots, 0.95]$. According our empirical results, we decided to use an ESN with the following parameters: $\alpha = 0.3$, $\rho(\mathbf{w}^{\mathrm{r}}) = 0.8$, and $N_{\mathrm{x}} = 90$. The ESN receives two types of input information: the meteorological explanatory variable $\mathbf{a}(t)$ and predicted values of the solar irradiance. We assume that after an arbitrary period Δ_t, we are able to have new measured values for the explanatory variables (**a**). Therefore at each instant t the ESN model has as inputs $\mathbf{a}(t - \Delta_t)$ and $\hat{y}(t)$, and the model output is the prediction $\hat{y}(t+1)$. In practice, most often we don't have the values of $\mathbf{a}(t)$ for predicting $\mathbf{y}(t+1)$, therefore we use as inputs Δ_t-lagged external meteorological inputs $\mathbf{a}(t - \Delta_t)$.

4 Experimental Results

This section starts with a description of the dataset. Next, we analyse our experimental results.

4.1 Data Description

We use the meteorological data provided by the National Renewable Energy Laboratory and Solar Technology Acceleration Center (SolarTAC) [9]. The collected data corresponds to the period that starts in January 1, 2015 till December 5, 2015. The temporal precision of the data is 1 min. The output variable is the global irradiance given by the *Global Horizontal Irradiance* in W/m^2, the input features are: Air Temperature, Wind Chill Temp, Dew Point Temp, Relative Humidity, Wind Speed, Pk Wind Speed, SDev Wind Speed, Wind Direction, Wind Dir at Pk WS, SDev Wind Direction, Station Pressure, Precipitation, Accumulated Precipitation, Zenith Angle, Azimuth Angle, Airmass, CMP22 Temp, and CR1000 Temp, CR1000 Battery, CR1000 Process Time. For more information about those variables and the protocol used for collecting the dataset, please see [9]. The preprocessing of the dataset consisted in changing the temporal precision from 1 to 10 min. Instead of using the variable information each minute, we consider the data each 10 min. The time-series covers has 50232 points in this period. All the variable were normalised in [0, 1]. Figure 1 presents the normalised solar irradiance during the year according to minutes and days. In Fig. 2 we can see the normalised air temperature. It is remarkable from both figures the correlation between the solar irradiance and the air temperature. Figure 3 shows the normalised humidity, following the same criteria that on the other figures, red colour represents high values of humidity and blue colour represents low values. We can see a similitude on the curves of solar irradiance and air temperature. On the other hand, the humidity shows an opposite trend that the solar power. These properties were also analysed on another dataset in [5]. Note that, we are not directly modelling the solar power, instead of that we are modelling the solar irradiance that has a strong correlation with solar power [2].

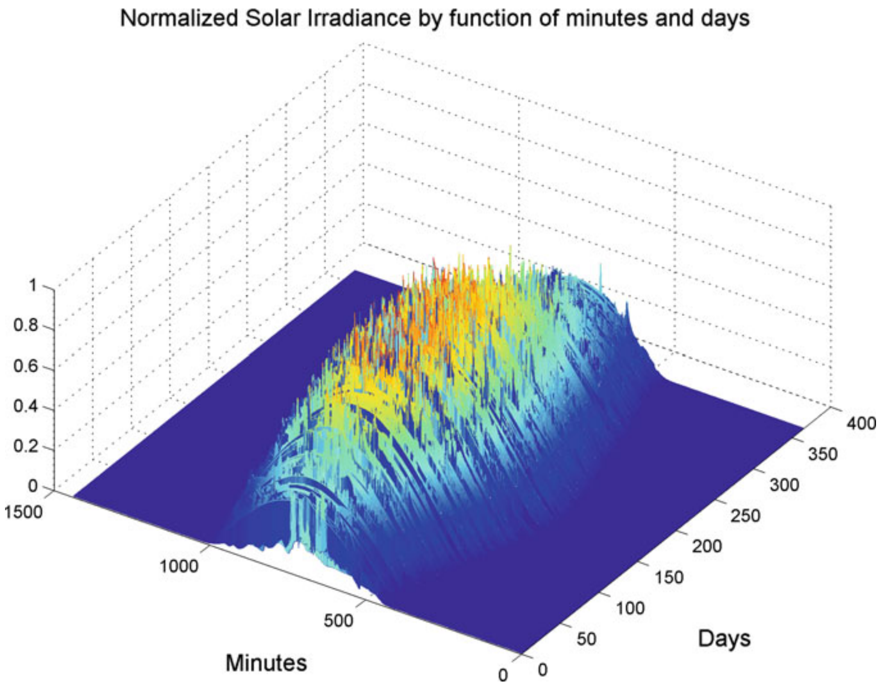


Fig. 1 Normalised solar irradiance according to minutes and days of the year 2015

4.2 Results Analysis

In order of having statistical results, we perform the GA 25 times with different initial populations. Table 1 presents the performance on the validation dataset for forecasting 3 days ahead, according to different number of generations of the GA. Table shows the minimum, mean and average of MSE reached among the 25 trials. Last column of the table presents the number of features used by the best chromosome at the iterations 10, 100, 200, 300 and 400. Figure 4 illustrates the MSE evolution of each trial of the GA. We can see that the error decreases with the number of generations. The best combination of input features reached with 400 iterations of GA has the following 7 features: solar irradiance, wind chill temperature, station pressure, precipitation, Zenith angle, air mass, and CMP22 temperature. In addition, the following variables don't appear as part of the solution in any of the configurations proposed by the 25 trials of GAs: wind direction, wind direction at Pk WS, accumulated precipitation. Besides, the following variables appear only in two final configurations of the 25 runs of the GAs: dew point temperature and Azimuth angle.

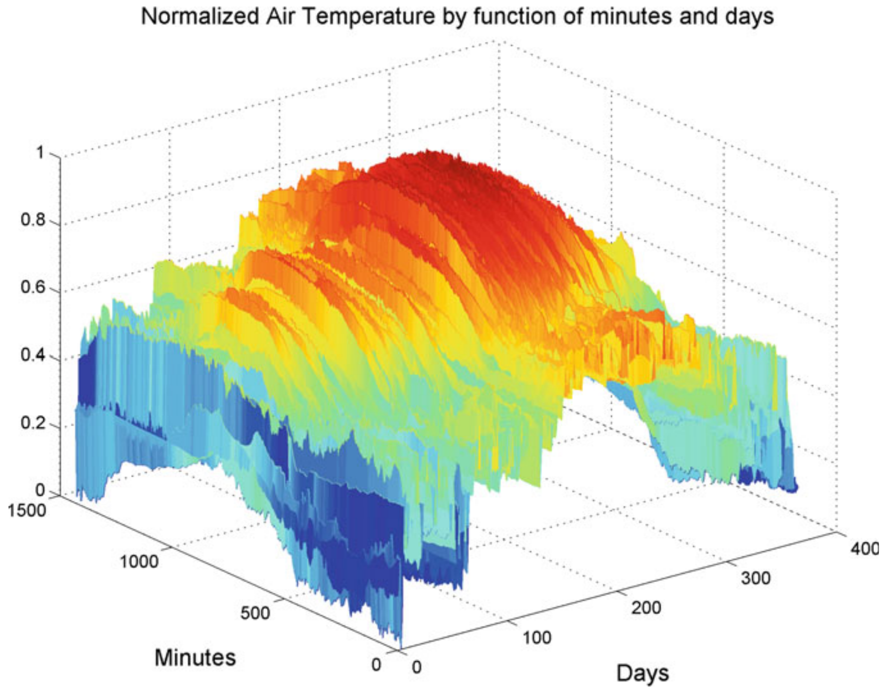


Fig. 2 Normalised air temperature according to minutes and days of the year 2015

In the following we compare the feature selection using GA with the Spearman's rank correlation coefficient. Table 2 shows in the first part the correlation factor of the selected variables of the best solution obtained by GA. The second part shows the Spearman correlation factor of the variables that aren't selected in any of the 25 runs of GA. The third column presents the ranking position of the variable when they are sorted according to the absolute value of correlation factor (the variable solar irradiance doesn't appear in the ranking because its correlation factor is 1). Note that the combination proposed by GA doesn't consider the air temperature and humidity, this is a difference with precedent related works.

5 Conclusions and Future Work

We presented a procedure for selecting the features using *Genetic Algorithms (GAs)* on a weather forecasting problem. We use GAs for selecting the *best* combination of external meteorological variables for forecasting the solar irradiance. We use a specific type of Recurrent Neural Network named *Echo State Networks (ESN)* for forecasting the time-series. The evaluation of our approach is done over a meteorological dataset provided by the Solar Technology Acceleration Center (SolarTAC),

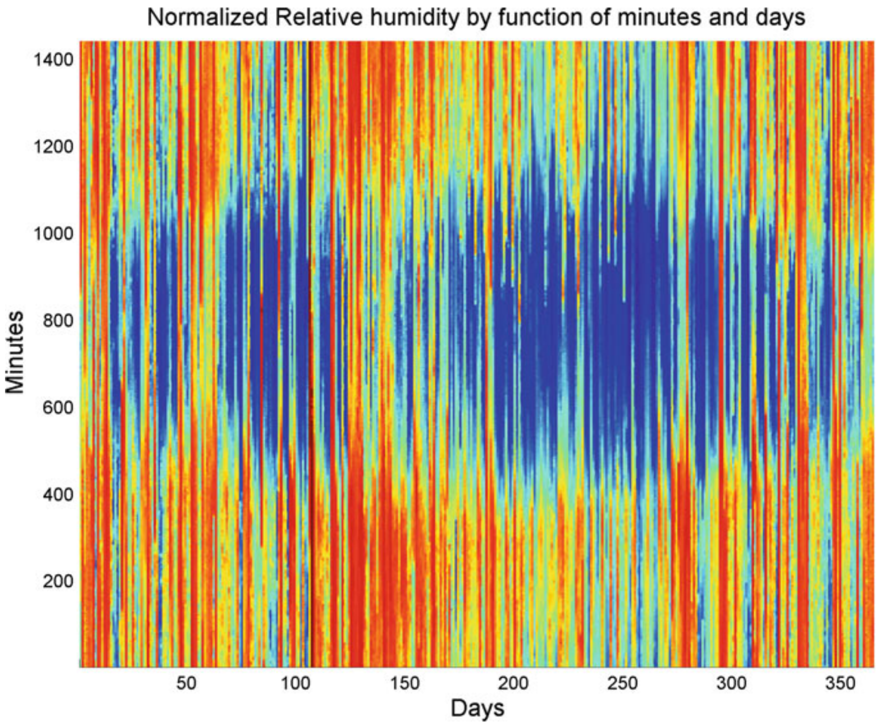


Fig. 3 Normalised humidity according to minutes and days of the year 2015

Table 1 Performance of 25 trials of GAs when we forecast three days ahead

Generations	Min (10^{-4})	Mean (10^{-4})	Var (10^{-9})	Number of features
10	4.7152	5.0415	5.1283	4
100	4.7152	4.9494	2.4654	4
200	4.6560	4.8775	1.3514	5
300	4.5372	4.8271	1.6695	7
400	4.5372	4.8216	1.5673	7

The accuracy, in scientific notation, is presented according the number of generations

Colorado, USA. We consider that we obtain promising results for a forecasting horizon of three days. It is remarkable that the best group of selected variables by the GA technique doesn't exactly matches with the group of variables with higher Spearman's rank correlation coefficient with the solar irradiance. In spite of that, the two most correlated variables with the solar irradiance are also selected with GA. We are interested in the near future to compare the results with other metahueristic techniques, as well as to extend the period used for training the network model.

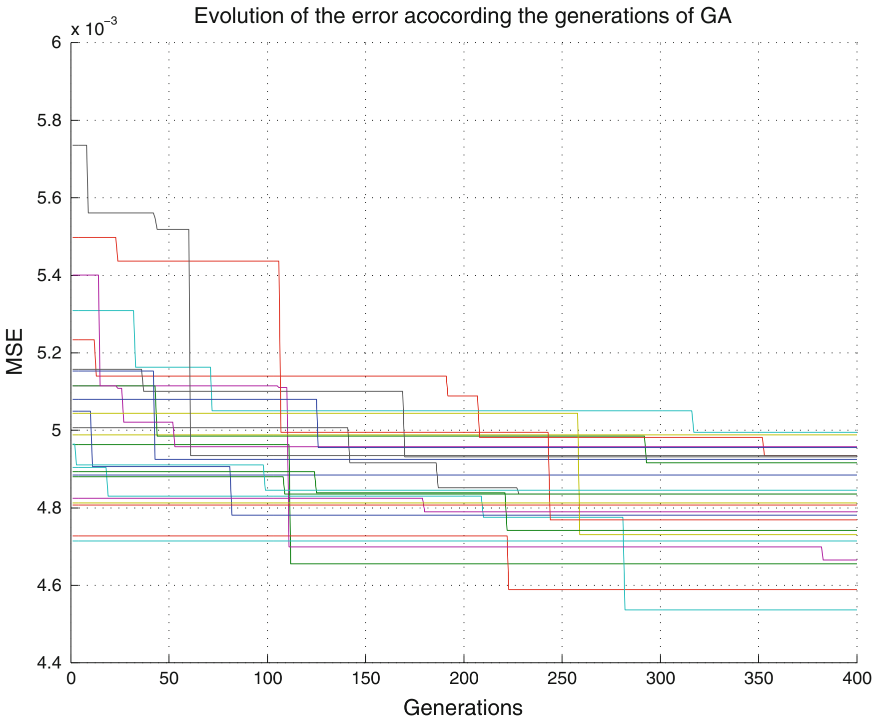


Fig. 4 Evolution of the model accuracy over the first 400 generations of GA. Each *curve* represents the evolution for different random initial trails of the GA

Table 2 Spearman's rank correlation factor computed for the solar irradiance in respect of other variables

Selected variables	Corr. factor	Rank
Wind chill temperature	0.4142	6
Station pressure	0.0542	14
Precipitation	−0.0468	15
Zenith angle	−0.7886	1
Air mass	−0.50754	2
CMP22 temperature	0.4588	3
Discarded variable	Corr. factor	Rank
Wind direction	−0.0132	21
Wind direction at Pk WS	−0.0390	17
Accumulated precipitation	−0.1320	12

The table shows the selected variables by GA

Acknowledgments This work was supported by Grant of SGS No. SP2016/97, VŠB-Technical University of Ostrava, Czech Republic.

References

1. Cox, B., Cohen, A.: Human Universe. William Collins (2014)
2. Pedro, H.T.C., Coimbra, C.F.M.: Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy **86**(7), 2017–2028 (2012)
3. Basterrech, S., Prokop, L., Burianek, T., Misak, S.: Optimal design of neural tree for solar power prediction. In: Proccedings of the 2014 15th International Scientific Conference on Electric Power Engineering (EPE), pp. 273–278, May 2014
4. Basterrech, S., Zjavka, L., Prokop, L., Misak, S.: Irradiance prediction using echo state queueing networks and differential polynomial neural networks. In: 2013 13th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 271–276, Dec 2013
5. Chen, C., Duan, S., Cai, T., Liu, B.: Online 24-h solar power forecasting based on weather type classification using artificial neural network. Solar Energy **85**, 2856–2870 (2011)
6. Bacher, P., Madsen, H., Nielsen, H.A.: Online short-term solar power forecasting. Solar Energy **83**, 1772–1783 (2009)
7. Letendre, S., Makhyoun, M., Taylor, M.: Predicting solar power production: irradiance forecasting models, applications and future prospects. Technical report, Solar Electric Power Association, Washington, DC, USA, March 2014. www.solarelectricpower.org
8. Pelland, S., Remund, J., Kleissl, J., Oozeki, T., De Brabandere, K.: Photovoltaic and solar forecasting: state of art. Technical Report IEA PVPS T14 01:2013, International Energy Agency Photovoltaic Power Systems Programme (2013). http://www.iea-pvps.org
9. Andreas, A., Wilcox, S.: Aurora, colorado (data). Technical Report DA-5500-56491, Solar Technology Acceleration Center (SolarTAC), Colorado, USA (2011). doi:10.5439/1052224
10. Robinzonov, N., Tutz, G., Hothorn, T.: Boosting techniques for nonlinear time series models. AStA Adv. Stat. Anal. **96**(1), 99–122 (2012)
11. Holland, J.H.: Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor (1975)
12. Reeves, C.R.: Genetic alogarithms for the opertions researcher. INFORMS J. Comput. **9**(3), 231–250 (1997)
13. Blum, C., Roli, A.: Metaheuristics in combinatorial optimization: overview and conceptual comparison. ACM Comput. Surv. **35**(3), 268–308 (2003)
14. Jaeger, H.: The “echo state” approach to analysing and training recurrent neural networks. Technical Report 148, German National Research Center for Information Technology (2001)
15. Verstraeten, D., Schrauwen, B., D’Haene, M., Stroobandt, D.: An experimental unification of reservoir computing methods. Neural Netw. **20**(3), 287–289 (2007)
16. Lukoševičius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. Comput. Sci. Rev. **3**, 127–149 (2009)
17. Butcher, J.B., Verstraeten, D., Schrauwen, B., Day, C.R., Haycock, P.W.: Reservoir computing and extreme learning machines for non-linear time-series data analysis. Neural Netw. **38**, 76–89 (2013)

Hybrid Intelligent Approach to Solving the Problem of Service Data Queues

L.A. Gladkov, N.V. Gladkova and S.N. Leiba

Abstract In the article the problem of design optimization maintenance schedule queue of applications is considered. The analysis of the main features of queuing systems is carried out. A modified evolutionary algorithm of plotting service applications is developed. The authors suggested a new approach on the basis of evolutionary algorithm integration and a fuzzy control model of algorithm parameters. A fuzzy logical controller structure is described in the article. To confirm the method effectiveness a brief program description is reviewed.

Keywords Queuing system · The organization of the service queue of applications · Optimization · Genetic algorithm · Fuzzy logic

1 Introduction

The task of improving decision-making methods used in the production facilities of various branches of the control systems is urgent and important. The presence of a significant number of poorly formalized factors requires the development of new mathematical methods and software systems. In a separate class should be allocated queuing network operating in complex data subsystems. Such information subsystems widely used in transport, data networks, the various sectors of the economy.

L.A. Gladkov (✉) · N.V. Gladkova · S.N. Leiba
Southern Federal University, Rostov-on-Don, Russia
e-mail: leo_gladkov@mail.ru

N.V. Gladkova
e-mail: nadyusha.gladkova77@mail.ru

S.N. Leiba
e-mail: lejba.sergej@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 450, DOI 10.1007/978-3-319-33609-1_38

The use for such tasks deterministic mathematical models and methods is almost impossible because of the large number of control parameters and limitations of the system resources. As practice shows, these problems are solved effectively with evolutionary computation techniques. It is also proposed to use a hybrid approach, based on a combination of approaches. Hybridization can achieve "synergy effect", enhancing the advantages of embedded methods.

Consider the elements of well-known models of queuing systems. To describe the model of queuing systems is necessary to define the following parameters of the system [1–6]:

1. The input parameters of the service requests arriving in the modeled system. To specify the input to be considered during receipt of the application and the number of applications received at the same time. Applications may act on the basis of the determined schedule or with a certain probability. The queuing theory is commonly used job stream in which the number n of applications at any time interval is distributed according to the Poisson law:

$$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \; n \geq 0, \; t \geq 0, \tag{1}$$

   where λ—the number of applications per unit of time.
2. Principles of queuing and selection of applications from the queue. For the organization of the queue of applications can use different heuristics. For example, a rule setting orders in a queue in order of arrival (FIFO—"First In—First Out"). Or the opposite rule "last arrived—first served" (LIFO—"Last In—First Out"). It is also possible queuing on the set parameters or random. The length of the queue limit can be imposed on the queue (with a limited number of places) or at the time of stay in it.
3. Terms of service applications. Terms of service may vary by time of service and the number of concurrent applications. Service time may suitably be either deterministic or set based on the probability distribution law. Service can be organized via one or more devices. If the device for service applications are connected in series, it turns multiphase system service applications.
   Terms of Service define the conditions for termination of service, range of applications for maintenance, etc. The rules may specify the service mode with the priority or no priority. When mode without priority order of service is determined based on r selection rules used applications from the queue. Priority mode means that the application is assigned to a parameter (priority) in the form of a number or function.
   The application can be assigned to an absolute or relative priority. When applications with absolute priority applications with lower priority service is interrupted. For regulation of this process in the system interrupt set sometimes depth.
4. The parameters of the output flow requirements. The output stream—this stream applications (served or not served) that leave the system. The structure of the output stream is important for multiphase systems, as these requests are input to

the next phase of service. time distribution applications in the output depends on the input flow stream density and maintenance work of the devices.

5. Possible modes of operation of the system. Changing the system performance can be caused by external influences or the duration of the equipment. It is also possible to use the lock mode, when the application service process stops or slows down.

For any queuing systems is true, the so-called Little's Law, which can be summarized as follows [2]. At any time allocation between the two applications, any distribution of their service time, any number of devices, and any parameters of the average number of service requests of N depends on the intensity of their receipt λ and the mean residence time in the system T:

$$N = \lambda T \tag{2}$$

2 Problem Definition

Consider the problem of determining the characteristics of an open network operation queuing in a stationary mode, the network consists of M parallel unrelated junctions, between junctions K circulate requests of different types, each type has priority d_j $(1 \leq j \leq K)$, each i-th junction $(1 \leq i \leq M)$ contains parallel service channels. Each service unit has all applications with a limited number of places.

The applications received in the input queue group are sent to one of the junctions for service or queuing. At the current time, each junction processes the application of only one type, so the input queue of the application can be sent only to devices, applications serving the same class.

To switch to any i-th junction on another $(j + 1)$-class of applications requires a certain processing time for restructuring $t_{i,\ j+1}$.

The system implemented a rule setting orders in a queue in order of arrival (*FIFO*), and usually queuing using applications with absolute priority. The service time of all types of applications to any device has an arbitrary distribution. Scheme of interaction of individual network junctions is given full mesh graph G. The network receives K arbitrary streams of applications. The route passing applications in queuing network is given by the matrix:

$$P = \|p_{im,\,i+1}\|, \tag{3}$$

where $P_{im,i+1}$—the probability that a given node after the service. m class of the application enters the next $i + 1$ node.

The route through each type of application is determined based on the need to visit a chance to a specified number of devices, the values of the probability of visiting these devices due to the type of application.

It is necessary to solve the problem of building a schedule of service devices for different classes of applications that maximize the efficiency of network operation. At the same time the possible costs must be considered.

It can be a variety of indicators to assess the effectiveness of queuing systems. The most commonly used are as follows.

1. The probability that the application service is performed to the instrument;
2. Degree of loading of the serving system can be estimated based on the average number of used devices:

$$N_3 = \sum_{m=1}^{n} mp_m, \tag{4}$$

3. Another important indicator is the average number of free devices from the service:

$$N_o = \sum_{m=0}^{n-1} (n-m)p_m, \tag{5}$$

3 Algorithm Description

As a method of optimization will use evolutionary algorithms. Using evolutionary algorithms traditionally requires the following tasks. Determination procedure set of possible coding decisions evolutionary algorithm selection control parameter selection and modification of genetic operators.

We will apply a fixed length chromosome *N*, comprised of a set of M vectors, each i-th vector is a sequence of service requests for a given junction. The genes are ordered in time, so that each gene is rigidly attached to a specific point in time [7]. Example of chromosomes is shown in Fig. 1.

Each line of matrix—is a order of service of various types of applications. For example, line number 2 defines the following order of service applications. First,

	Application type				
Device 1	s_{1j}	$t_{1j,j+1}$	s_{1j+1}	…	s_{1N}
Device 2	s_{2j}	$t_{2j,j+1}$	s_{2j+1}	…	s_{2N}
………	…	…	…	…	…
Device *i*	s_{ij}	$t_{ij,j+1}$	s_{ij+1}	…	s_{iN}
………	…	…	…	…	…
Device *M*	s_{Mj}	$t_{Mj,j+1}$	s_{Mj+1}	…	s_{MN}

Fig. 1 The structure of chromosomes

		Application types				
		1	2	3	…	K
Applicat ion types	1	2	3	4	…	3
	2	1	1	5	…	1
	3	5	2	3		4
	………	…	…		…	…
	K	2	1	2	…	4

Fig. 2 An example of the switching matrix

the application is processed by the j-th class, then the device switches to the application process ($j + 1$)-th class, etc.

The duration of switching between the different classes of applications defined by the switching matrix (Fig. 2).

If P—population size, and K—the number of generations, the number of treated individuals $L = KP$. We assume the process of evolution complete, if the past generations N (N is assumed equal to 100) fitness of the best individual is not increased.

The main genetic operators are selection, crossover and mutation. The main genetic operators are selection, crossover and mutation. The selection operator performs in the genetic algorithm are two important tasks. Firstly, the range of solutions from the current population to which the crossover and mutation operators are used. After all the genetic operators in the current iteration of the algorithm, a new expanded population is formed. For further operation of the genetic algorithm is necessary to restore the original size of the population. In this algorithm, we will use a modified selection operator based on a combination of tournament and a elite selection in which of the two randomly selected individuals in the next generation of the best moves.

In our algorithm, we will use a modified single-point crossover operator. The operator is applied to each row of the matrix individually. In this case, randomly selected break point of the possible positions of the L-1 (L—length of the string). The selected point must always coincide with one of the positions corresponding to the time switching device between different classes. If randomly selected point corresponds to a different position, then we shift it to the left to the desired position. The chromosomes (parents) are divided at the breaking point. After an exchange of fragments we get two new chromosomes (offspring). After the crossover operator in the resulting offspring, in the break point insert new switching time between modes taken from the switching matrix. An example of the crossover operator is shown in Fig. 3.

Mutation operator is one of the main operators of the genetic algorithm. Using mutations tended to increase diversity in the population of solutions, and prevents premature convergence. Our algorithm uses a modified exchange mutation operator. To do this randomly selected position on the analogous lines in different

Parent 1	App_1	***3***	App_2	*5*	App_3	…
Parent 2	App_2	***1***	App_1	*4*	App_3	…
Offspring 1	App_1	***2***	App_1	*4*	App_3	…
Offspring 2	App_2	***1***	App_2	*5*	App_3	…

Fig. 3 Example of the crossover operator

Parent 1	App_1	$\boldsymbol{t_{1,5}}$	App_5	$t_{5,2}$	App_2	…
Parent 2	App_3	$t_{3,4}$	App_4	$\boldsymbol{t_{4,6}}$	App_6	…
Offspring 1	App_4	$t_{4,6}$	App_6	$t_{6,2}$	App_2	…
Offspring 2	App_3	$t_{3,1}$	App_1	$t_{1,5}$	App_5	…

Fig. 4 Example of the mutation operator

chromosomes and are exchanged. After an exchange of produce the necessary changes in the vicinity of the point of rupture and get two new lines. An example of the mutation operator shown us Fig. 4.

Selection and control values of the control algorithm parameters is carried out through the use of fuzzy logic controller. Control parameters in this case are the likelihood of implementing genetic operators [8, 9].

Thus, taking into account the use of the fuzzy logic controller (FLC) block diagram of the modified genetic algorithm can be represented as follows (Fig. 5).

Let us consider the work of fuzzy logic controller. It consists of the following blocks [10].

The rule base. Knowledge used for the correct functioning of the fuzzy control module, written in the form of fuzzy rules that represented as

$$R^k\colon \mathbf{IF}\left(x_1 \text{ is } A_1^k \ \mathbf{AND} \ldots \mathbf{AND}\ x_n \text{ is } A_n^k\right) \mathbf{THEN} \left(y \text{ is } B^k\right).$$

It is also possible to present this knowledge in the form of fuzzy sets with membership function defined by the expression:

$$\mu_{R^k}(x, y) = \mu_{A^k \to B^k}(x, y) \tag{6}$$

The output block. The membership function particular form of a fuzzy set is dependent on the applicable T-norms, definitions of fuzzy implication and the representation of fuzzy sets Cartesian product.

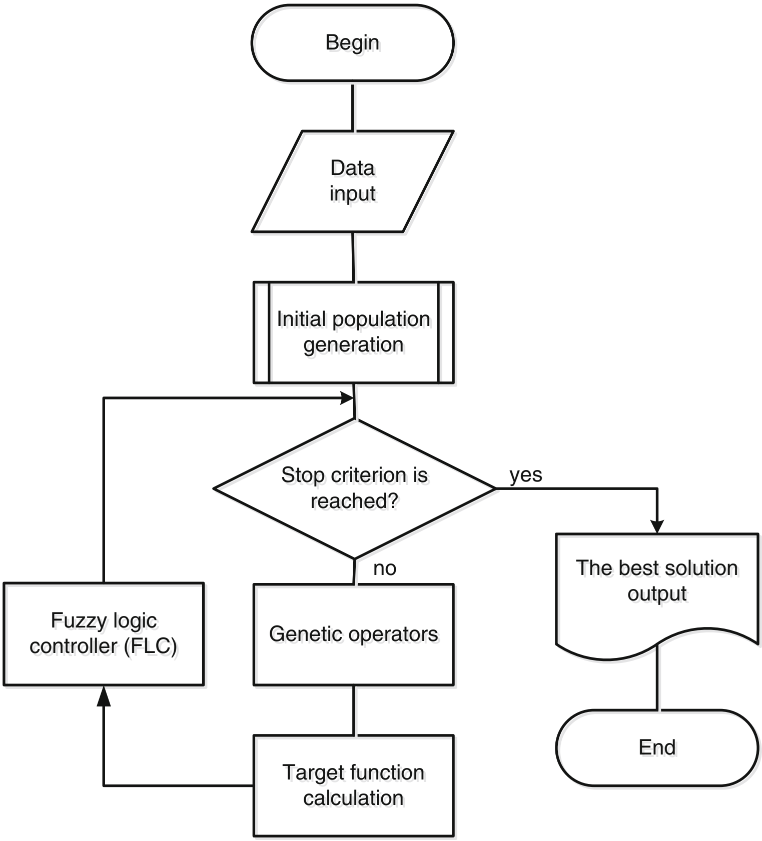


Fig. 5 Block diagram of the modified genetic algorithm

The fuzzification block. As a results of fuzzification operation to the FLC input parameter $\bar{x} = (\overline{X_1}, \overline{X_2}, \ldots, \overline{X_n})^T \in X$ put into correspondence the fuzzy set $A' \subseteq X = X_1 \times X_2 \times \cdots \times X_n$.

The defuzzification block. The reflection problem of fuzzy sets $\overline{B_k}$ (or fuzzy set B′) into a single value $\bar{y} \in Y$ is solved. This value represents the control action that affects on the FLC control objects. In the proposed algorithm the central average defuzzification is used.

To improve the quality of the genetic search results expert information is included in the evolution circuit by the FLC creation that adjust evolution parameters on the basis of expert knowledge.

As input parameters are used the best, average and worst values of the objective function [11, 12].

$$e_1(t) = \frac{f_{ave}(t) - f_{best}(t)}{f_{ave}(t)};$$

$$e_2(t) = \frac{f_{ave}(t) - f_{best}(t)}{f_{worst}(t) - f_{best}(t)};$$

$$e_3(t) = \frac{f_{best}(t) - f_{best}(t-1)}{f_{best}(t)};$$

$$e_4(t) = \frac{f_{ave}(t) - f_{ave}(t-1)}{f_{ave}(t)};$$

where t is a number of the current iteration; $f_{best}(t)$ and $f_{best}(t-1)$ are the best value of the target function at iterations t and $(t-1)$ accordingly; $f_{worst}(t)$ is the worst value of the target function at the iteration t; $f_{ave}(t)$ and $f_{ave}(t-1)$ are average value of the target function at iterations t and $(t-1)$ accordingly.

The allowed value range for e_1, e_2, e_3, e_4 is within intervals: $e_1 \in [0; 1]$; $e_2 \in [0; 1]$; $e_3 \in [-1; 1]$; $e_4 \in [-1; 1]$.

The values of variables e_1, e_2, e_3, e_4, affect the probability of performing crossover and mutation operators

$$Pc(t) = Pc(t-1) + \Delta Pc(t),$$

$$Pm(t) = Pm(t-1) + \Delta Pm(t).$$

4 Software Implementation and Experiments

To develop an application was selected Microsoft Visual Studio 2010. The simple program interface that shows the basic data, which should be output during program execution, was developed. This program interface is represented on Fig. 6.

The experiments were carried out. FLC parameters were obtained as a result of the learning algorithm. The set of experiments reflects the dependence of the probability of crossover and mutation operators on the values of e_1, e_2, e_3, e_4.

In this case the change in each parameter on the condition that other parameters remain unchanged, has little effect on the output value is determined only FLC ratios and defuzzification block y^k.

For example, if $y^k = \{0.4, 0.5, 0.6\}$ than e_4 has the following effect on the output value (Fig. 7).

To compare the effectiveness the test problems solved using the FLC and without it are investigated. Figure 8a, b showed that the efficiency of the algorithm with use the controller is much higher than the efficiency of the algorithm without it.

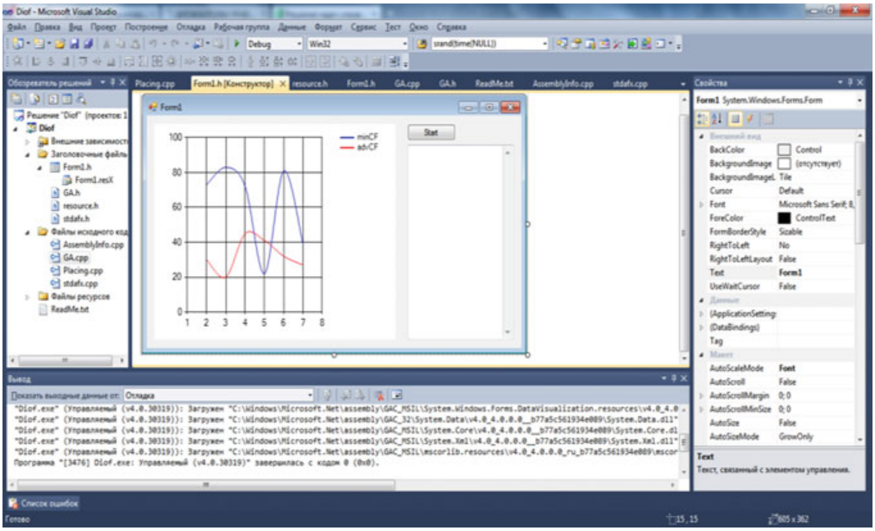


Fig. 6 The program interface

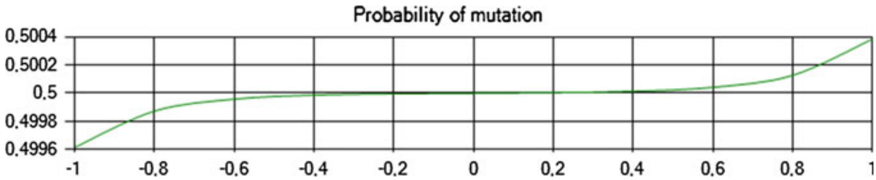


Fig. 7 The dependence of the probability values of the coefficients y^k

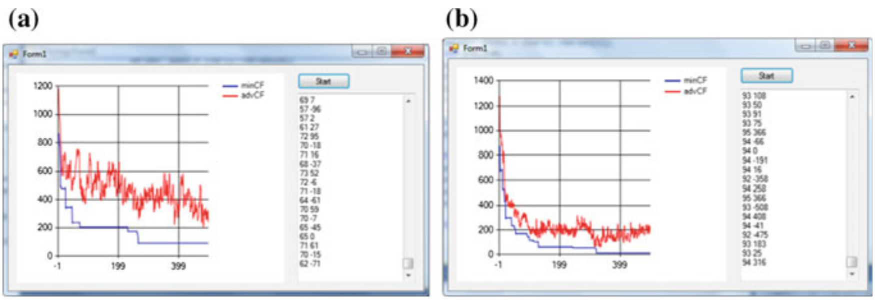


Fig. 8 The effectiveness of the algorithm without using FLC (**a**) and with it (**b**)

Efficiency of the controller is increased after the introduction of the training unit on the basis of an artificial neural network model. Coefficients parameters (x_i^k, σ_i^k and y^k) are given randomly or based on the direct search.

5 Conclusion

FLC provides a probability value crossover and mutation operators based on the 4 input variables which characterize the process of search. In this case, the dependence of the output values of the input variables is determined by the parameters of the FLC. The number of these parameters depends on the number of input variables, the number of membership functions and parameters which define each membership function.

The studies were conducted for 4 input variables and 3 membership functions for each variable. In addition, each membership function defined by two parameters. Therefore, fuzzy logic controller, when determining the output value handles 28 parameters. Since in our case 2 output parameters were calculated (probability of crossover and mutation), the FLC behavior is determined by the 56 parameters. For an expert it would be problematic manually set these parameters so as to obtain the required behavior of FLC.

In this situation, it is necessary to develop an interface that would allow operating the linguistic variables to specify the desired behavior. FLC parameters that were used in the study were obtained using genetic algorithm learning. Training was carried out on the basis of statistical information on the dependence of the FLC parameters and the efficiency of the algorithm placement. This information is collected during the learning process.

Acknowledgment This research is supported by grants of the Russian science foundation (project № 14-11-00242) in Southern Federal University.

References

1. Dshalalow, J.H. (ed.): Advances in Queueing Theory, Methods, and Open Problems (Probability and Stochastics Series). CRC Press (1995)
2. Asmussen, S. (ed): Applied Probability and Queues (Stochastic Modeling and Applied Probability Series Vol. 51). Springer (2010)
3. Gross, D., Shortle, J.F., Thompson, J.M., Harris, C.M.: Solutions Manual to Accompany Fundamentals of Queueing Theory, 4th edn. Wiley (2008)
4. Bhat, N.U.: An Introduction to Queueing Theory: Modeling and Analysis in Applications (Statistics for Industry and Technology). Birkhäuser (2008)
5. Bose, S.K.: An Introduction to Queueing Systems (Network and Systems Management). Kluver Academic/Plenum Publisher, New York (2002)
6. Ng, C.-H., Boon-Hee, S.: Queueing Modelling Fundamentals: With Applications in Communication Networks, 2nd edn. Wiley (2008)
7. Gladkov, L.A., Kureychik, V.V., Kureychik, V.M.: Genetic Algorithms. Fizmatlit, Moscow (2010)
8. Herrera, F., Lozano, M.: Fuzzy adaptive genetic algorithms: design, taxonomy, and future directions. Journal of Soft Computing, pp. 545–562. Springer-Verlag (2003)
9. Michael, A., Takagi, H.: Dynamic control of genetic algorithms using fuzzy logic techniques. In: Proceedings of the Fifth International Conference on Genetic Algorithms, pp. 76-83. Morgan Kaufmann (1993)

10. Batyrshin, I.Z., Nedosekin, S.A.: Fuzzy Hybrid Systems. Theory and Practice. Fizmatlit, Moscow (2007)
11. Gladkov, L., Gladkova, N., Leiba, S.: Manufactoring scheduling problem based on fuzzy genetic algorithm. In: Proceeding of IEEE East-West Design & Test Symposium—(EWDTS'2014), pp. 209–212, Kiev, Ukraine (2014)
12. Glagkov, L.A., Glagkova, N.V., Leiba, S.N.: Electronic computing equipment schemes elements placement based on hybrid intelligence approach. In: Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015), Vol. 2: Software Engineering in Intelligent Systems, № 348, pp. 35–45

Artificial Bee Colony Algorithm—A Novel Tool for VLSI Placement

Daria Zaruba, Dmitry Zaporozhets and Vladimir Kureichik

Abstract The paper discusses a modified algorithm based on the bees' behavior in nature. We suggest to apply this algorithm for solving the element placement problem—one of the most difficult problem in the VLSI design. This problem belongs to the NP-class problem that is there are no precise methods to solve this problem. Also we formulate the placement problem and choose an optimization criterion. The developed artificial bee colony optimization (ABC) algorithm obtains optimal and quasi-optimal solutions during polynomial time. The distinguish feature of the algorithm is to split search space into subareas. After, we activate optimization process for each subarea parallel. To compare obtained results with known analogous algorithms we developed software which allows to carry out experiments on the basis of IBM benchmarks. Conducted experiments shown that the ABC algorithm is better than the other algorithms an average of 8 %.

Keywords VLSI · Placement problem · ABC algorithm · Parallel optimization

1 Introduction

Currently, rapid development of VLSI manufacturing is a reason for creation a novel effective tools for automated design which allow to design integrated circuit with millions transistors on a chip.

D. Zaruba · D. Zaporozhets · V. Kureichik (✉)
Southern Federal University, Rostov-on-Don, Russia
e-mail: vkur@sfedu.ru; vkur@tgn.sfedu.ru

D. Zaruba
e-mail: daria.zaruba@gmail.com

D. Zaporozhets
e-mail: elpilasgsm@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 450, DOI 10.1007/978-3-319-33609-1_39

There is a accepted sequence of design stage: a system specification, a functional design, an algorithmic design, an engineering design, a process design, a packaging, a testing and verification [1, 2].

The engineering design contains results obtained at the functional design stage. But the process design is connected with results at the engineering design stage, in particular, a development of technological process of device manufacturing.

At the engineering design stage a circuit representation of each component is transformed to a geometrical representation (a set of geometrical images with special functions according to components). Geometrical images are connection between different circuit components.

The VLSI placement is a one of the most complex problem in the course of engineering design because of modern VLSI circuits contain several hundreds of logical units, and a circuit occupies about 80 % of its area. Current rapid development of information technologies requires new automated algorithms which would allow to find effective solutions within the reasonable CPU time. Heuristic swarm algorithms meet these conditions and allow to find optimal solutions of engineering design problems. So, the development of such algorithm is important and promising problem under current conditions. In this paper the authors suggest an innovative artificial bee colony (ABC) algorithm for the VLSI placement problem.

2 VLSI Placement Problem Formulation

In general case initial information include:

- connection field;
- electric schematic diagram;
- VLSI fragments;
- a net list.

The formulation of VLSI placement problem is as follows. Let $X_1, \ldots, X_n$ be VLSI fragments which should be placed within connection field. Each element $X_i \mid 1 \le i \le N$ is set by geometrical sizes: height h_i and width w_i. Let $N = \{n_i \mid i = 1, m\}$ be a net list connecting fragments, L_i is a length of the net $n_i \mid i = 1, m$. The placement problem involves a search of rectangular areas within connection field for each fragment and defines by a set of areas $R = \{r_i \mid i = 1, n\}$ such that each fragment is situated in corresponding rectangular area according with the following statement:

$$r_i = \langle u_i, t_i, w_i, h_i \rangle, \tag{1}$$

where r_i has a height h_i, a width w_i and coordinates u_i, t_i.

Each net n_i is a sequential list of circuit elements connection (tuple) and represented as

$$n_i = \langle R_{n_i} \rangle, R_{n_i} \subseteq R, \tag{2}$$

where R_{n_i} is areas contains in net n_i.

As a criterion we suggest a classical criterion with respect to VLSI placement problems—total length of connections. This criterion is widely used for estimation of solutions and simplified the conducting of experiments. So, the total length criterion is the following:

$$F(N) = \sum_{i=1}^{\mathrm{m}} f(n_i), \tag{3}$$

where N is a set of placed fragments, n_i is a i-th net, $m = |N|$ is a number of nets.

$$f(n_i) = L_i = \sum_{j=1}^{|n_i|-1} d(r_j, r_{j+1}), \tag{4}$$

where $d(r_j, r_{j+1})$ is a distance between two neighbor fragments in the net.

$$d(r_j, r_{j+1}) = \sqrt{\left(u_{r_j} - u_{r_{j+1}}\right)^2 + \left(t_{r_j} - t_{r_j+1}\right)^2}. \tag{5}$$

3 The Artificial Bee Algorithm for VLSI Placement Problem

The main idea of the ABC algorithm was borrowed from bee behavior when they looking for promising areas with nectar. Bees can occupy a large area and use several source of food simultaneously [3–5]. Bees-foragers are placed on flowers area with nectar, collection of which requires less effort. These areas should be visited by a large number of bees, while areas with fewer quantity of nectar visit fewer bees [5].

Scout-bees start foraging i.e. they are looking for promising flowers areas. They are moving from one area to another in a random way. Colony continues investigation of search area maintaining the permanent percent of scouts. A model of bees' behavior in nature is presented in Fig. 1.

As mentioned previously, the ABC algorithm is an optimization algorithm, inspired by bees' behavior during the food search, which allow to find optimal solutions of given problem. A global extremum is an area with the greatest number

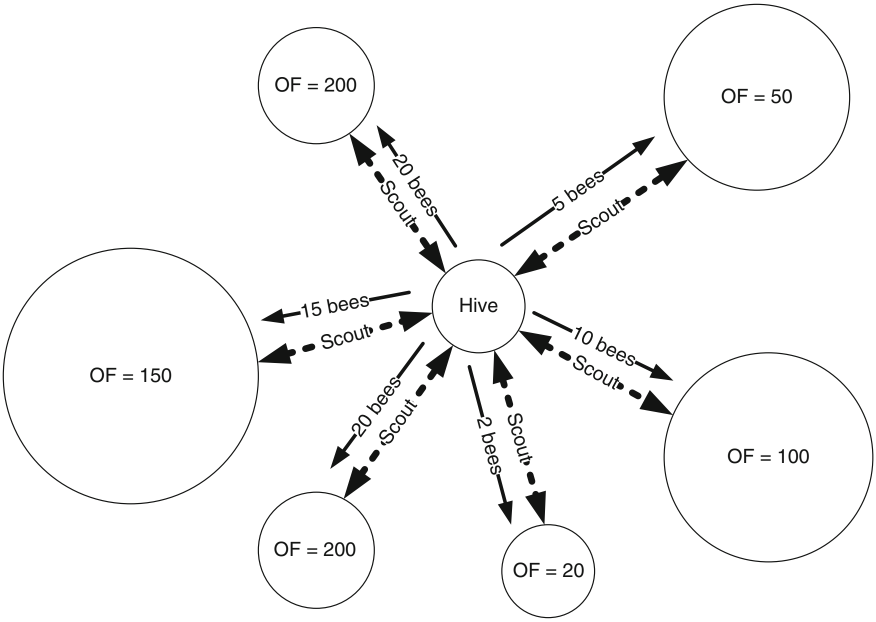


Fig. 1 The model of bees' behavior in nature (OF—objective function)

of nectar. This area is unique, i.e. other area have nectar too but fewer. Bees are in multidimensional search space, i.e. each coordinate assign one parameter of objective function (OF). Nectar denotes a value of OF in this point.

In the ABC algorithm each alternative solution is determined by "bee" with coordinates and parameters of a multidimensional function which corresponds to a particular area in search space [5].

At the first step of the ABC algorithm scouts are exploring areas with random coordinates. The objective function is denoted by bee's coordinates on the basis of which there are chosen promising areas near global optimum.

After assignment of foragers on best areas, remain bees (scouts) explore other random points of search space.

After this, the best value, which is an intermediate solution, is saved. Than the algorithm is repeated until a spot criterion has been reached.

To place VLSI fragments in optimal way the authors suggest the algorithm based on bees' behavior (Fig. 2).

The first stage is to set control parameters such as a number of iteration T, a number of forager-agents A_f, a number of scout-agents A_s. Next, an initial population is generated in any way, for example, random generation of alternative solutions or results of execution of other algorithms. On the basis of OF values

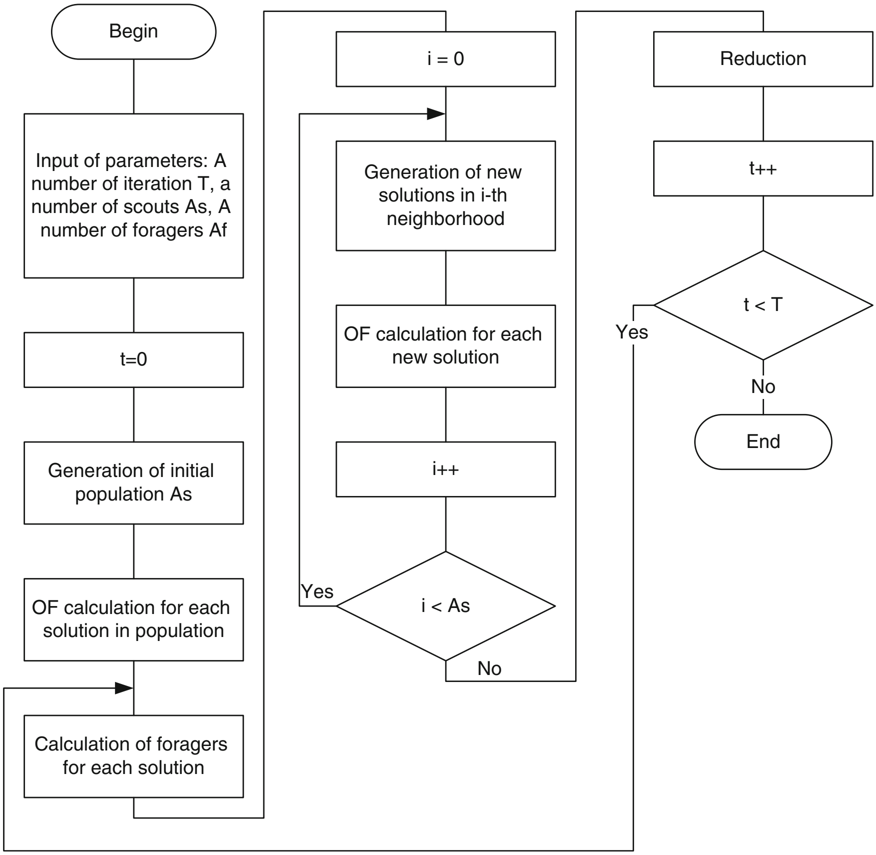


Fig. 2 VLSI fragments placement based on the bees' behavior

there is calculated a number of foragers which will explore search space in solution neighborhoods. A number of foragers for *i*-th population is defined as follows:

$$A_F^i = A_F * \frac{F_i}{\sum_{j=1}^{As} F_j}, \tag{6}$$

where F_i is an OF value for *i*-th population; $\sum_{j=1}^{As} F_j$ is a total OF value for each population.

This involves generation of new solutions differing from initial one only a pair permutation. Permutations conform to the following rule: a gene-operator can

1	2	-1	4	3	-1	-1	6	5	-2	-2	F_1=10
1	2	-1	4	-1	3	-1	5	-2	6	-1	F_2=15
1	2	-1	4	-1	3	6	-1	5	-2	-1	F_3=40
1	2	-1	4	3	-1	-2	6	-2	5	-1	F_4=21
1	2	-1	3	4	-2	5	-1	6	-2	-1	F_5=5

Fig. 3 Initial population of alternative solutions

interchange only with a gene-operator, and a gene-operand—only with a gene-operand. These genes are selected in random way. As an example, we have consider the population with five solutions $A_S = 5$ (Fig. 3).

Let a total number of forages be $A_f = 20$. So, we can calculate a number of foragers for each solution.

$A_f^1 = 20 \times 10/(10 + 15 + 40 + 21 + 5) = 200/91 \approx 2$
$A_f^2 = 20 \times 15/(10 + 15 + 40 + 21 + 5) = 300/91 \approx 3$
$A_f^3 = 20 \times 40/(10 + 15 + 40 + 21 + 5) = 800/91 \approx 9$
$A_f^4 = 20 \times 21/(10 + 15 + 40 + 21 + 5) = 410/91 \approx 5$
$A_f^5 = 20 \times 5/(10 + 15 + 40 + 21 + 5) = 100/91 \approx 1$

Let perform pair permutations. On the basis of each initial solution there is generated a set of new solutions A_f^i by pair permutation of genes. As a result we obtain:

For the first solution:

$\langle$ 3, 2, −1, 4, 1, −1, −1, 6, 5, −2, −2 $\rangle$ $F = 6$, $\langle$1, 2, −2, 4, 3, −1, −1, 6, 5, −1, −2$\rangle$ $F = 11$.

For the second solution:

$\langle$1, 4, −1, 2, −1, 3, −1, 5, −2, 6$\rangle$ $F = 16$, $\langle$1, 2, −1, 4, −1, 3, −2, 5, −1, 6$\rangle$ $F = 33$, $\langle$1, 2, −1, 4, −1, 6, −1, 5, −2, 3$\rangle$ $F = 12$.

For the third solution:

$\langle$1, 2, −1, 3, −1, 4, 6, −1, 5, −2, −1$\rangle$ $F = 25$, $\langle$1, 2, −2, 4, −1, 3, 6, −1, 5, −1, −1$\rangle$ $F = 31$, $\langle$1, 6, −1, 4, −1, 3, 2, −1, 5, −2, −1$\rangle$ $F = 15$, $\langle$1, 2, −1, 4, −1, 3, 6, −1, 5, −1, −2$\rangle$ $F = 43$, $\langle$5, 2, −1, 4, −1, 3, 6, −1, 1, −2, −1$\rangle$ $F = 40$, $\langle$1, 2, −1, 3, −1, 4, 6, −1, 5, −2, −1$\rangle$ $F = 12$, $\langle$1, 2, −1, 4, −1, 6, 3, −1, 5, −2, −1$\rangle$ $F = 35$, $\langle$1, 4, −1, 2, −1, 3, 6, −1, 5, −2, −1$\rangle$ $F = 31$, $\langle$3, 2, −1, 4, −1, 1, 6, −1, 5, −2, −1$\rangle$ $F = 17$.

For the forth solution:

$\langle$1, 2, −1, 4, 3, −2, −1, 6, −2, 5, −1$\rangle$ $F = 16$, $\langle$3, 2, −1, 4, 1, −1, −2, 6, −2, 5, −1$\rangle$ $F = 25$, $\langle$1, 2, −1, 4, 3, −1, −2, 5, −2, 6, −1$\rangle$ $F = 15$, $\langle$1 ,2, −1, 6, 3, −1, −2, 4, −2, 5, −1$\rangle$ $F = 6$, $\langle$5, 2, −1, 4, 3, −1, −2, 6, −2, 1, −1$\rangle$ $F = 18$.

1	2	-2	4	3	-1	-1	6	5	-1	-2	F_1=11
1	2	-1	4	-1	3	-2	5	-1	6	-1	F_2=33
1	2	-1	4	-1	3	6	-1	5	-1	-2	F_3=43
3	2	-1	4	1	-1	-2	6	-2	5	-1	F_4=25
1	2	-1	3	6	-2	5	-1	4	-2	-1	F_5=9

Fig. 4 Population of alternative solutions after the first iteration

For the fifth solution:

$\langle 1, 2, -1, 3, 6, -2, 5, -1, 4, -2, -1 \rangle$ $F = 9$.

To ensure the algorithm convergence the reduction operator is applied. At each iteration this operator selects the best solution to for a new population of solutions (Fig. 4). This process continues iteratively until a determiner number of iteration will be reached. As a result we obtain one or more quasi-optimal solutions with the best OF value.

4 Experiments

To confirm the efficiency of the ABC algorithm we developed software which allow to estimate the suggested algorithm with known analogs such as Capo 8.6, Feng Shui 2.0, Dragon 2.23 [6–8] on the basis of IBM benchmarks [9]. Obtained results are represented in Table 1 and on Figs. 5, 6.

5 Conclusion

In this paper we suggested and researched the original swarm algorithm for VLSI fragments placement on the basis of bees' behavior in nature. This method allows to control the algorithm convergence by means of search space dividing on subareas and implementation of parallel search in it. Besides that parallel search increase a probability of the algorithm convergence in a global optimum. Experiments shown that the algorithm is less efficient for low dimension problems (less than 15000 elements), but in the average quality of the obtained solutions more than 8 %.

Table 1 Comparison of benchmarks ibm01–ibm13 placement by Capo 8.6, Feng Shui 2.0, Dragon 2.23 and ABC algorithms

Benchmarks		Capo 8.6	Feng Shui 2.0	Dragon 2.23	ABC algorithm						
Name	A number of elements	Length, m	Length, m	Length, m	Length, m	Increment in length (Capo 8.6), m	Increment in length (Capo 8.6), %	Increment in length (Feng Shui 2.0), m	Increment in length (Feng Shui 2.0), %	Increment in length (Dragon 2.23), m	Increment in length (Dragon 2.23), %
ibm01	12752	4.97	4.87	4.42	4.43	−0.54	−12.19	−0.44	−9.93	0.01	0.23
ibm02	19601	15.23	14.38	13.57	13.6	−1.63	−11.99	−0.78	−5.74	0.03	0.22
ibm03	23136	14.06	12.84	12.33	12.3	−1.76	−14.31	−0.54	−4.39	−0.03	−0.24
ibm04	27507	18.13	16.69	15.41	15.22	−2.91	−19.12	−1.47	−9.66	−0.19	−1.23
ibm05	29347	44.73	37.3	36.38	36.2	−8.53	−23.56	−1.1	−3.04	−0.18	−0.49
ibm06	32498	21.96	20.27	20.38	19.85	−2.11	−10.63	−0.42	−2.12	−0.53	−2.60
ibm07	45926	36.06	31.5	29.97	28.95	−7.11	−24.56	−2.55	−8.81	−1.02	−3.40
ibm08	51309	37.89	34.14	32.2	31.4	−6.49	−20.67	−2.74	−8.73	−0.8	−2.48
ibm09	53395	30.28	29.86	28.1	27.1	−3.18	−11.73	−2.76	−10.18	−1	−3.56
ibm10	69429	61.25	57.99	57.2	57.3	−3.95	−689 %	−0.69	−1.20	0.1	0.17
ibm11	70558	46.45	43.28	40.77	41.1	−5.35	−13.02	−2.18	−5.30	0.33	0.81
ibm12	71076	81.55	75.91	71.03	68.88	−12.67	−18.39 %	−7.03	−10.21 %	−2.15	−3.03 %
ibm13	84199	56.47	54.09	50.57	48.55	−7.92	−16.31 %	−5.54	−11.41 %	−2.02	−3.99 %

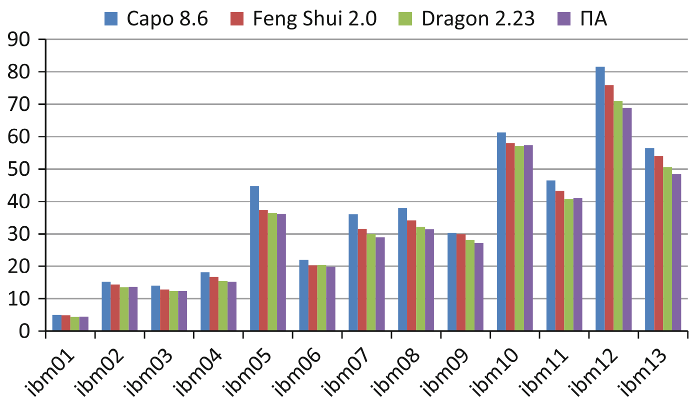


Fig. 5 The length of the wire when placing benchmarks ibm01–ibm 13 by different algorithms

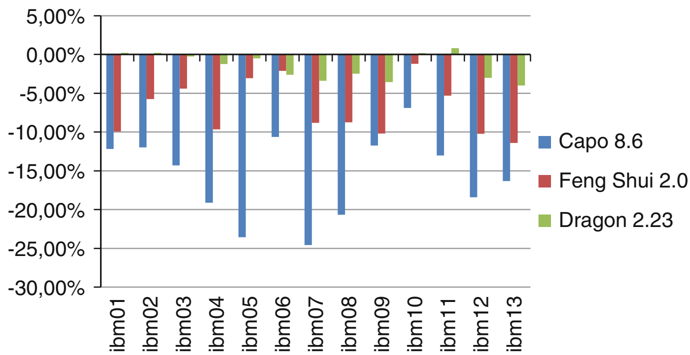


Fig. 6 The increment of wire length when placing benchmarks ibm01–ibm13 by different algorithms

Acknowledgements This research is supported by grants of the Ministry of Education and Science of the Russian Federation, the project # 8.823.2014.

References

1. Alpert, C., Mehta, D., Sapatnekar, S.: Handbook of Algorithms for Physical Design Automation. Auerbach Publications, New York, USA (2009)
2. Kureichik, V.V., Kureichik, V.M., Malioukov, S.P., Malioukov, A.S.: Algorithms for Applied CAD Problems. Springer, Heidelberg (2009)
3. Markov, I.L., Hu, J., Kim, M.-C.: Progress and Challenges in VLSI Placement Research. Proc. IEEE **103**, 1985–2003 (2015)
4. Karaboga, D.: An idea based on honey bee swarm for numerical optimization. Technical Report TR06, Erciyes University, Engineering Faculty, Computer Engineering Department (2005)

5. Kureichik, V.V., Zaruba, D.V.: The bioinspired algorithm of electronic computing equipment schemes elements placement. In: Silhavy, R., Senkerik, R., Oplatkova, Z. K., Prokopova, Z., Silhavy, P. (eds.) 4th Computer Science On-line Conference, CSOC 2015б, vol. 347, pp. 51–58. Springer (2015)
6. Caldwell, A.E., Kahng, A.B., Markov, I.L.: Can recursive bisection alone produce routable placements? DAC **2000**, 477–482 (2000)
7. Wang, M., Yang, X., Sarrafzadeh, M. Dragon2000: Standard-cell Placement Tool for Large Industry Circuits, ICCAD2000, pp. 260–263 (2000)
8. Yang, X., Choi, B.-K., Sarrafzadeh, M.: Routability driven white space allocation for fixed-die standard-cell placement. ISPD **2002**, 42–50 (2002)
9. IBM-PLACE 2.0 benchmark suits. http://er.cs.ucla.edu/-benchmarks/-ibm-place2/bookshelf/-ibm-place2-all-bookshelf-nopad.tar.gz

Mechanisms of Adaptive Ant Colony Behavior in Placement Problem

Boris K. Lebedev, Oleg B. Lebedev and Tatiana Y. Kudryakova

Abstract New technologies, principles and mechanisms to solve placement problem are provided in the paper. They are based on adaptive ant colony behavior modeling. Solution search graph structure, methods of pheromone precipitation and evaporation. Concept of element placement by pair of consequences is used in the paper. Comparing with existing results some improvement is gained. The paper provides a solution for placing different-sized elements. Method is based on pair of element consequence and on adaptive ant colony behavior modeling Dorigo (Ant Colony Optimization. MIT Press, Cambridge 2004, [8]). Obtained results allow us to make a conclusion on potential further development and application of ant colony algorithms to tasks of optimal placement.

Keywords Location problem · Adaptive behavior · Pair of sequences

1 Introduction

An important stage in topology design is an element placement in the connecting field (CF). To locate means to define a distribution of a module, cell or block in a designed device or just single topological elements on a plate of integrated chip [1]. Recently nanotechnology methods and new packaging schemes (like multichip modules for complicated tasks and placement tasks) have appeared. Such technologies actively use methods based on artificial intelligence. Computational complexity and high dimensionality of placement tasks sometimes do not allow to

B.K. Lebedev (✉) · O.B. Lebedev · T.Y. Kudryakova
Southern Federal University, Rostov-on-Don, Russia
e-mail: lebedev.b.k@gmail.com

O.B. Lebedev
e-mail: lebedev.ob@mail.ru

T.Y. Kudryakova
e-mail: tatiana.kudryakova@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 450, DOI 10.1007/978-3-319-33609-1_40

get an optimal solution within a reasonable time. As a result to develop approximate solution methods becomes a key importance. High interest is shown to algorithms based on natural and physical processes and especially to algorithms inspired by natural systems. The newest tendencies are multiagent methods of intellectual optimization based on swarm intellect modeling. These methods are genetic algorithms, simulated annealing algorithm, swarm methods—ant and bee colony algorithms [2–5].

The most widespread method to place different-size rectangle elements is connected with constructing restriction graph for mutual element placement or for ordered tree [6]. An effective way to represent restriction graph might be a proposed method of pair element sequences [7]. A common idea is firstly to place blocks in the grid and then to use a longer route in restriction graph in order to generate a nested placement. To define block placement, 2 "name" block sequences are generated, they correspond to horizontal and vertical grid lines. As a result the solution search in placement task appears to be a synthesis of pair element sequences. Mutual placement of each pair of elements in both sequences defines a mutual placement of element pair in the connecting field.

The paper provides a solution to place different-sized elements. Method is based on pair element consequence and on adaptive ant colony behavior modeling [8]. Obtained results allow us to make a conclusion on potential further development and application of ant colony algorithms to tasks of optimal placement.

2 Fundamentals

Placement problem is formed the following way. Set of elements (modules) with terminals(outputs) is given. Set of chains connecting module terminals is given. Connecting field (CF) where elements might be placed is given. It is needed to place elements in the CF by optimizing some quality criteria. Input information includes module description with form, sizes and terminal placement in modules; list of chains determining the module intercircuit connection links and CF description. Output information is a list of X, Y coordinates in CF of all the modules.

The main aim of placement problem is to minimize total chip area, to create facilities for the subsequent (to solve routing problem) and to minimize the estimation of total intercircuit connection length. To minimize intercircuit connection length is required in order to reduce time delays in long chains and to speed up information processing in GSI.

Let us examine element placement concept based on pair of sequences [7]. Pair of sequences A_1 and A_2 is a pair of ordered lists of the same set of elements. Pair of sequences (A_1 = 4, 3, 1, 6, 2, 5; A_2 = 6, 3, 5, 4, 1, 2) is given. Formation of restriction graph for the given pair of lists is proceeded by sequential meta-grid construction with a slope of 45 degrees as shown on the Fig. 1.

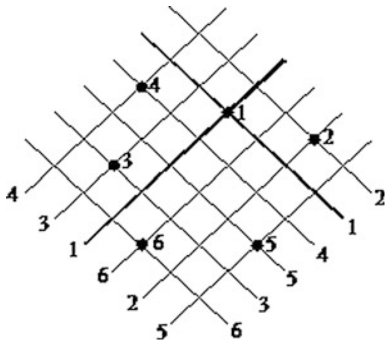


Fig. 1 The structure of sequence pair

For each element area is divided by 2 inclined lines into 4 sectors as it is shown for the 1st element on the Fig. 1. Sequential pair determines a relation between element pair a_i and a_j containing in these lists the following way:

$$\text{If } \left(A_1 = \langle \ldots, a_i, \ldots, a_j, \ldots \rangle, A_2 = \langle \ldots, a_i, \ldots, a_j, \ldots \rangle\right), \tag{1}$$

then a_j is placed to the right from $a_{i;}$

$$\text{If } \left(A_1 = \langle \ldots, a_i, \ldots, a_j, \ldots \rangle, A_2 = \langle \ldots, a_j, \ldots, a_i, \ldots \rangle\right), \tag{2}$$

then a_j is placed below a_i.

Let us examine mutual placement of elements concerning element 1.

Concerning pair of sequences element 2 is placed in the right sector from the element 1 as in both sequences 2 is situated to the right from 1 (Fig. 1). Elements 6 and 5 are placed in the lower sector concerning element 1, and so on.

Thus having a pair of sequences (A_1, A_2) determining horizontal relation between the elements, one can construct a horizontal restriction graph $G_h(V_h, E_h)$ as follows:

$$V_h = \{ s_h\}*\{ t_h\}*\{v_i | i = 1..n \}, \tag{3}$$

$$E_h = \{(s_h, v_i) | i = 1 \ldots n \}*\{(v_i, t_h) | i = 1 \ldots n \}*\left\{\left(v_i, v_j\right) |\ a_i \text{ is to the left from } a_j\right\}; \tag{4}$$

where v_i corresponds to element, s_h is initial vertex representing a left border, t_h is the end vertex representing a right border; vertex weight v_i is equal to element a_i width. Weight of vertexes s_h, t_h is equal to 0.

Vertical restriction graph $G_v(V_v, E_v)$ is constructed the same way.

For example, Fig. 2 represents both graphs for placement depicted on the Fig. 1. Both graphs are directed, acyclic and have weighted vertexes. That is why we might apply searching-the-longest-route algorithm to determine XУ coordinates for every element. Let's consider left lower corner's element coordinate as element's coordinate.

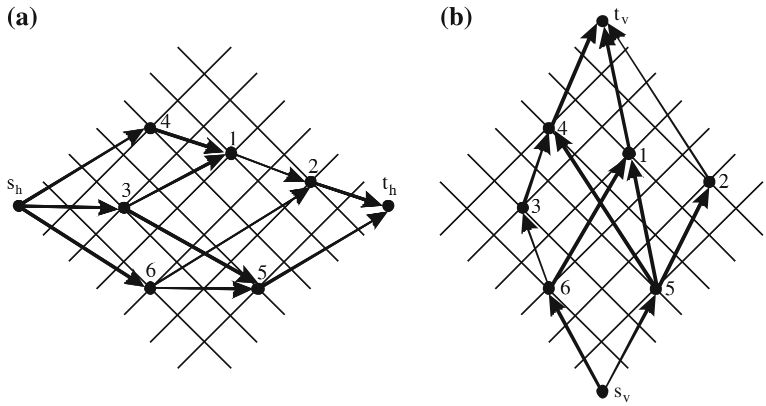


Fig. 2 Horizontal (**a**) and vertical (**b**) restriction graphs

Time $O(n^2)$ is required to construct restriction graph. To compute the longest route $O(n + m)$ time is needed, where n is the number of vertexes, and m is the number of edges in the graph. Thus the total time required to convert sequential pair to placement is $O(n^2)$.

Transition from pair of sequences to placement is made in 2 stages using concept of pair of sequences. Firstly transition to restriction graph or to ordered restriction tree (OT) covering restriction graph is performed. Then using the trivial algorithm walking through the restriction tree the plan or element placement are built.

If the vertexes *a, b* and *c* are in such relations that $a \rightarrow b$, $b \rightarrow c$ and $a \rightarrow c$, then the transitive edge $a \rightarrow c$ is redundant and might be deleted in restriction graph. To delete redundant edges the procedure of deleting transitive edges is performed. Trivial procedure of obtaining acceptable placement over OT is described in the paper [6]. The essence consists in successive sampling of the element in restriction graph and shifting it down to the limit and then to the left limit. With the initial placement we might proceed to acceptable element placement along the left and low sides. When acceptable element placement cannot be moved neither down nor to the left (Fig. 3).

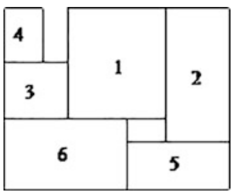


Fig. 3 Acceptable location

3 Organizing Search Procedures Based on Ant Colony Adaptive Behavior Modeling

Element set $A = \{a_i | i = 1, 2,\ldots,n\}$ is given. As a scheme model hypergraph $H = (X, E)$ is used where $X = \{x_i | i = 1, 2,\ldots,n\}$ is a set of vertexes that model elements, and $E = \{e_j | e_j \subset X, j = 1, 2,\ldots,m\}$ is a set of hyper edges that model chains linking the elements. Elements have rectangle shape. Description of element sizes is given. Solution search is carried on 2 solution search graphs $G^*_1 = (X^*_1, U_1)$ and $G^*_2 = (X^*_2, U_2)$, where $X^*_1 = X_1 \cap S_1$, $X^*_2 = X_2 \cup S_2$ (Fig. 4).

Vertexes from the set $X_1 = \{x_{1i} | i = 1, 2,\ldots, n\}$ correspond to modules a_{1i} placed in the first sequence. Vertexes from the set $X_2 = \{x_{2i} | i = 1, 2,\ldots, n\}$ correspond to modules a_{2i} placed in the second sequence. Sets X_1 and X_2 form complete subgraphs G_1 and G_2 with non-directed edges.

Vertexes S_1 and S_2 are initial. Directed edges connect vertex S_1 with every vertex from the set X_1, and the same way vertex S_2 is connected with every vertex from the set X_2. Generally synthesis of pair of sequences is carried by pairs of ants $Z = \{z_k | k = 1, 2,\ldots,n\}$, where $z_k = \langle z_{k1}, z_{k2} \rangle$. At every iteration of the ant algorithm every ant pair z_k starts from initial vertexes S_1 and S_2. They construct their solution—pair of routes M_1 and M_2 on subgraphs G_1 and G_2 (Fig. 5).

Route M_1 includes all the vertexes of the set X_1. Route M_2 includes all the vertexes of the set X_2.

Number of solutions constructed by ants at every iteration equals n. Modeling ant behavior is related to pheromone distribution on graphs' G^*_1 and G^*_2 edges. Initially all the edges are marked with the same (small) quantity of pheromone $\theta = Q/v$, where $v = |U|$. Parameter Q is set a priori. Solution search process is iterative. Every iteration l includes 3 stages [9].

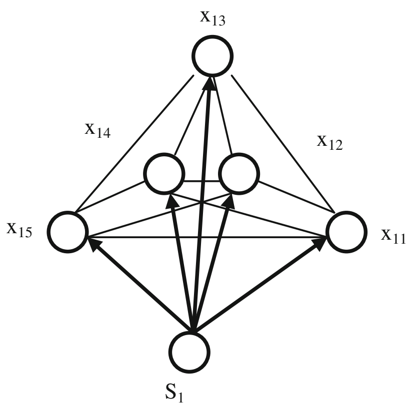


Fig. 4 Solution search graphs

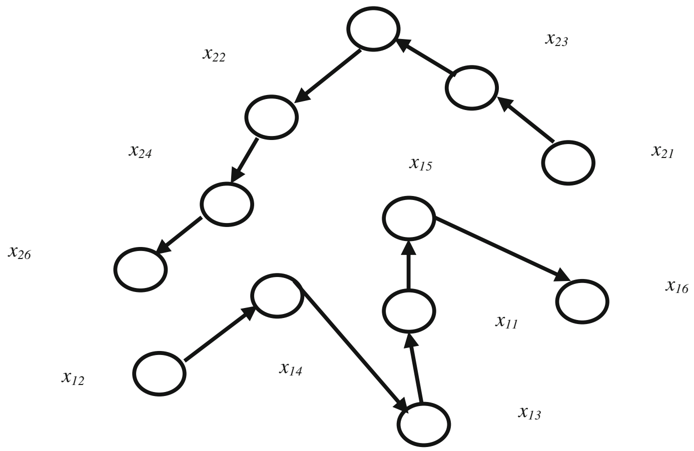


Fig. 5 Routes M_1 *and* M_2 constructed on graphs by pair of ants

Table 1 Experiments

Benchmark	Algorithm					
	GEN	AD	HY	MA	AT	SPP
apte	49425136,4	49425136,4	49396354	49396354	49396354	49468310
xerox	24954567,8	24954567,8	24895479	24873723	24957366	24946938
hp	12232485	12232485	12194845	12176343	12393236	11980127
ami33	3510173,4	3510173,4	3538838	3525858	3490074	3465157
ami49	136133614,4	136133614,4	137343334	134285950	138591325	132488414

At first stage each ant pair finds solution—routes M_1 and M_2 in graphs G^*_1 and G^*_2. At the second stage each ant pair $\langle z_{k1}, z_{k2} \rangle$ marks with pheromone edges of the constructed routes M_1 and M_2. One should mention that all the edges are marked with the same quantity of pheromone that corresponds to solution estimate. At the third stage pheromone evaporates from the graphs' G^*_1 and G^*_2 edges. Cyclic method of ant systems (ant-cycle) is used in the paper. In this case pheromone is laid on the edges just after the complete solution is constructed.

Process of constructing routes M_{1k} and M_{2k} is held step by step. Let's examine the constructing of route M_{1k}. (M_{2k} is constructed similarly). At every step t agent z_{k1} applies probability rule of selecting next vertex to include it into the constructing route $M_{1k}(t)$. That is why vertex set $X_{1k}(t) \in X_1$ is formed where every vertex $x_{1i} \in X_{1k}(t)$ might be included into the constructing route $M_{1k}(t)$. $e_{k1}(t)$ is the last vertex in the route $M_{1k}(t)$. For each vertex $x_{1i} \in X_{1k}(t)$ parameter f_{1ik} is computed. f_{1ik} is the total quantity of pheromone laid in the graph's G^*_1 edge, connecting x_{1i} with the vertex $e_{k1}(t)$.

Probability P_{1ik} of including vertex $x_{1i} \in X_{1k}(t)$ into the constructed route $M_{1k}(t)$ is determined by the following ratio

$$\mathrm{P_{1ik}} = \mathrm{f_{1ik}} / \sum \mathrm{f_{1ik}} \quad (5)$$

With the probability P_{1ik} agent z_{k1} selects one of the vertexes that is included then into the route $M_{1k}(t)$. Similarly route $M_{2k}(t)$. is constructed by ant z_{k2}.

At the second step of iteration every ant z_{k1} lays pheromone on the edges of the constructed route $M_{1k}(t)$.

Pheromone quantity $\tau_k(l)$ laid by ants z_{k1} and z_{k2} on every edge of the routes $M_{1k}(t)$ and $M_{2k}(t)$ is defined as following:

$$\tau_k(l) = Q/F_k(l). \quad (6)$$

where l is the iteration count, Q is the total quantity of pheromone laid by ant on the edges of routes $M_{1k}(t)$ and $M_{2k}(t)$. $F_k(l)$ is an objective function for solution constructed by ant pair $\langle z_{k1}, z_{k2} \rangle$ at the iteration l. The smaller $F_k(l)$ the more pheromone is laid on the edges of constructed routes and therefore the probability of selecting these edges on the next iteration is higher.

At the third step pheromone evaporates from the graphs' G^*_1 and G^*_2 edges according to formula

$$\mathrm{f_{ik}} = \mathrm{f_{ik}}(1 - \rho), \quad (7)$$

where ρ is renewal coefficient.

After performing all the actions pair of agents with the best solution is found and solution is remembered. Further transition to the next iteration is carried out.

Time complexity of this algorithm depends on colony lifetime l (number of iterations), number of graph vertexes n and number of ants m. Time complexity is defined as $O(l*n^2*m)$.

4 Experimental Evaluation

The experiments of ant colony placement algorithm have shown that less than 150 generations are required to get optimal solution. For schemes consisting of 100 elements it will take approximately 160 s.

3 times to launch program with random initial populations is enough to get a solution with average deviation of 2 %.

GEN—genetic algorithm; AD—adaptive algorithm; HY—hybrid algorithm; MA; AT—ant tree algorithm; SPP—sequential pair placement (Table 1).

In order to compare the results held by placement experiments most wide-known genetic algorithms GASP, ESP and annealing modeling algorithm TimberWolf F 3.3. were used.

5 Conclusion

Developed nature-inspired allocating algorithm is based on concept of a pair of sequences. When implementing ant colony mechanisms a new approach is suggested. A pair of element sequences is constructed by a pair of ants on 2 solution search graphs. It allows to parallelize the process of constructing a pair of element sequences and to reduce the time of algorithm performing. Combining the composition of pairs and using hybridization with genetic algorithms is possible.

Experimental research were held on IBM PC. Compared to existing algorithms based on the concept of pair of element sequences some improvements are achieved.

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. Lebedev, B.K., Lebedev, O.B.: Locating Methods. TRTU publishing house, Monograph. Taganrog (2006)
2. Kureichik, V.M., Lebedev, B.K., Lebedev, O.B.: Solving Location Problem by Evolutional Modeling. Izvestiya of RAS. Theory and Control Systems. № 4. pp. 78–90 (2007)
3. Lebedev, B.K., Lebedev, V.B.: Coverage based on swarm algorithm .In: Proceedings of the 3rd All-Russian Scientific and Technical Conference. Neuro-informatics-2011, vol. 2, Fizmalit Press, pp. 93–103 (2011)
4. Lebedev, V.B.: Construction of the Shortest Linking Nets Based on Swarm Intellect. Izvestiya of SFU. TTI SFU publishing house. № 7. pp. 37–44 (2011)
5. Lebedev, B.K., Lebedev, VB.: Location Based on Bee Colony Algorithm. Izvestiya of SFU. TTI SFU publishing house, № 12. pp. 12–18 (2010)
6. Chang, C.C., Cong, J. и Xie, M.: Opimality and scalability study of existing placement algorithms. In: Proceedings of the Asia South Pacific Design Automation Conference. pp. 621–627 (2003)
7. Xu, J., Guo, P.-N., Cheng, C.-K., Cheng, C.-K.: Sequence-pair approach for Rectiliner module placement. IEEE Trans. Comput.-Aided Design Integr. Circuits Syst. **18**(4), 484–493 (1999)
8. Dorigo, M., Stützle, T.: Ant Colony Optimization. MIT Press, Cambridge (2004)
9. Lebedev, O.B.: Tracing in Channel Using Ant Colony Algorithm. Izvestiya of SFU. TTI SFU publishing house, № 4. pp. 46–52 (2009)

Multiagent Self-Organizing Interval Bacterial Colony Evolution Optimization Algorithm

Andrei Panteleev and Valentin Panovskiy

Abstract This work considers multiagent Self-Organizing Interval Bacterial Colony Evolution Algorithm—new global optimization technique which is a fusion of interval analysis and heuristic procedures and rules. This algorithm, imitates process of bacteria colony evolution and implements features of self-tuning and chaos control. The efficiency of this method is demonstrated on two applied problems: group interception (where a system of interceptors should be navigated to a system of targets) and satellite stabilization (rotational motion dampening).

Keywords Interval analysis · Nonlinear programming · Optimization · Optimal control · Self-organizing algorithm · Multiagent algorithm

1 Introduction

Existing numerical methods of optimal control include various methods which use Pontryagin's maximum principle and Bellman's equation, direct methods (e.g. gradient methods—first-ordered methods), second-ordered methods which base on Taylor approximation of Krotov-Bellman function, different improvement methods and etc. [1, 2]. There is a need in development of new approaches in order to improve the efficiency and accuracy. Possible approach consists in the use of interval analysis coupled with metaheuristic strategies [3, 4].

Interval analysis [5–9] is used as the main component of Self-Organizing Interval Bacterial Colony Evolution Algorithm (SOIBCEA). Existing interval optimization methods can be divided into two groups:

A. Panteleev (✉) · V. Panovskiy
Moscow Aviation Institute, Applied Mathematics and Physics,
Volokolamskoe Shosse 4, 125993 Moscow, Russia
e-mail: avpanteleev@inbox.ru
URL: http://www.mai.ru

V. Panovskiy
e-mail: panovskiy.v@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 450, DOI 10.1007/978-3-319-33609-1_41

- methods of unconstrained optimization (Moore-Skelboe algorithm, Ichida-Fujii algorithm, Dussel algorithm, interval simulated annealing, graph subdivision method and etc. [10–14]),
- methods of constrained optimization (Hansen method, Moore method and etc. [6, 8]).

SOIBCEA belongs to the group of so called metaheuristic algorithms. Metaheuristic algorithms give high-quality solutions in the most practically significant cases. Such algorithms do not guarantee determination of the best solution, but the main advantage is that they have lower computational complexity. Thus they can be applied to complicated problems without additional constraints on problem statement (differentiability, convexity and etc.). Also these methods do not use necessary and sufficient optimality conditions [15]. Metaheuristic algorithms combine one or several heuristic procedures which base on the higher level strategy.

It should be noted that all metaheuristic algorithms have a problem of parameter tuning. Results of such algorithms may vary significantly according to different values of parameters. In the majority of cases it is impossible to predict appropriate values. To get accurate result you need to restart algorithm with different parameter values and choose the best result. That is the point why development of self-organizing algorithms is very urgent [16, 17]. This feature allows you to get rid of mistakes which occur due to bad parameter choice.

In the following work it will be shown how general open-loop control problem can be transformed into the form of interval constrained global optimization problem which will be solved numerically by SOIBCEA. Method efficiency is demonstrated on the solution of applied aircraft control problems.

2 Optimal Control of Continuous Deterministic Systems

Behaviour of control object is described by differential equation [18]:

$$\dot{x}(t) = f\left(t, x(t), u(t)\right), \tag{1}$$

where $t \in T = \left[t_0; t_1\right]$ is continuous time, T is system operational time, $x \in \mathbb{R}^n$ is system state vector, $u \in [\mathbf{u}] \subset \mathbb{R}^q$—control vector, $[\mathbf{u}]$—set of admissible control values (box), $f(t, x, u) = \left(f_1(t, x, u), \ldots, f_n(t, x, u)\right)^T$—continuous vector-function. Hereafter, lowercase letters in square brackets ($[a] = \left[a_l; a_u\right]$) will be used to denote intervals and bold lowercase letters in square brackets ($[\mathbf{a}] = \left[a_1\right] \times \cdots \times \left[a_n\right]$)—for boxes.

Initial state is given

$$x\left(t_0\right) = x_0. \tag{2}$$

Terminal state $x\left(t_1\right)$ should satisfy the following interval terminal conditions:

$$\Gamma_i\left(x\left(t_1\right)\right) \in \left[G_i\right], i = 1, \ldots, l, \tag{3}$$

where $0 \le l \le n$, $\left[G_i\right]$—given intervals, functions $\Gamma_i(x)$ are continuously differentiable, system of vectors $\left\{\frac{\partial \Gamma_i(x)}{\partial x_1}, \ldots, \frac{\partial \Gamma_i(x)}{\partial x_n}\right\}, i = 1, \ldots, l$ is linearly independent $\forall x \in \mathbb{R}^n$.

The only information which is used to form control is continuous time t. Hence, we consider open-loop control.

Set of admissible processes $D\left(t_0, x_0\right)$ is defined as a set of pairs $d = (x(\cdot), u(\cdot))$ which include trajectory $x(\cdot)$ and admissible piecewise-continuous control $u(\cdot)$, where $\forall t \in T : u(t) \in [\mathbf{u}]$ that satisfy state equation (1), initial condition (2) and terminal condition (3).

On the set of admissible processes we consider cost functional

$$I(d) = \int_{t_0}^{t_1} f(t, x(t), u(t))\, dt + F\left(x\left(t_1\right)\right), \tag{4}$$

where $f^0(t, x, u)$ and $F(x)$ are given continuous functions.

In order to transform the given problem into the problem without terminal constraints (3) we can use penalty function:

$$\bar{I}(d) = I(d) + \sum_{i=1}^{l} R_i^{(1)} \cdot H_i^{(1)}, \tag{5}$$

where $H_i^{(1)} = h_\infty^2\left(\left[\Gamma_i\left(x\left(t_1\right)\right); \Gamma_i\left(x\left(t_1\right)\right)\right], \left[G_i\right]\right)$—penalty function, $h_\infty([a],[b]) = inf\{r | [a] \subset [b] + [-r; r]\}$—Hausdorff metric, $R_i^{(1)}$—penalty parameters.

If there are interval phase constraints $\left\{p_j(x(t), u(t), \dot{x}(t), \dot{u}(t)) \in \left[P_j\right], \forall t \in T\right\}_{j=1}^{N_p}$, where $p_j(\cdot)$—is constraint function, $\left[P_j\right]$—given intervals, N_p—is a number of constraints, then we have another form of modified cost functional (4):

$$\bar{I}(d) = I(d) + \sum_{i=1}^{l} R_i^{(1)} \cdot H_i^{(1)} + \sum_{j=1}^{N_p} R_j^{(2)} \cdot H_j^{(2)}, \tag{6}$$

where $H_j^{(2)} = \int_{t_0}^{t_1} h_\infty^2\left(\left[p_j(x(t), u(t), \dot{x}(t), \dot{u}(t)); p_j(x(t), u(t), \dot{x}(t), \dot{u}(t))\right], \left[P_j\right]\right) dt$, $R_j^{(2)}$—penalty parameters.

The task is to find such a pair $d^* = (x^*(\cdot), u^*(\cdot))$ that

$$d^* = arg \min_{d \in D\left(t_0, x_0\right)} I(d). \tag{7}$$

To solve problem (7) it is suggested to propose parametric form of control and convert initial problem to a problem of parametric optimization. The control which is obtained by such a procedure is suboptimal.

3 Application of Interval Optimization Technique

The main idea of interval analysis [5–9] is that real numbers are surrounded with intervals and real vectors—with interval vectors, or boxes.

Each interval has the following characteristics:

- lower bound: $\underline{[a]} = sup\,\{\xi \in \mathbb{R} \cup \{-\infty, \infty\}\,|\forall \zeta \in [a] : \xi \le \zeta\}$,
- upper bound: $\overline{[a]} = inf\,\{\xi \in \mathbb{R} \cup \{-\infty, \infty\}\,|\forall \zeta \in [a] : \zeta \le \xi\}$,
- width: $\omega([a]) = \overline{[a]} - \underline{[a]}$,
- midpoint of nonempty and finite interval: $mid\,([a]) = \frac{\underline{[a]}+\overline{[a]}}{2}$,

The same parameters are defined for boxes. Lower and upper bounds and midpoint become vectors, width is now calculated as maximum of components' widths.

Interval hull of a set $X \subset \mathbb{R}^n$ is the smallest box (box with minimum width) $[\mathbf{X}]$, that contains X. If the set is surrounded by square brackets it means that we consider the interval hull.

Let $\circ$ be a binary operation, then $[x] \circ [y] = \left[\left\{\xi_1 \circ \xi_2 | \xi_1 \in [x], \xi_2 \in [y]\right\}\right]$; let f be an unary operator, then $f([x]) = \left[\{f(\xi)\,|\,\xi \in [x]\}\right]$.

The set of intervals is designated by $\mathbb{IR}$, set of boxes—by $\mathbb{IR}^n$. Consider a function f from $\mathbb{R}^n$ to $\mathbb{R}$. The interval function $[f]$ from $\mathbb{IR}^n$ to $\mathbb{IR}$ is an inclusion function for f if $\forall\,[x] \in \mathbb{IR}^n, f([x]) \subset [f]([x])$. Inclusion function allows to get a priori estimate of function's range even if it is nonconvex or disconnected. If all variables are substituted with intervals and all operations are done by the rules of interval arithmetic such a value is called an estimation of a direct function image.

We will work with the following interval ε-minimization problem: consider given box $[\mathbf{s}]$, cost function $f : \mathbb{R}^n \to \mathbb{R}$. Interval ε-minimization problem consists in determination of such a box $[\mathbf{p}]^* \subset [\mathbf{s}]$ that

$$\omega\left([\mathbf{p}]^*\right) \le \varepsilon, \nexists\,[\mathbf{x}] \subseteq [\mathbf{s}], \omega([\mathbf{x}]) \ge \varepsilon, \overline{[f]([\mathbf{x}])} < \underline{[f]\left([\mathbf{p}]^*\right)}, \tag{8}$$

where the value ε affects the width of the $[\mathbf{p}]^*$ box.

We will consider interval control as piecewise-constant or piecewise-linear interval function $[u](t)$ and interval trajectory as piecewise-continuous interval function which satisfies (2)–(4).

Piecewise-constant control (Fig. 1a) can be uniquely associated with interval vector $\underbrace{[u_1(\tau_0)] \times \cdots \times [u_q(\tau_0)]}_{[u(\tau_0)]} \times \cdots \times \underbrace{[u_1(\tau_{N-1})] \times \cdots \times [u_q(\tau_{N-1})]}_{[u(\tau_{N-1})]}$, where $\tau_i =$

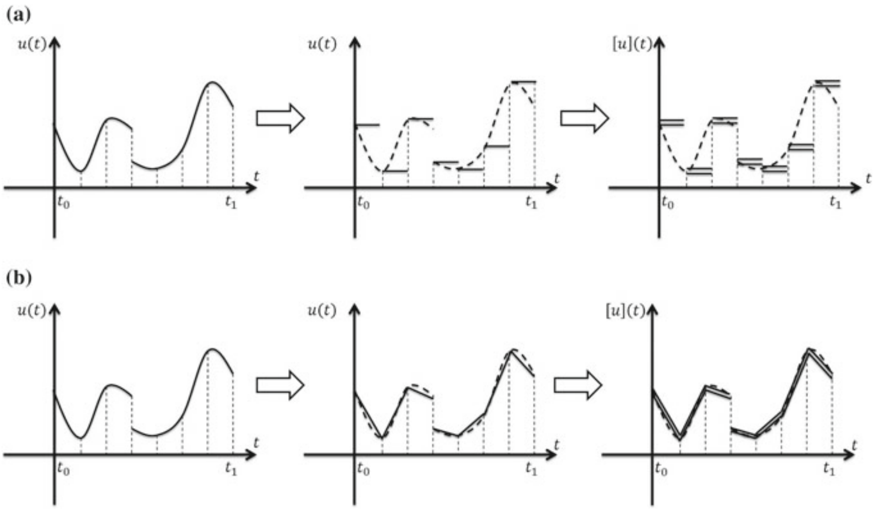


Fig. 1 Piecwise-continuous interval control

$t_0 + \frac{t_1-t_0}{N} \cdot i$, N is number of discretization steps. Corresponding interval control can be found as follows: $[u](t) \in [u(\tau_i)]$, $\tau_i \le t < \tau_{i+1}$, $i = 0, \ldots, N-1$.

Piecewise-linear control (Fig. 1b) can be uniquely associated with interval vector $\underbrace{[u_1(\tau_0)] \times \cdots \times [u_q(\tau_0)]}_{[u(\tau_0)]} \times \cdots \times \underbrace{[u_1(\tau_N)] \times \cdots \times [u_q(\tau_N)]}_{[u(\tau_N)]}$. Corresponding interval control can be found as follows: $[u](t) \in \frac{\tau_{i+1}-t}{\tau_{i+1}-\tau_i} \cdot [u(\tau_i)] + \frac{t-\tau_i}{\tau_{i+1}-\tau_i} \cdot [u(\tau_{i+1})]$, $\tau_i \le t < \tau_{i+1}$, $i = 0, \ldots, N-1$.

It is suggested to find the control $[u^*](t)$ by the SOIBCEA algorithm (optimization will be done on the search area $[\mathbf{s}] = \underbrace{[\mathbf{u}] \times \cdots \times [\mathbf{u}]}_{q \cdot N}$—for piecewise-constant control and $[\mathbf{s}] = \underbrace{[\mathbf{u}] \times \cdots \times [\mathbf{u}]}_{q \cdot (N+1)}$—for piecewise-linear control).

The value of the cost function criterion (6) is calculated by the numerical methods using Euler, Euler-Cauchy or Runge-Kutta formulas [4] (all calculations on the right hand side of equations are done by the rules of interval arithmetic [5–9]).

4 SOIBCEA Strategy

SOIBCEA is interval global optimization algorithm which imitates bacteria colony evolution process. Each bacteria (agent) is associated with a box that describes its habitat area. The main purpose is to assure that all bacterias are situated in the

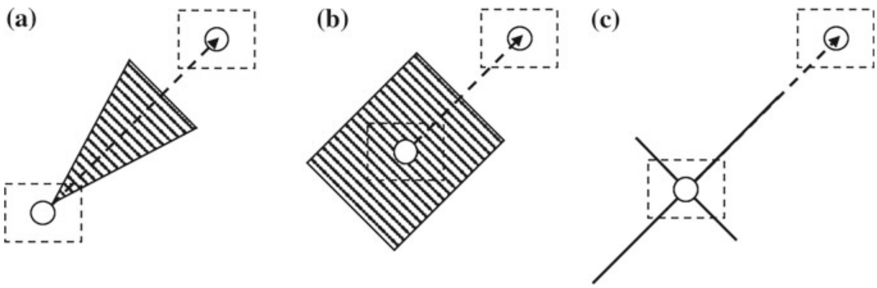


Fig. 2 Moving strategies: **a** pyramid, **b** zone, **c** star

neighbourhood of the possibly best location (areas with smaller lower bound of the direct function image estimation are treated as better ones).

Each bacteria has the following attributes:

- moving strategy,
- character,
- parameters which describe possible movement,
- current moving direction,
- habitat area (expressed as a box).

Current version of SOIBCEA supports three types of moving strategies (Fig. 2):

- pyramid,
- zone,
- star.

For described moving strategies bacteria can change its location to any in the shaded area (for pyramid and zone moving strategies) or new location can be generated on axes which are collinear to current direction vector (for star moving strategy).

Each bacteria has its own character from the list:

- greedy (bacteria moves to a first found area which is better than the current one),
- smart (bacteria chooses new habitat area from the list of areas),
- prudent (bacteria can move multiply times during one iteration).

Each iteration of the algorithm corresponds to the evolution cycle. During this cycle all bacterias find their best habitat area, produce bacterias of a new generation and die.

Feature of self-organization is implemented via tuning process. Possible movement of any bacteria is represented via numeric vector. Exact values of this vector coordinates are dynamically tuned according to current bacteria effectiveness.

Next generation is formed basing on the following heuristic rules:

- 30 % of new colony consists of random (newly generated) bacterias,
- 30 % of new colony consists of bacterias which are situated on previous generation best habitat areas,
- 30 % of new colony consists of bacterias which are situated in the same area as their predecessors but with opposite moving direction,
- 10 % of new colony consists of bacterias which are situated in the best (over all iterations) habitat area.

Implementation of these rules provides opportunities to keep both sufficient level of chaos and inheritance in a colony.

The box which is associated with the best found bacteria is chosen as a problem solution $[\mathbf{p}^*]$ and $[f]\left([\mathbf{p}^*]\right)$—as a cost function interval.

5 Applied Problems

Problem 1 (*Group Interception*).

Consider A targets (T) and A interceptors (I). Differential equations which describe such system has the following view [19]:

$$\begin{aligned} \dot{x}_T^i &= \left(V_T^i - \left(V_T^i - U_T^i\right) \cdot \left|u_T^i\right|\right) \cdot \cos\psi_T^i, \\ \dot{y}_T^i &= \left(V_T^i - \left(V_T^i - U_T^i\right) \cdot \left|u_T^i\right|\right) \cdot \sin\psi_T^i, \\ \dot{\psi}_T^i &= \alpha_T^i \cdot u_T^i, i = 1, \dots, A, \\ \dot{x}_I^j &= \left(V_I^j - \left(V_I^j - U_I^j\right) \cdot \left|u_I^j\right|\right) \cdot \cos\psi_I^j, \\ \dot{y}_I^j &= \left(V_I^j - \left(V_I^j - U_I^j\right) \cdot \left|u_I^j\right|\right) \cdot \sin\psi_I^j, \\ \dot{\psi}_I^j &= \alpha_I^j \cdot u_I^j, j = 1, \dots, A, \end{aligned} \tag{9}$$

where V_T^i, U_T^i, V_I^j, U_I^j, $i,j = 1, \dots, A$—maximum and minimum velocities of targets and interceptors accordingly, x_T^i, y_T^i, x_I^j, y_I^j, $i,j = 1, \dots, A$—coordinates of targets and interceptors accordingly, ψ_T^i, ψ_I^j, $i,j = 1, \dots, A$—relative bearings of targets and interceptors accordingly α_T^i, α_I^j, $i,j = 1, \dots, A$—maximum turn rates of targets and interceptors accordingly. Control vector $u = \left(u_I^1, \dots, u_I^A\right)^T$.

The purpose is to find such a control, that will minimize total loss:

$$I = \sum_{i=1}^{A} \min_{j=1,\dots,A} r_{i,j}\left(t_1\right), \tag{10}$$

Table 1 Targets and interceptors

Targets				
i	1	2	3	4
V_T^i	170	160	150	140
U_T^i	100	110	120	130
α_T^i	2°	3°	4°	5°
x_T^i	−2500	2500	2500	−2500
y_T^i	2500	2500	−2500	−2500
ψ_T^i	135°	45°	−45°	−135°
u_T^i	−0.1	−0.2	−0.3	−0.4
Interceptors				
j	1	2	3	4
V_I^j	250	230	200	230
U_I^j	150	140	110	100
α_I^j	7°	5°	3°	3°
x_I^j	0	0	0	0
y_I^j	0	0	0	0
ψ_I^j	180°	180°	45°	−45°

where $r_{i,j}() = \sqrt{\left(x_T^i(t_1) - x_I^j(t_1)\right)^2 + \left(y_T^i(t_1) - y_I^j(t_1)\right)^2}$—distance between i-th target and j-th interceptor.

Information about targets and interceptors is described in Table 1.

The solution which was obtained by SOIBCEA (piecewise-linear controls with $N = 25$ and corresponding trajectories of targets and interceptors) is illustrated on Fig. 3.

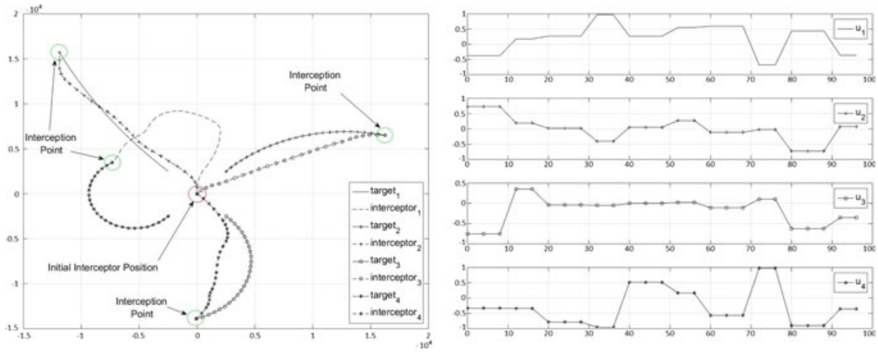


Fig. 3 Trajectories and controls found by SOIBCEA, $[I] = [0.9941; 0.9957]$

Problem 2 (*Satellite Stabilization*).

The problem of satellite rotational motion dampening is considered [20]. After the conversion to non-dimensional variables system of differential equations, which describes solid body motion relative to center of mass, has the following view:

$$\begin{cases} \dot{p} = \dfrac{u_1}{6}, \\ \dot{q} = u_2 - 0,2 \cdot p \cdot r, \\ \dot{r} = 0,2 \cdot \left(u_3 + p \cdot q\right). \end{cases} \tag{11}$$

Control vector constraints: $[\mathbf{u}] = [-200; 200] \times [-200; 200] \times [-200; 200]$.

Cost functional:

$$I_1(d) = \int_0^1 \left(\left|u_1(t)\right| + \left|u_2(t)\right| + \left|u_3(t)\right|\right) dt. \tag{12}$$

Initial state:

$$p(0) = 24, q(0) = 16, r(0) = 16. \tag{13}$$

Terminal constraints:

$$p(1) \in [0; 0], q(1) \in [0; 0], r(1) \in [0; 0]. \tag{14}$$

Results obtained by local variation method [20] are illustrated on Fig. 4.

Let's consider the following transformation of the cost functional:

$$I(d) = I_1 + 0,05 \cdot I_2 + 1000 \cdot I_3, \tag{15}$$

where $I_2 = \int_0^1 \left(u_1(t)^2 + u_2(t)^2 + u_3(t)^2\right) dt$—regularizing component smoothing, $I_3(d) = p^2(1) + q^2(1) + r^2(1)$—component for terminal constraint satisfaction.

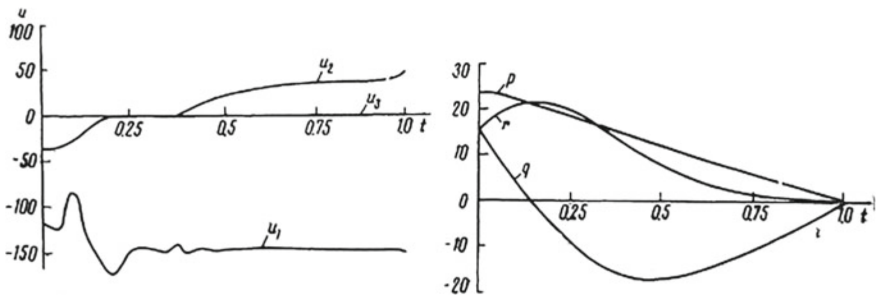


Fig. 4 Results from [20], $I = 169.42$

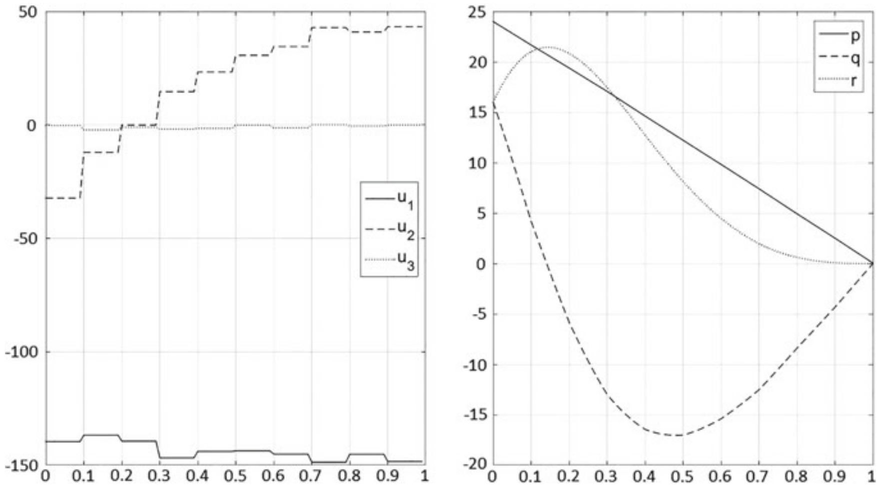


Fig. 5 Trajectories and controls found by SOIBCEA, components of cost functional: $[I_1] = [172.0956; 172.0973]$, $[I_2] = [21620.3657; 21620.8742]$, $[I_3] = [0.0020; 0.0024]$

Results obtained by SOIBCEA are illustrated on Fig. 5, piecewise-constant control, $N = 10$.

6 Conclusion

In the present work multiagent Self-Organizing Interval Bacterial Colony Evolution Algorithm (SOIBCEA) was considered. This newly developed interval optimization method was applied to optimal open-loop control problems: group interception and satellite stabilization. Obtained results prove the efficiency of SOIBCEA, all problems were solved with admissible accuracy.

References

1. Gurman, V.I., Rasina, I.V., Blinov, A.O.: Evolution and prospects of approximate methods of optimal control (Jevoljucija i perspektivy priblizhennyh metodov optimal'nogo upravlenija). J. Program Syst.: Theory Appl. **2**(6), 11–29 (2001)
2. Fedorenko, R.P.: Approximate Solution of Optimal Control Problems (Priblizhennoe reshenie zadach optimal'nogo upravlenija). Nauka, Moscow (1978)
3. Panteleev, A.V., Metlitskaya, D.V., Aleshina, E.A.: Methods of Global Optimization: Metaheuristic Strategies and Algorithms (Metody global'noj optimizacii. Metajevristicheskie strategii i algoritmy). Vuzovskaya Kniga, Moscow (2013)

4. Panteleev, A.V.: Application of Evolution Method of Global Optimization to Optimal Deterministic System Control Problems (Primenenie jevoljucionnyh metodov global'noj optimizacii v zadachah optimal'nogo upravlenija determinirovannymi sistemami). Izdatelstvo MAI, Moscow (2013)
5. Shary, S.P.: Finite-Dimensional Interval Analysis (Konechnomernyj interval'nyj analiz). XYZ, Novosibirsk (2010)
6. Hansen, E.: Global Optimization Using Interval Analysis. Marcel Dekker, New York (2004)
7. Jaulin, L., Kieffer, M., Didrit, O., Walter, E.: Applied Interval Analysis. Springer, London (2001)
8. Moore, R.E.: Interval Analysis. Prentice Hall, Englewood Cliffs (1966)
9. Moore, R.E.: Methods and Applications of Interval Analysis. SIAM, Philadelphia (1979)
10. Ratschek, H.: New Computer Methods for Global Optimization. Horwood, Chichester (1988)
11. Panovskiy, V.N.: Interval genetic algoritm of global constrained optimization (Interval'nyj geneticheskij algoritm global'noj uslovnoj optimizacii). J. Achievements Mod. Radioelectronics **4**, 71–75 (2014)
12. Panovskiy, V.N.: Application of Interval Explosion Search (Prikladnoe primenenie interval'nogo metoda vzryvov). J. MAI Proc. **73** (2014) http://www.mai.ru/science/trudy/published.php?ID=48451
13. Shary, S.P.: A surprising approach in interval global optimization. J. Reliable Comput. **7**, 497–505 (2001)
14. Shary, S.P.: Randomized algorithms in interval global optimization. J. Numer. Anal. Appl. **1**, 376–389 (2008)
15. Panteleev, A.V., Letova, T.A.: Methods of Optimization (Metody optimizacii). Logos, Moscow (2011)
16. Zelinka, I.: SOMA—Self-organizing migrating algorithm. J. Stud. Fuzziness Soft Comput. **141**, 167–217 (2004)
17. Nolle, L., Zelinka, I., Hopgood, A.A.: Comparison of an self-organizing migration algorithm with simulated annealing and differential evolution for automated waveform tuning. J. Adv. Eng. Softw. **10**, 645–653 (2005)
18. Panteleev, A.V., Bortakovskiy, A.S.: Control Theory in Examples and Problems (Teorija upravlenija v primerah i zadachah). Viskaya Shkola, Moscow (2003)
19. Shapira, I., Ben-Asher, J.Z.: Near-optimal horizontal trajectories for autonomous air vehicles. J. Guidance Control Dyn. **20**(4), 735–741 (1997)
20. Krylov, I.A.: Numerical solution of optimal satellite stabilization problem (Chislennoe reshenie zadachi ob optimal'noj stabilizacii sputnika). J. Comput. Math. Math. Phys. **8**(1), 203–208 (1968)

Effectiveness Evaluation of Memetics and Biogeography Algorithms Using Benchmark and Trans Computational Tasks of Combinatorial Optimization

S.I. Rodzin and O.N. Rodzina

Abstract The article discusses the elements of the theory of population metaheuristics. Original biogeographical and memetic algorithms for solving trans computational optimization problem are presented for the traveling salesman problem. Authors presented the method of biogeography and its modifications, as well as results of the comparative analysis of genetic, biogeographic and memetic algorithms. The authors conducted experimental verification of the effectiveness of the algorithms on known test functions. Experiments were carried out on certain benchmarks from the library TSPLIB. Efficiency, operating time and the diversity of the population were the criteria for comparison algorithms mentioned above.

Keywords Metaheuristics · Population · Memetics · Biogeography · Optimization · The traveling salesman problem · Benchmarking

1 Introduction

The task of searching for optimal solutions under various constraints are countless. Each process in science and technology, economics and business, has the potential for optimization and can be formulated as an optimization problem. Optimization is, in one case, minimization of time, cost, risk, or, in other case, maximization of profits, quality, and efficiency. Most real-world optimization problems are complex and difficult to decide accurately within a reasonable time. The main alternative for these

S.I. Rodzin (✉) · O.N. Rodzina
Southern Federal University, Rostov-on-Don, Russia
e-mail: srodzin@sfedu.ru; srodzin@yandex.ru

O.N. Rodzina
e-mail: orodzina@sfedu.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 450, DOI 10.1007/978-3-319-33609-1_42

tasks are approximate algorithms. The most common class of approximation algorithms are metaheuristic algorithms and, at the same time, are applicable to a variety of optimization problems. It can be adapted to solve any problem of optimization.

The term "metaheuristic algorithm" refers to the search algorithm designed to automate the process of selecting, combining, generalizing, or adaptation of several more simple heuristics for effective solution of the computing problem. The main advantage of heuristic algorithms consists in building effective procedures to solve a class of problems (not a single specific task). Metaheuristic algorithm has to be fast and easy to implement, but, at the same time, it should be a sufficiently reliable solution for a class of problems. Metaheuristics organize the process of finding suboptimal solutions and sometimes include quite complex training procedures. Metaheuristics are approximate, non-deterministic procedures using domain-specific knowledge.

Metaheuristic algorithms are one of the areas of optimization techniques in computer science and applied mathematics. These algorithms are booming the past 30 years in such areas as artificial intelligence, soft computing, mathematical programming, operations research, and decision-making [1]. It can be attributed to bioinspired metaheuristics as the follows: genetic algorithms [2], algorithms ant colony [3], swarm of bees [4], flock of birds [5], school of fish [6], swarm of salmonella bacterium [7], behavior of cuckoos during the forced nest parasitism [8], glowworm swarm [9], weed farmland colonization [10], behavior of frog groups during the food searching process [11], swarm of flies [12], flock of bats [13], memetic algorithms [14], biogeography-based algorithms [15], etc.

The article has two purposes. First is to provide the reader with computational models two versions population metaheuristic algorithms: memetic and biogeographic trans computational solutions for the traveling salesman problem. Second is to compare the proposed algorithms on benchmarks with known solutions obtained by genetic algorithm.

2 Statement of the Traveling Salesman Problem

The traveling salesman problem consists in finding the most profitable routes through the city indicated at least once, then return to the starting city (Hamiltonian cycle). In the context of the task, we specify the criterion profitability of the route (the shortest, cheapest, set of criteria, etc.) and the corresponding distance matrix, cost, etc. There are several varieties of setting the traveling salesman problem: geometric problem, it is when the distance matrix reflects the distance between the points on the plane for the metric traveling salesman problem, when the matrix of values the inequality triangle symmetric and asymmetric traveling salesman problem.

Along with the ease of determining the relative simplicity of finding and making good the traveling salesman problem is characterized in that the determination of the optimal path is really trans computational problematic: even at a of cities n (n = 66 and more) it cannot be solved by brute force.

Traveling salesman problem can be represented as a model on a graph $G = (X, V)$, consisting of vertices-cities (X) and edges of paths (V) and between them.

If $i, j \in X$—vertex, then the ribs $\{i, j\} \in V$ you can compare the profitability of the route criterion, for example, the distance between the cities, the time or cost of the trip. The decision of the traveling salesman problem is to find a Hamiltonian cycle of minimum weight in its weighted graph.

Asymmetric Traveling Salesman Problem is modeled directed graph. In a directed graph the weight c_{ij} of a rib $\{i, j\}$ can differ by weight c_{ji} of a rib $\{j, i\}$. In symmetric problem all pairs of edges between the same nodes are of equal length ($c_{ij} = c_{ji}$), the number of possible routes is half-symmetric case. Symmetric problem modeled by the undirected graph. Symmetric traveling salesman problem will call metric if the relative lengths of the edges of the triangle inequality holds: $c_{ij} \leq c_{ik} + c_{kj}$. In such problems, workarounds of long straight: from the top edge i to the top j never longer than the path through the intermediate top k.

The traveling salesman problem is a problem of combinatorial optimization. Its solutions are represented as variables. For example, a symmetric problem can be represented as a plurality of ribs V. Each edge $\{i, j\}$ is compared with binary variable $x_{ij} \in \{0, 1\}$, that is equal 1, if the edge belongs to the route, and 0—in the opposite case. Custom route can be represented as a set of variables values toiletries, but each set determines the route of a salesman.

The best solution is to correct the traveling salesman problem that satisfies all the constraints of the problem the route, the length of which is equal to

$$min\{\sum_{i \in X}\sum_{j \in X \setminus \{i\}} c_{ij}x_{ij} | \in x_{ij}\{0, 1\}. \quad (1)$$

Since the salesman selects in each city the next of the city that he had not yet visited, there are $(n - 1)!$ routes and asymmetric $(n - 1)!/2$ routes for symmetric traveling salesman problem. Thus, the size of the search space is exponentially dependent on the number of cities. Different versions of the traveling salesman problem (metric, symmetric and asymmetric) are *NP*-equivalent.

There are algorithms for finding approximate solutions to a metric problem in polynomial time and route finding twice as long as the maximum optimum. So far, no known polynomial-time algorithm that guarantees the accuracy of at least 1.5 times worse than the optimal length of the route.

All the efficiency and at the same time reduces the exhaustive search methods for solving the traveling salesman problem is heuristic and metaheuristic. Heuristic methods find not the best optimal route, but the approximate solution.

3 Theory of Population Metaheuristics

The theory of population-based metaheuristics of trans computational tasks for combinatorial optimization problems based on the following hypotheses and laws.

Population-based metaheuristic algorithms (*PMA*) represent a consistent transformation of the set of solutions using the information stored in the search. *PMA* applied to determine problems where the phase space variables is not obligatorily a metric. *PMA* imitates the process of discovering the optimal solutions by evolutionary operators of crossover, mutation, inversion, translocation, selection, transposition, segregation, etc., and support populations of structures that is developed in the environment. Selection concentrates on the selection of individuals with upper values of the objective function. As well, reproduction, crossover, mutation and other operators generate new individuals.

Basics of the population-based metaheuristics theory for optimal solutions represent the following system. Set of objects *X* (population-based solutions) is fixed. Among the capability, you need to select the best for the optimality criterion *F*. The optimal criterion is based on the properties of objects is a mapping:

$$F\colon X \rightarrow R, \tag{2}$$

where $x \in X$ from *X* assigns the value $F(x)$.

Phenotypic nature of a set of objects is random, so, we build a coded representation of a set of objects in the final vector set *S* (genotype). Mapping of the form *φ: X S* describes the relationship between the subject, which are solutions of the search. There is some inverse mapping

$$\varphi^{-1}\colon S \rightarrow X, \tag{3}$$

where each element of the population $s \in S$ meet to an element of the set *X*. Then the optimization process using *PMA* algorithm consists in construction of a set of objects-solutions $X_{opt} \in X$, for following conditions are satisfied:

$$X_{opt} = argmax\, F\left[\varphi^{-1}(s)\right], s \in S. \tag{4}$$

In the process of optimizing the set *X* progress to an optimal state changing its composition and parameters of its constituent objects. The method of constructing is a set of objects $s \in S$ determined by the *PMA* algorithm. Evolution of *X* is determined by the evolution of the population *S*. On the set *S* is defined subset P_0—random initial population. The decision on each step of evolution of the difference computing circuit is determined

$$P_{t+1} = A(P_t), \tag{5}$$

where A is the composition different evolutionary operators.

Structure of the population metaheuristic search is adaptive using a composition of various evolutionary operators A in the subtraction computational scheme $P_{t+1} = A(P_t)$. The only way to reduce the time and improve the quality of population-based heuristic search for optimal solutions is a specialty search pattern. We show this below by the example of the traveling salesman problem, *PMA* memes algorithms and bioinformatics.

4 Memetic Population-Based Algorithm

Memetic population algorithm (*MPA*) uses the concept of memes, and the principle of evolution [14, 16]. A meme is an information measure that is able to reproduce itself by means of people. People store and transmit this information to others through communication [17]. In *MPA*, meme is realization of a local optimization algorithm that clarifies current solution on each iterations [1].

MPA has the following algorithm:

Step 1. Creating the initial population of solutions $P_0 = \{p_1^0, \ldots, p_n^0\}$. Forming a plurality of memes. As memes use a variety of algorithms $M = \{M_1, \ldots, M_k\}$, for example, Threshold Accepting, Record-to-Record Travel, Great Deluge Algorithm, Demon Algorithms, Guided Local Search, Smoothing Methods, Noisy Method, GRASP, etc. [1, 18].

Step 2. For each solution, we find the fitness function F_i.

Step 3. Collaboration by exchanging information between solutions (similar to crossing-over in genetic algorithms), usage of evolutionary operators.

Step 4. Local search. The best meme M_i is selected from swarm of memes based on h launches of local search by calculating the values of the objective function optimization.

Step 5. Creation of a new population through the competition/selection.

Step 6. Examination nearest results (number of iterations, improving the outcome evaluation), go to step 2, otherwise, the choice of the best solutions from the population.

5 Biogeographic Population-Based Algorithm

Biogeographical population algorithm (*BPA*) uses a model of island biogeography [19]. The main factors that affect the species richness of the island are migration and mutation. The balance between emigration, immigration, and extinction determines the number of inhabiting island's species. The quality of the island in terms of its suitability for the habitat a species determines the suitability index (*HSI*) [15]. Variables that affect the suitability index are the area of the island, rainfall, topographic diversity, temperature, variety of vegetation, etc. Islands with a high *HSI* are diverse and have a large number of species of wildlife. They have high levels of emigration and low immigration in contrast to the "poor" islands with low *HSI*. However, the island with low fitness are more dynamic.

Model of species distribution on the island is characterized by $\lambda(n)$—immigration function, $\mu(n)$—emigration function, n—the number of species of the island. With $n = 0$ emigration is zero, but with an increase in the number of species island overflows and a greater number of species have an incentive to leave the island and explore the other islands. Emigration reached its maximum value when the number of species that equal $n_{\max}$, Maximum immigration takes place with a minimum number of species on the island. Immigration becomes zero with $n_{\max}$.

A number of species n_0 achieves the equality of levels of immigration and emigration. The metastable number that can randomly vary depending on the climate changes of the island, epidemics, etc. [15] offers a simplified theoretical model explaining the processes of immigration/migration on the island. In this model of function $\lambda(n)$ and $\mu(n)$ are linear functions.

If we denote p_n(t) the probability that the island contains reporting n species at time t, then, according to [15], we can estimate the probability of a change p_n in time [t, t + Δt]:

$$p_n(t+\Delta t) \cong p_n(t) + \frac{dp_n(t)}{dt}\Delta t \tag{6}$$

Formula (6) a main settlement ratio is used in *BPA* algorithm.

We interpret the traveling salesman problem in terms of the notation and *BPA* algorithm. Each *i*–island (i = 1, 2, …, k) characterized by the vector of the integer parameters $X_{\rm i}$ = $SIV_{\rm i}$, each of which takes a value in the interval [$a_{\rm i}$, $b_{\rm i}$]. Then the index fitness of the island

$$HSI(X_{\rm i}) = F(X_{\rm i}) = F(SIV_1, SIV_2, \ldots, SIV_{\rm d}). \tag{7}$$

In [15] we can see a proposed canonical biogeographical algorithm. However, studies [20] have shown that the canonical version of the biogeographic algorithm is not without some disadvantages: often generates solutions that are far from optimal; there is no mechanism for the selection of the best decisions; generated a significant number of invalid decisions.

We offer a *BPA*, in other words, a modification of the canonical biogeographic algorithm based on the use of symmetrical paths in the traveling salesman problem.

To reduce the computation time we can compare a decision and symmetrical him to opt for a solution with the best objective function value. This is a definite potential for accelerating the convergence of optimum search procedure. The same approach we can apply not only to the initial population of solutions, but also to each solution in the current population.

Here the concept of symmetrical path according to [21].

Definition Suppose P = (x_1, x_2,…, x_d) is a some current optimum point in d-dimensional space, where $x_1, x_2, \ldots, x_d \in R$ (R—set of real numbers) and $x_i \in [a_i, b_i]$, $i \in \{1, 2, \ldots, d\}$. The point $\hat{P} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_d)$ called symmetric relative to the point of this optimum, if its coordinates are defined as

$$\hat{x}_i = 2x_{co} - x_i, \tag{8}$$

where x_{co} is the optimal solution in the current population.

From (8) it follows that as the point of symmetry using not the middle interval [*a, b*], and the coordinates of the optimal solutions in the current population of solutions. This increases the chances of finding the global optimum, especially in the later stages of the method, because the points are symmetric near the optimum current.

Modified *BPA* algorithm includes the following steps.

Step 1. We set the parameters required for the operation of the algorithm: maximum number of islands k, the maximum number of species on the island n_{max}, the maximum intensity of immigration and emigration $\lambda_{\max}$, $\mu_{\max}$; the maximum level of mutation (random imbalance species on the island) $m_{\max}$; the maximum number of $\rho_{\max}$ elite island ecosystem value $\lambda = 0$.

Step 2. We randomly initialize the initial population R. Initialize ecosystem, accidentally ran over an initial population P in k islands (potential solutions to the traveling salesman problem) and the initial value of the index fitness islands.

Step 3. We form a symmetrical population $\breve{P}$.

Step 4. We keep in the initial population the most suitable solution from the set $\{P \cup \breve{P}\}$.

Step 5. Perform a predetermined number of iterations.

Step 5.1. We produce migration and mutation solutions according to the canonical *BBO* algorithm [15]. Remove from the current population duplicate solutions. Calculate the current population of the index of fitness solutions P.

Step 5.2. We form a symmetrical population $\breve{P}$. Compute the suitability index of the solutions and leave the current population of the most suitable solution from the set $\{P \cup \breve{P}\}$.

Step 5.3. Add the elite solutions.

Step 6. If the stop condition is not satisfied, repeat step 5, otherwise, the output of the best solutions and stop of the algorithm.

6 Experimental Results

First, we have checked the effectiveness of the *MPA* and *BPA* algorithms on test functions. The question of defining an adequate plurality of test problems for the presented algorithms is not easy; although principles seem obvious:

- tasks in the test set should be similar to each other;
- situations that usually cause difficulties should be contained in test tasks;
- test tasks should be non-linear, separable and scalable.

These principles are performed on the test set, which includes the following functions:

- $F_1(x, y) = 0,1x^2 + 0,1y^2 - 4\cos(0,8x) - 4\cos(0,8y) + 8$, $x, y \in [-16; 16]$, $min = F_1(0,0) = 0$, Rastrigin function;

- $F_2(x,y) = 100(y - x^2)^2 + (1 - x)^2$,
  $x, y \in [-2, 2], min = F_2(1,1) = 0$, Rosenbrock function;
- $F_3(x,y) = -\frac{10}{0,005(x^2+y^2) - \cos(x)\cos\left(\frac{y}{\sqrt{2}}\right) + 2} + 10$,
  $x, y \in [-16, 16], min = F_3(0,0) = 0$, Griewank function;
- $F_4(x,y) = -\frac{100}{100(x^2 - y) + (1 - x)^2 + 1} + 100$,
  $x, y \in [-5; 5], min = F_4(1,1)$, De Jong function;
- $F_5(x,y) = (1 - \frac{sin^2\left(\sqrt{x^2+y^2}\right)}{1 + 0,001(x^2+y^2)}$,
  $x, y \in [-10; 10], min = F_5(0,0) = 0$, Sombrero function;
- $F_6(x,y) = 0,5(x^2+y^2)\lfloor 2A + Acos(1,5x)\cos(3,14y) + Acos(\sqrt{5}x)\cos(3,5y)\rfloor$,
  $A = 0,8; x, y \in [-2,5; 2,5], min = F_6(0.0) = 0$, Katkovnik function;
- $\frac{1}{F_7(x,y)} = \frac{1}{K} + \sum_{j=1}^{25} 1/(c_j + \sum_{i=1}^{2}(x_i - a_{ij})^6, \begin{cases} a_{1,5(i-1)+1} = a_{2,i} = -32 \\ a_{1,5(i-1)+2} = a_{2,i+5} = -16 \\ a_{1,5(i-1)+3} = a_{2,i+10} = 0 \\ a_{1,5(i-1)+4} = a_{2,i+15} = 16 \\ a_{1,5(i-1)+5} = a_{2,i+20} = 32, \end{cases}$

  $K = 500; c_j = j; x, y \in [-65, 65], min = F_7(-32, 32) \approx 1$, Shekel function.

Numerical coefficients in the Rastrigin function selected in the way that its value at the global optimum slightly different from the values at the points of neighboring local optima, because these optima are located very “deep”. Rosenbrock function has the large ravine in the optimum (Rosenbrock’s “banana”). Griewank function is multiextremal with many local optima. De Jong function was developed for testing in such a way that it has weight on almost entire domain. At the same, time the distance between local and global optima are very small. Sombrero function has a global optimum in the coordinate origin and value of the function in the points equidistant from the global optimum are equal. Katkovnik function is multiextremal and has very complicated system of “deep” local optima. The surface of Shekel function resembles foxholes. This function is multiextremal and has many local optima and small between local and global optima. These distances are slightly different, and in the remaining part of the domain of the function has an equal value.

The effectiveness of the developed memetic and biogeographic algorithms are experimentally tested on test functions. The algorithms compared between themselves and with the genetic algorithm. The genetic algorithm is applied evenly mating, breeding and rank high mutation. The size of the population—50, the number of launches each algorithm—50.

To assess the effectiveness of the algorithms the authors have used two measures: the reliability and the average number of iterations. Reliability is the percentage of successful launches (global optimum is found) to the total number of launches of the algorithm. The average number of iterations is the number of iterations averaged over the successful launch in case when the global optimum is found. The algorithm with the highest reliability is considered as the best. Among

Table 1 Experimental results

Test function	GA		MPA		BPA	
	Reliability (%)	Average number of iterations	Reliability (%)	Average number of iterations	Reliability (%)	Average number of iterations
F_1	20	16	29	18	**33**	**15**
F_2	42	18	48	21	**51**	**12**
F_3	24	20	32	24	**35**	**19**
F_4	6	**20**	12	29	13	21
F_5	53	20	62	20	**69**	**18**
F_6	36	11	44	**10**	46	13
F_7	100	8	**98**	9	92	**7**

the algorithms with the same reliability, we call the best one with the lowest average number of iterations.

Experimental test results are shown in Table 1 (the algorithm who is the "winner" is bolded on the test function). The authors emphasize that these results have been obtained within the test only for a set of test functions F_1–F_7.

The proposed algorithms *MPA* and *BPA* had been compared with well-known solutions that we obtain by genetic algorithm (*GA*) [22]. Authors had been performed comparative assessment of *GA*, *MPA* and *BPA,* along with experimental research of four well-known benchmark from the library TSPLIB for the traveling salesman problem [23]: *kroA*100, *ch*130, *kroA*200 and *lin*318. CBBO, end OCBBO: *kroC*100, *ch*130, *kroA*200 and *lin*318. For a correct comparison following general settings were used: population size—100; maximum of iterations—500.

The criteria for comparison of the test algorithm is efficiency (the length path of a TSP), operation time (CPU time per second), and diversity of the population.

Direct communication between these criteria is absent, but it allows you to make the comparison more objective. It is known as the calculation time depends on many factors (used data structures, programming style, and performance of the computer). However, it reflects the rate of convergence of the test algorithm.

For population algorithms diversity of the population it is an important indicator, preventing premature convergence. The entropy measure for estimating population diversity proposed in [18] was used here.

Experiments on benchmarks *kroA*100, *ch*130, *kroA*200 and *lin*318 have showed the following: efficiency of *BPA* on average for all benchmarks is higher than *GA* and *MPA* on 29.3 and 37.75 % accordingly.

Comparison between *GA*, *MPA* and *BPA* based on time-parameter have showed the following: work-time of *GA* and *MPA* grows linearly with the dimension of the problem. The work-time value of *BPA* is worse than *GA* and *MPA*. The work-time value of *BPA* increased approximately from 57 s for 100 cities up to 300 s for 318 cities. Therefore, there is no confidence in the linear nature of time depending on the method of operation of the number of cities that must visit a traveling salesman.

The main reason here is a proposed procedure for the formation of symmetric ways, i.e. its implementation requires additional resources.

Finally, comparison between *GA, MPA* and *BPA* by the criterion of making the diversity of the population leads to the conclusion in favor of the *GA* that has the lowest level of entropy, which indicates that the diversity of populations among all the tested benchmark. We can say it represents a little unexpected result, because usually we believe that a variety of populations leads to improved efficiency solutions trans computational tasks. However, this is not always the case.

7 Conclusion and Future Work

The authors consider the novel ideas proposed in the article are elements of the general theory of population metaheuristics and modified algorithms memetics and biogeography. Metaheuristics memes and biogeography are promising tools for solving combinatorial trans computational optimization problems. The experiments for the traveling salesman problem benchmark confirm this conclusion. The results of the comparison test algorithms biogeography and memetics according to the criteria of efficiency, working time and diversity of populations showed that the proposed algorithms *MPA* and *BPA* algorithms are superior genetic algorithm performance but inferior to him in working hours.

Development and research of bio-geographical and memetic algorithms for solving other problems trans computational for analysis and processing of big data in computer science, biology, medicine, management, as well as for many engineering applications are important and perspective tasks.

Acknowledgment The study was performed by the grant from the Russian Science Foundation (project # 14-11-00242) in the Southern Federal University.

References

1. Kureichik, V.V., Kureichik, V.M., Rodzin, S.I.: Theory of evolutionary computation. Moscow, Fizmatlit (2012)
2. Holland, J.H.: Adaptation in Natural & Artificial Systems. University of Michigan Press (1975)
3. Dorigo, M., Maniezzo. V., Colorni. A.: The ant system: optimization by a colony of cooperating objects. IEEE Trans. Syst. Man Cybern. Part B, **26**(1), 29–41 (1996)
4. Pham, D.T., Ghanbarzadeh, A., Koc, E., Otri, S., Rahim, S., and Zaidi, M.: The Bees Algorithm. Technical Notes. Manufacturing Engineering Centre, Cardiff University, UK (2005)
5. Reynolds, C.: Flocks, herds, and schools: a distributed behavioral model. Comput. Graph. **4**(21), 25–34 (1987)

6. Bastos-Filho, C.J.A., Lima-Neto, F.B., Lins, A., Nascimento, A., Lima, M.: Fish School Search. Nature-inspired Algorithms for Optimization (NISCO'2010), vol. 193, pp. 261–277. Springer, Heidelberg (2009)
7. Das, S., Biswas, A., Dasgupta, S., Abraham, A.: Bacterial Foraging Optimization Algorithm: Theoretical Foundations, Analysis, and Applications. Foundations of Computational Intelligence, vol. 203, pp. 23–55. Springer (2009)
8. Yang, X.-S., Deb, S.: Cuckoo search via l'evy flights. In: Proceedings of the World Congress NaBIC'2009, India. IEEE Publication, USA, 210–214 (2009)
9. Yang, Xin-She: Firelly algorithm, stochastic test functions and design optimization. Int. J. Bioinspired Comput. **2**(2), 78–84 (2010)
10. Mehrabiana, A.R., Lucase, C.: A novel numerical optimization algorithm inspired from weed colonization. Ecol. Inform. **1**, 355–366 (2006)
11. Mucherino, A., Seref, O.: Monkey search: a novel meta-heuristic search for global optimization. In: Proceedings of AIP Conference Data Mining, System Analysis and Optimization in Biomedicine, pp. 162–173 (2007)
12. Bova, V.V., Legebokov, A.A., Gladkov, L.A.: Problem-oriented algorithms of solutions search based on the methods of swarm intelligence. J. World Appl. Sci. J. **27**, 1201–1205 (2013)
13. Yang, X.-S.: A new metaheuristic sat-inspired algorithm. Nature Inspired Cooperative Strategies for Optimization, vol. 284, pp. 65–74. Springer, Berlin (2010)
14. Moscato, P.: Memetic algorithms. Handbook of Applied Optimization. Oxford University Press, Oxford (2002)
15. Simon, D.: Biogeography-based optimization. IEEE Trans. Evol. Comput. **12**(6), 702–713 (2008)
16. Rodzin, S., Rodzina, L.: Search for optimal solutions and its application for the processing of problem-oriented knowledge. In: Proceedings of IEEE Conference AICT'14, Astana, Kazakhstan, pp. 142–146 (2014)
17. Rodzina, L., Kristofferson, S.: Context-dependent car Navigation as kind of human-machine collaborative interaction. In: Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS'2013, May 20–24, 2013, San Diego, California, USA.). Publications of the IEEE, pp. 253–259 (2013)
18. Maekawa, K., et. al.: A genetic solution for the TSP by means of a selection rule. In: Proceedings of IEEE International Conference on Evolutionary Computation. Nagoya, Japan, pp. 529–534 (1996)
19. MacArthur, R.H.: The Theory of Island Biogeography. Princeton University Press (1967)
20. Ma, H., Simon, D.: Blended biogeography-based optimization for constrained optimization. Eng. Appl. Art. Intell. **24**(6), 517–525 (2010)
21. Rodzin, S., Rodzina, O.: New computational models for big data and optimization. In: Proceedings of IEEE Conference Application of Information and Communication Technologies (AICT'15), Rostov-on-Don, Russia, pp. 3–7 (2015)
22. Kravchenko, Y.A., Kureichik, V.V.: Bioinspired algorithm applied to solve the travelling salesman problem. World Appl. Sci. J. **22**(12), 1789–1797 (2013)
23. Reinelt, G.: TSPLIB—a TSP problem library. J. Comput. **3**, 376–384 (1991)

Performance Investigation of Mind Evolutionary Computation Algorithm and Some of Its Modifications

Maxim Sakharov and Anatoly Karpenko

Abstract Three new modifications of Mind Evolutionary Computation (*MEC*) algorithm were proposed in this paper. They are based on the concepts of co-evolution and memetic algorithms; modular software implementation of the specified methods was also presented. Paper contains results of performance investigation of the algorithms and their software implementation that was carried out using 8D benchmark functions of various classes. The influence of the free parameters' values on the performance of proposed algorithm was also studied; recommendations on the selection of those parameters' values were given based on the obtained results.

Keywords Global optimization · Mind evolutionary computation · Memetic algorithms · Co-evolution algorithms

1 Introduction

During last decades the population algorithms gained a reputation of powerful optimization methods. Their popularity is caused by simplicity of implementation and flexibility of application as they are based on the universal idea of evolution. In the meantime, the main disadvantage of such algorithms is their slow convergence to the neighborhood of global optimum because these methods don't utilize any local information about the objective function's landscape [1, 2]. This fact often restricts their usage in real world problems, where computation time is a crucial factor.

M. Sakharov (✉) · A. Karpenko
Bauman MSTU, Moscow, Russia
e-mail: max.sfn90@gmail.com

A. Karpenko
e-mail: apkarpenko@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 450, DOI 10.1007/978-3-319-33609-1_43

Nowadays, researchers in the field of population optimization methods are trying to improve their performance by means of either hybridization or meta-optimization. A development of hybrid algorithms implies a combination of various or same methods with different values of free parameters in such a way that advantages of one method would overcome disadvantages of another one. Meta-optimization, on the other hand, implies the adjustment of free parameters' values that would provide the maximum efficiency of an algorithm being investigated [2, 3].

The subject of this work is the Mind Evolutionary Computation algorithm (*MEC*), used by authors as a base algorithm for developing multi-memetic algorithms [2, 4]. A concept of the *MEC* algorithm was firstly proposed in 1998 [5]. The algorithm simulates some aspects of human behavior in the society; every individual is an intellectual agent which operates within a group of other individuals. In order to achieve a high position within its group, an individual has to study from the most successful individuals in this group. And groups themselves should follow the same principle to stay alive in the intergroup competition.

It was demonstrated in the work [6] that traditional *MEC* algorithm is bad at optimizing high-dimensional functions, in other words the algorithm isn't suitable for real world problems. In this regard, authors attempted to improve the efficiency of the algorithm without increasing its computational complexity.

It was proposed in this work to utilize a concept of co-evolution, based on the simultaneous evolution of several populations within one search domain, which solve one problem with different values of free parameters and compete for certain common resource. Authors proposed a co-modification of the *MEC* algorithm where a total number of individuals is used as resource for competition. As evolution proceeds more successful subpopulations get more agents for search domain exploration, taking them from less successful subpopulations. In such a manner, the total number of agents remains the same.

Mind Evolutionary Computation co-algorithm (*CoMEC*) and its memetic modification were proposed in this work along with the simple memetic modification of traditional *MEC*. Modular software implementation of those algorithms was also presented. Performance investigation of algorithms and their software implementation was carried out using multidimensional benchmark functions and with various values of free parameters.

2 Problem Statement and Simple *MEC*

In this paper a deterministic global unconstrained minimization problem is considered

$$\min_{X \in R^{|X|}} \Phi(X) = \Phi(X^*) = \Phi^*. \tag{1}$$

Here $\Phi(X)$—the scalar objective function, $\Phi(X^*) = \Phi^*$—its required minimal value, $X = (x_1, x_2, \ldots, x_{|X|}) - |X|$-dimensional vector of variables, $R^{|X|} - |X|$-dimensional arithmetical space.

A domain D is used for generating the initial values of vector X and determined with inequality constraints

$$D = \{X | x_i^{min} \leq x_i \leq x_i^{max}, i \in [1 : |X|]\} \subset R^{|X|}. \tag{2}$$

The *MEC* algorithm can be easily considered as a population one with a single multi-population made of leading groups $S^b = (S_1^b, S_2^b, \ldots, S_{|S^b|}^b)$ and lagging groups $S^w = \left(S_1^w, S_2^w, \ldots, S_{|S^w|}^w\right)$, which include $|S^b|$ and $|S^w|$ subpopulations correspondingly. In the traditional *MEC* algorithm a number of individuals within each subpopulation is set to be the same and equals $|S|$. Each of subpopulations S_i^b, S_j^w has its own communication environment named a local blackboard and denoted as C_i^b, C_j^w correspondingly. Besides, a multi-population as a whole $S = \{S^b, S^w\}$ has a common global blackboard C^g.

Canonical *MEC* is composed of three main stages: initialization of groups, similar-taxis and dissimilation. Operations of similar-taxis and dissimilation are repeated iteratively while the best obtained value of an objective function $\Phi(X)$ is changing. When the best obtained value stops changing, the winner of the best group from a set of leading ones is selected as a solution to the optimization problem [7, 8].

3 Co-Algorithm Based on *MEC*

The dissimilation stage of the *MEC* algorithm was modified in accordance with a concept of co-evolution [9]. After arranging groups and determining leading and lagging ones the redistribution of individuals within each group takes place following the rules.

- Define a minimum allowed size of a group $|S_{min}|$. Within the scope of this paper this value equals 20 % of the initial number of agents in each group $|S|$.
- A number of agents $|S_j^w|$ in every lagging group S_j^w is reduced proportionally to their scores at the global blackboard. A number of individuals in the best group from S^w is reduced by one, and a number of individuals in the worst group from S^w is reduced by a value of p_A, which can be calculated as follows

$$p_{|S^w|}^A = \left\lfloor \frac{(1 - |S_{min}|) \cdot |S|}{\omega} \right\rfloor. \tag{3}$$

Here ω—remove frequency of lagging groups, in other words if within ω iterations a group remains a lagging one, it is removed from a population; $\lfloor \cdot \rfloor$—a symbol of nearest least whole number.

- A number of individuals, being removed from intermediate groups is calculated using a linear approximation

$$p_k^A = \left\lfloor \frac{(k-1)\cdot(p_{|S^w|}^A - 1)}{|S^w| - 1} + 1 \right\rfloor, k \in [2 \colon |S^w| - 1]. \tag{4}$$

- When a number of agents in a particular group is below a minimum allowed value $|S_{min}|$ this number is set to be $|S_{min}|$.
- A number of agents in all leading groups except for the best one doesn't change.
- A number of individuals in the best group is increased by the number of individuals removed from other groups $\sum_{k=1}^{|S^w|} p_k^A$.

In such a manner, the total number of agents within a whole multi-population remains unchanged in the progress of computational process. A general scheme of *CoMEC*, can be described as follows.

1. Initialization of groups within the search domain D.
   (a) Generate a given number γ of groups $S_i, i \in [1 \colon \gamma]$, where γ—free parameter of the algorithm.
   (b) Generate a random vector $X_{i,1}$, whose components are distributed uniformly within the corresponding search subdomain. Identify this vector with the individual $s_{i,1}$ of the group S_i.
   (c) Determine the initial coordinates of the rest of individuals in the group $S_{i,j}, j \in [2 \colon |S|]$ following the formula

$$X_{i,j} = X_{i,1} + N_{|X|}(0, \sigma), \tag{5}$$

   in other words, they are placed randomly around the main individual $s_{i,1}$ in accordance with $|X|$-dimensional normal distribution law $N_{|X|}(0, \sigma)$, with zero mathematical expectation along all $|X|$ coordinates and standard deviation σ (another free parameter of the algorithm).
   (d) Calculate the scores of all individuals in every population S_i and put them on the corresponding local blackboards.
   (e) Create leading S^b and lagging S^w groups on the basis of obtained information.

2. Similar-taxis operation is performed in every group.
   (a) Take information on the current best individual $s_{i,j}, j \in [1 : |S_i|]$ of the group S_i from the blackboard C_i.
   (b) Create leading groups $S^b = (S_1^b, S_2^b, \ldots, S_{|S^b|}^b)$ and lagging groups $S^w = \left(S_1^w, S_2^w, \ldots, S_{|S^w|}^w\right)$ on the basis of obtained information.
   (c) Put information on the new winners in every group S_i of the multi-population on the corresponding local and global blackboards.
3. Dissimilation operation.
   (a) Read the scores of all groups $\Phi_i^b, \Phi_j^w, i \in \left[1 : \left|S^b\right|\right], j \in [1 : |S^w|]$ from the global blackboard C^g (scores of the best individuals in the groups).
   (b) Compare those scores. If score of any leading group S_i^b appeared to be less than score of any lagging group S_j^w, than the latter becomes a leading group and the first group becomes a lagging one. If score of a lagging group S_k^w is lower than scores of all leading groups for ω consecutive iterations, then it's removed from the population.
   (c) Redistribution of individuals between groups takes place in accordance with the rules specified above.
   (d) Using the initialization operation and formula (5) each removed group is replaced with a new one.
4. Evaluate the termination criteria. If either a number of stagnation iterations λ_{stop} or maximum allowed number of iterations λ_{max} exceed their limits then the computational process should be stopped, and the best current individual is set to be a solution X^* to the optimization problem. Otherwise it continues and goes to point 2.

4 Memetic Modifications of *MEC* and *CoMEC*

Efficiency of a memetic approach to the hybridization of population algorithms was demonstrated in the works [2, 4, 10]. In this regard, authors also proposed memetic modifications [11, 12] of classic *MEC*, named *mMEC* and *CoMEC*, named *mCoMEC*.

A usage of memes leads to the increase in computational complexity in the algorithms, that's why in order to maintain relatively low level of computational expenses it was decided to use only one meme at the stage of local search. Nelder-Mead method was selected for this purpose [13].

Memetic Mind Evolutionary Computation algorithm differs from simple *MEC* by a modified similar-taxis operation, which can be described as follows.

(a) Take information on the current best individual $s_{i,j}, j \in [1:|S_i|]$ of the group S_i from the blackboard C_i.
(b) Launch a selected meme from the current positions of each individual. Local search is carried out for l_s iterations.
(c) Determine new best individuals $s_{i,j}, j \in [1:|S_i|]$ in each group S_i with the lowest values of objective function.
(d) Create leading groups $\boldsymbol{S}^b = (S_1^b, S_2^b, \ldots, S_{|S^b|}^b)$ and lagging groups $\boldsymbol{S}^w = \left(S_1^w, S_2^w, \ldots, S_{|S^w|}^w\right)$ on the basis of obtained information.
(e) Put information on the new winners in every group S_i of the multi-population on the corresponding local and global blackboards.

This modified operation of similar-taxis is also used in memetic *CoMEC*.

5 Software Implementation and Numeric Experiments

All of the described algorithms were implemented by authors in *Wolfram Mathematica*. Software implementation has a modular structure, which helps to modify algorithms easily.

Taking into account that performance of the algorithms significantly depends on the random initial position of individuals, every computational experiment was repeated 50 times, i.e. the multi-start method was utilized. In these experiments the best obtained value of objective function Φ^* was used as the main performance index along with the probability estimation p of the global minimum localization. In addition, the average value of objective function¯ based on the results of all launches was utilized in order to estimate the influence of free parameters' values on the performance of the algorithms.

5.1 Benchmark Functions

Eight-dimensional ($|X| = 8$) multi-extremal benchmark Rastrigin and Griewank functions were considered in this work. Performance investigation was carried out within the following parallelepiped

$$D = \{X| - 5 \le x_i \le 5, i \in [1:|X|]\}, \quad (6)$$

where the objective functions reach their global minimum at the point $X^* = (0, 0, \ldots, 0)$; the global minima equal zero [14].

Furthermore, eight-dimensional ($|X| = 8$) ravine benchmark functions of Rosenbrock and Zakharov were considered within the parallelepiped (6). Global minima within that parallelepiped are reached at the points $X^* = (1, 1, \ldots, 1)$ and $X^* = (0, 0, \ldots, 0)$ correspondingly and equal zero [14].

5.2 Analysis of the Results

Computational experiments for *MEC* and *CoMEC* algorithms were carried out with the following values of free parameters: initial number of individuals in a group $|S| = 50$; number of groups $\gamma = 100$. In order to maintain approximately the same level of computational expenses those values were reduced for the experiments with the memetic modifications: $|S| = 20$, $\gamma = 25$. The number of iterations for local search was limited to $l_s = 10$. The number of stagnation iterations $\lambda_{stop} = 50$ was used as a termination criterion for the algorithms. Tolerance used for identifying stagnation was equal to $\varepsilon = 10^{-5}$.

The best obtained values of objective functions Φ^* as well as the probability estimation p of global minimum localization in multi-start mode and the corresponding values of the free parameters are presented in Table 1. As one can see from the results, none of the algorithms could localize the known global optimum of eight-dimensional Rastrigin function with the specified settings of algorithms. At the same time, known global optima for every other benchmark functions were localized with a high probability. Best performance was demonstrated by the *mMEC*.

Table 1 Best obtained values of objective function Φ^* and estimation of localization probability p

	MEC	*CoMEC*	*mMEC*	*mCoMEC*
Rastrigin function	$\Phi^* = 4.81$; $p = 0$ %; $\sigma = 0.75$; $\omega = 20$; $\eta = 50$ %;	$\Phi^* = 4.60$; $p = 0$ %; $\sigma = 0.1$; $\omega = 20$; $\eta = 50$ %;	$\Phi^* = 3.05$; $p = 0$ %; $\sigma = 0.1$; $\omega = 15$; $\eta = 50$ %;	$\Phi^* = 3.33$; $p = 0$ %; $\sigma = 1.0$; $\omega = 15$; $\eta = 50$ %;
Rosenbrock function	$\Phi^* = 0.34$; $p = 16$ %; $\sigma = 0.1$; $\omega = 20$; $\eta = 50$ %;	$\Phi^* = 0.24$; $p = 82$ %; $\sigma = 0.1$; $\omega = 25$; $\eta = 25$ %;	$\Phi^* = 0.15$; $p = 60$ %; $\sigma = 0.1$; $\omega = 25$; $\eta = 50$ %;	$\Phi^* = 0.21$; $p = 26$ %; $\sigma = 0.1$; $\omega = 20$; $\eta = 50$ %;
Griewank function	$\Phi^* = 1.2E-4$; $p = 100$ %; $\sigma = 0.1$; $\omega = 25$; $\eta = 50$ %;	$\Phi^* = 8.0E-5$; $p = 100$ %; $\sigma = 0.1$; $\omega = 30$; $\eta = 75$ %;	$\Phi^* = 8.0E-5$; $p = 100$ %; $\sigma = 0.1$; $\omega = 25$; $\eta = 50$ %;	$\Phi^* = 1.2E-4$; $p = 92$ %; $\sigma = 0.1$; $\omega = 20$; $\eta = 50$ %;
Zakharov function	$\Phi^* = 6.9E-4$; $p = 100$ %; $\sigma = 0.1$; $\omega = 30$; $\eta = 75$ %;	$\Phi^* = 2.5E-4$; $p = 100$ %; $\sigma = 0.1$; $\omega = 30$; $\eta = 50$ %;	$\Phi^* = 8.3E-4$; $p = 100$ %; $\sigma = 0.1$; $\omega = 25$; $\eta = 50$ %;	$\Phi^* = 1.3E-3$; $p = 100$ %; $\sigma = 0.1$; $\omega = 20$; $\eta = 50$ %;

5.3 Estimating the Influence of Free Parameters

One of the distinctive features of population algorithms is a large number of free parameters. Their performance significantly depends on the values of those parameters, but despite that, as a rule there are no recommendations on the choice of such values that would effectively deal with different classes of optimization problems.

Due to the absence of such recommendations for *MEC* algorithm, a comparison of the efficiency of this algorithm and its three proposed modifications was performed for different values of the free parameters. In this work, the performance of proposed algorithms and their software implementation was investigated depending on the values of the following free parameters: standard deviation σ, remove frequency of lagging groups ω, ratio η between lagging $|S^w|$ and leading groups $|S^b|$.

The values of variable parameters were chosen from discrete sets: $\sigma = \{0.25; 0.5; 0.75; 1.0\}$; $\omega = \{10, 15, 20, 25, 30\}$; $\eta = \{10, 25, 50, 75, 90\}$ %. In the beginning, the optimum value of the standard deviation was determined for each benchmark function; then with a particular σ the optimal values or remove frequency ω were find out. Finally, the optimal values of η were determined. This study was based on the average value of the objective function¯ only. Results of computational experiments are presented in Figs. 1, 2 and 3. Here, the quantity¯, % indicates the relative change in the average value of objective function.

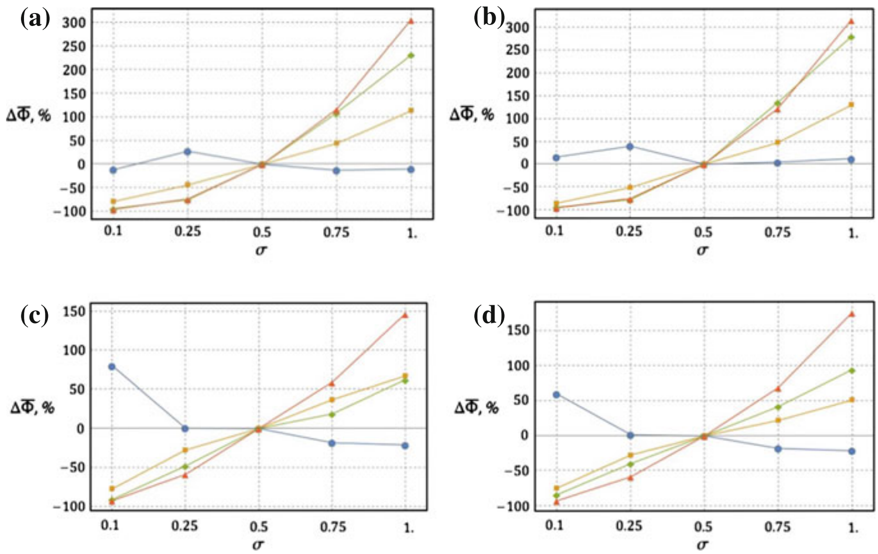


Fig. 1 Percentage change in obtained minimum of objective function while varying standard deviation σ: —Rastrigin; —Rosenbrock; —Griewank; —Zakharov. **a** *MEC*. **b** *CoMEC*. **c** *mMEC*. **d** *mCoMEC*

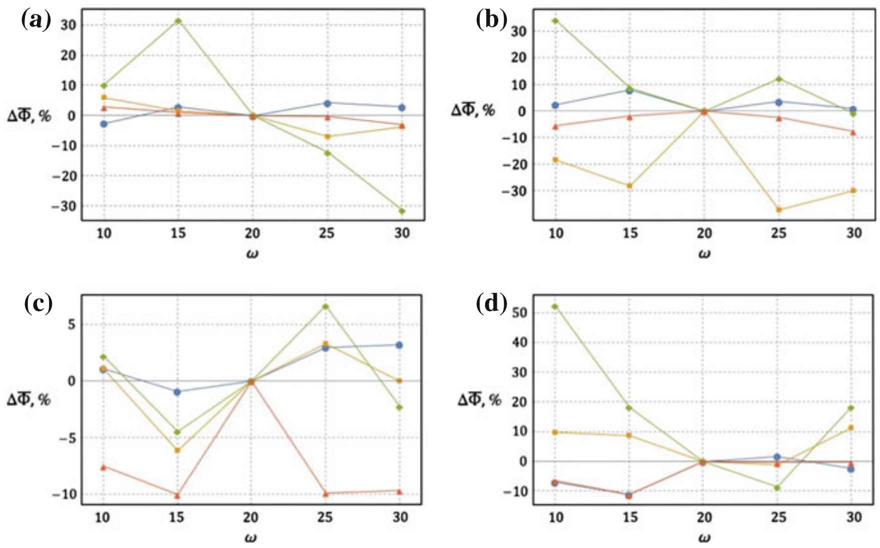


Fig. 2 Percentage change in obtained minimum of objective function while varying remove frequency ω: —Rastrigin; —Rosenbrock; —Griewank; —Zakharov. **a** *MEC*. **b** *CoMEC*. **c** *mMEC*. **d** *mCoMEC*

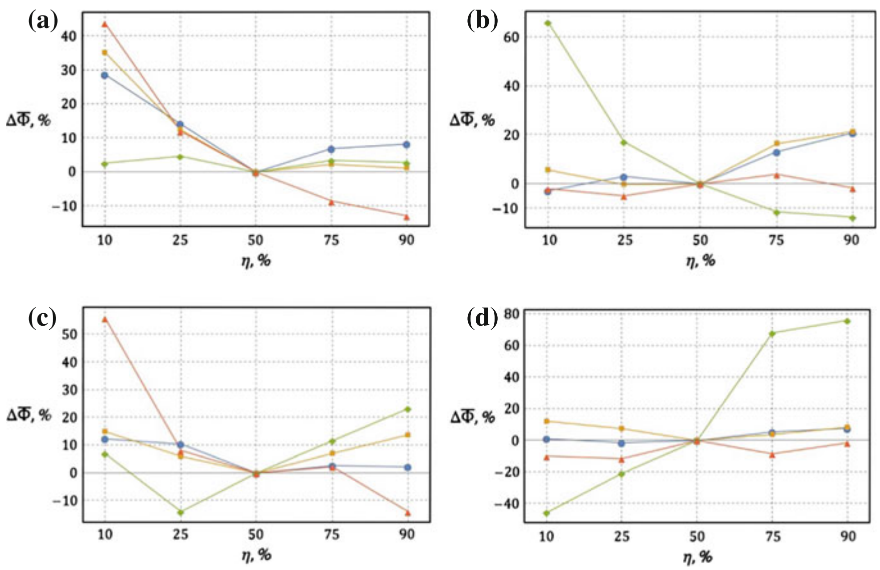


Fig. 3 Percentage change in obtained minimum of objective function while varying a ratio η: —Rastrigin; —Rosenbrock; —Griewank; —Zakharov. **a** *MEC*. **b** *CoMEC*. **c** *mMEC*. **d** *mCoMEC*

Results presented in Fig. 1 demonstrate that standard deviation σ affects the performance of algorithms significantly. Average value of the objective function grows by 3–4 times with an increase in standard deviation, compared to the base level for algorithms without the local search stage, namely *MEC* and *CoMEC*. The stage of local search in *mMEC* and *mCoMEC* allows one to smooth this effect, but it still remains—average value of the objective function may grow by 2.5 times. Despite this effect, it is recommended to select small values of the standard deviation.

At the same time, analyzing results obtained with a use of the Rastrigin function it can be concluded that the performance of algorithms also depends on the topology of objective function—final result is better if the distribution of individuals covers the area which exceeds the neighborhood of the nearest local minimum. Otherwise, an algorithm won't be able to "get away" from a local optimum, which results in the premature stagnation of computational process.

Results of computational experiments with a variation of remove frequency ω for lagging groups are demonstrated in Fig. 2. For *MEC* and *CoMEC* algorithms it is recommended to select a sufficiently high frequency; in such a manner lagging groups would have enough time to fully investigate their search domain. For memetic modifications of the specified algorithms a value of remove frequency ω does not cause any particular effect on the performance, because it is overlapped by the effect caused by a meme at the stage of similar-taxis.

Regarding the results obtained by varying the ratio between lagging and leading groups (Fig. 3), it appeared to be essential to have a relatively large number of leading groups for *MEC* and *mMEC* algorithms. Presence of a large number of lagging groups which are periodically removed from a multi-population with a certain frequency ω, results in the insufficient number of individuals required to locate the global optimum. On the other hand, in co-modifications this parameter is insignificant due to the re-distribution of individuals in favor of the best group. As a result, leading groups have a sufficient number of agents for successful localization of the global minimum.

6 Conclusions

Traditional Mind Evolutionary Computation algorithm was studied in this paper along with its three modifications proposed by authors. Modular software implementation of *MEC*, *CoMEC*, *mMEC* and *mCoMEC* was also presented. A wide performance investigation of those algorithms was carried out using benchmark functions. In addition, influence of the free parameters' values on the efficiency of those algorithms was analyzed.

Presented results demonstrate the superiority of proposed co-modification of Mind Evolutionary Computation algorithm over its classical version. *CoMEC* allows one to obtain better results without increasing a number of agents or groups. At the same time, co-algorithm is inferior to simple memetic hybridization scheme used in *mMEC* and *mCoMEC* in terms of efficiency. It also should be noticed that a combination of co-evolution strategy and the concept of memes in *mCoMEC* didn't bring the desired synergetic effect—the algorithm is inferior to *mMEC* and sometimes even to other non-memetic algorithms in terms of performance.

Results obtained during the investigation of the influence of the free parameters' values on the algorithms' performance confirmed a significant effect of these parameters on the value of best-obtained optimum. Identified dependencies can be used to give recommendations on selecting initial values of these free parameters, as well as to classify test functions by their topology. As an example, for Rosenbrock, Griewank and Zakharov functions best values of global minimum were achieved with standard deviation $\sigma = 0.1$. This can be explained by the fact that local minimums of these functions are located sufficiently far from each other; as a result, algorithms are capable of searching subdomains intensively and locating global optimum with high probability. However, this is not applicable to the Rastrigin function where local minimums are located close enough to each other. It can be seen from the results (Fig. 1) that for small values of σ algorithms converge to an arbitrary local optimum. For functions, similar to the Rastrigin one, selection of small values for σ can lead to non-optimal results.

Further research will be devoted to the analysis of mechanisms for preventing premature convergence of algorithms and evaluating the diversity of agents in populations.

References

1. Weise, T.: Global Optimization Algorithms—Theory and Application, 758 p. University of Kassel (2008)
2. Sakharov, M.K., Karpenko, A.P., Velisevich, Ya.I. Multi-Memetic Mind Evolutionary Computation Algorithm for Loosely Coupled Systems of Desktop Computers. Science and Education of the Bauman MSTU, 2015, vol. 10, pp. 438–452. doi:10.7463/1015.0814435
3. Karpenko A.P.: Modern algorithms of search engine optimization. Nature-inspired optimization algorithms, 446 p. Moscow, Bauman MSTU Publication (2014) (in Russian)
4. Karpenko, A.P., Sakharov, M.K.: Multi-Memes Global Optimization Based on the Algorithm of Mind Evolutionary Computation. Inform. Technol. **7**, 23–30 (2014) (in Russian)
5. Chengyi, S., Yan, S., Wanzhen, W.: A Survey of MEC: 1998–2001. In: 2002 IEEE International Conference on Systems, Man and Cybernetics IEEE SMC2002, Hammamet, Tunisia, vol. 6, pp. 445–453, October 6–9. Institute of Electrical and Electronics Engineers Inc. (2002)
6. Sakharov, M.K.: Study on Mind Evolutionary Computation. Technologies and Systems 2014, pp. 75–78. Moscow, Bauman MSTU Publication (2014)
7. Jie, J., Zeng, J.: improved mind evolutionary computation for optimizations. In: Proceedings of 5th World Congress on Intelligent Control and Automation, pp. 2200–2204. Hang Zhou, China (2004)
8. Jie, J., Han, C., Zeng, J.: An extended mind evolutionary computation model for optimizations. Appl. Math. Comput. **185**, 1038–1049 (2007)

9. Vorobeva, E.Y., Karpenko, A.P., Seliverstov, E.Y.: Co-evolutionary algorithm of global optimization based on particle swarm optimization. Science and Education of the Bauman MSTU, vol. 4, pp. 431–474 (2012). doi:10.1007/1113.0619595
10. Karpenko, A., Posypkin, M., Rubtsov, A., Sakharov, M.: Multi-memetic global optimization based on the mind evolutionary computation. In: Proceedings of the IV International Conference on Optimization Methods and Application "Optimization and applications" OPTIMA-2013, pp. 83–84. Moscow (2013)
11. Neri, F., Cotta, C., Moscato, P.: Handbook of Memetic Algorithms, 368 p. Springer, Berlin, Heidelberg (2011). doi:10.1007/978-3-642-23247-3 (Ser. Studies in Computational Intelligence; vol. 379)
12. Nguyen, Q.H., Ong, Y.S., Krasnogor, N.A Study on the design issues of memetic algorithm. In: IEEE Congress on Evolutionary Computation, pp 2390–2397 (2007)
13. Nelder, J.A., Meade, R.: A simplex method for function minimization. Comput. J. **7**, 308–313 (1965)
14. Floudas, A.A., Pardalos, P.M., Adjiman, C., Esposito, W.R., Gümüs, Z.H., Harding, S.T., Klepeis, J.L., Meyer, C.A., Schweiger, C.A.: Handbook of Test Problems in Local and Global Optimization, p. 441. Kluwer, Dordrecht (1999)

Three Types of Differential Evolution Applied to the Facility Location Problem

Pavel Krömer, Jan Platoš and Václav Snášel

Abstract Facility location problem is a combinatorial optimization problem with many variants and a number of real-world applications in operations research and industry. A number of exact and approximate methods has been developed to tackle this complex task. Populational nature-inspired metaheuristic optimization methods have been investigated in context of the facility location problem as well. They are able to find excellent problem solutions, but often rely on domain-specific local search and employ heuristic steps. In this work, we study the application of three different flavours of the differential evolution algorithm to a hard variant of the facility location problem. The methods are used as pure metaheuristics without any domain specific knowledge and the efficiency of different optimization strategies they represent is evaluated and compared on a test data set.

Keywords Differential evolution · Facility location problem · Experiments · Comparison

1 Introduction

Many industrial applications require solving complex assignment, allocation, and transportation problems. The problems are in real-world scenarios often tightly connected with many internal links and dependencies. Practical industrial tasks can be mapped to a number of abstract models formulated within the areas of operations research, scheduling, and planning. Facility location problem is a many-sided

P. Krömer (✉) · J. Platoš · V. Snášel
Department of Computer Science, VŠB Technical University of Ostrava,
Ostrava, Czech Republic
e-mail: pavel.kromer@vsb.cz

J. Platoš
e-mail: jan.platos@vsb.cz

V. Snášel
e-mail: vaclav.snasel@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 450, DOI 10.1007/978-3-319-33609-1_44

problem that combines allocation and transportation elements into a single ... task. Various exact and approximate methods have been developed to solve this hard problem.

Nature-inspired metaheuristic methods are among the most universal optimization algorithms that can be applied to a variety of different problems. They possess the ability to solve complex problems and to adapt to different characteristics of problem instances. However, no single metaheuristic optimization method is suitable for all types of problems [17]. Therefore, experimental evaluation of different types of optimization strategies, represented by different metaheuristics and their particular variants, is needed to assess their performance on different problems. In this work, we compare the results and performance of three popular varieties of differential evolution on a hard variant of the facility location problem.

2 Single Source Capacitated Facility Location Problem

Facility location problem (FLP) is a combinatorial optimization problem that consists of selecting a set of facility locations on their setup costs, production capacities, customer demand, and service (e.g. transportation) costs. Capacitated FLP is a realistic FLP variant which assumes that each facility has a fixed production capacity [15]. There are many other variants of this generic hard optimization problem. Single source capacitated facility location problem (SSCFLP) [5, 7, 15] is a challenging NP-hard FLP version which requires that each customer is served from exactly one facility.

Consider a set of potential facility locations, $J = \{1, \dots, m\}$, with opening costs and production capacity in each location denoted by f_j and s_j, respectively. For each customer, $i \in I = \{1, \dots, n\}$, with demand, d_i, assume service costs, c_{ij}, as the price of serving its full demand from a facility, j. SSCFLP can be then defined as a binary integer linear programming problem that minimizes a cost function [5, 7]

$$\sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} + \sum_{j \in J} f_j y_j, \tag{1}$$

subject to

$$\sum_{i \in I} x_{ij} = 1 \quad \forall j \in J, \qquad \sum_{j \in J} d_j x_{ij} \leq s_j y_j \quad \forall j \in J, \tag{2}$$

$$y_j = \{0, 1\} \quad \forall j \in J, \quad \text{and } x_{ij} = \{0, 1\} \quad \forall i \in I, \forall j \in J. \tag{3}$$

Binary decision variables y_i and x_{ij} are defined as

$$y_j = \begin{cases} 1, & \text{if } j \text{ is open} \\ 0, & \text{otherwise} \end{cases}, \quad x_{ij} = \begin{cases} 1, & \text{if } i \text{ is served from } j \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

Different approaches have been adopted to solve the SSCFLP. Heuristic methods based on e.g. subgradient optimization (Lagrangean heuristic) [7] and kernel search [5] have shown good ability to find SSCFLP solutions. Metaheuristic algorithms, such as genetic algorithms (GAs) [3], and the iterated tabu search (ITS) algorithm [6], have been used to tackle this complex problem as well. In contrast with heuristic methods, they scale well and are able to find reasonable solutions for a wide range of SSCFLP instances with different properties. In the many cases, they are coupled with some sort of local search that is used to intensively explore promising solution space regions identified by the algorithms [].

2.1 Nature-Inspired Metaheuristics for the Facility Location Problem

Add some final motivation (we want to investigate the power of pure metaheuristics).

Caserta and Rico [2] presented a three-phase algorithm for solving CFLP. The first phase is based on the cross-entropy scheme for selection of the opened facility. The second phase is a local search algorithm for searching around actually selected solution. The third phase stochastically optimizes parameters. The suggested algorithm can find a globally optimal solution for tested datasets.

Sun [14] described an algorithm which uses a Tabu search principle for heuristic local search. His algorithm uses an efficient memory structure for optimization of the search procedure, and avoidance of the reusing already found the solution.s

Arostegui Jr. et al. [1] compared three different meta-heuristic algorithms on the CFLP. These three algorithms (Tabu Search, Simulated Annealing, and Genetic Algorithms) attacks the problem of CFLP from the different point of view, but all of them can solve it. According to the performed experiments, the TS is the best for time-limited evaluation constraint, Simulated Annealing is the best when a solution evaluation count is set as a constraint.

3 Differential Evolution

The DE is a population-based optimizer that evolves a population of real encoded vectors representing the solutions to given problem. The DE was introduced by Storn and Price in 1995 [12, 13] and it quickly became a popular alternative to the more traditional types of evolutionary algorithms. It evolves a population of candidate solutions by iterative modification of candidate solutions by the application of the differential mutation and crossover [10]. In each iteration, so called trial vectors are created from current population by the differential mutation and further modified by various types of crossover operator. At the end, the trial vectors compete with existing candidate solutions for survival in the population.

The DE starts with an initial population of N real-valued vectors. The vectors are initialized with real values either randomly or so, that they are evenly spread over the problem space. The latter initialization leads to better results of the optimization [10]. During the optimization, the DE generates new vectors that are scaled perturbations of existing population vectors. The algorithm perturbs selected base vectors with the scaled difference of two (or more) other population vectors in order to produce the trial vectors. The trial vectors compete with members of the current population with the same index called the target vectors. If a trial vector represents a better solution than the corresponding target vector, it takes its place in the population [10].

The two most significant parameters of the DE are scaling factor and mutation probability [10]. The scaling factor $F \in [0, \infty]$ controls the rate at which the population evolves and the crossover probability $C \in [0, 1]$ determines the ratio of bits that are transferred to the trial vector from its opponent. The size of the population and the choice of operators are another important parameters of the optimization process.

The classic DE has shown the ability to solve a wide range of problems. However, its performance in particular domains strongly relies on the selection of differential mutation and crossover operators as well as parameters F and C [11]. A number of self-adaptive DE variants was designed to mitigate this dependence. Among them, the Self-Adaptive Differential Evolution (SaDE) algorithm has shown good results for many types of tasks [4, 11]. Another interesting parameter-free DE variant is called Gaussian Bare-bones Differential Evolution (GBDE) [16].

3.1 Self-Adaptive Differential Evolution

SaDE is based on probabilistic selection of trial vector generation strategy based on historical performance of different strategies, randomization of scaling factor, F, and adaptation of crossover probability, C.

In each generation, G, select for every target vector, $\boldsymbol{x}^i$, trial vector generation strategy s_k from a pool of strategies, $S = \{s_1, s_2, \dots s_K\}$, with respect to strategy selection probability $p_{k,G}$. The strategy selection probability, $p_{k,G}$, is adapted on the basis of the number of successes (i.e. number of times a trial vector, $\boldsymbol{v}^i$, is better solution than target vector, $\boldsymbol{x}^i$) and failures (i.e. number of times $\boldsymbol{v}^i$ is worse solution than $\boldsymbol{x}^i$) of trial vectors generated by s_k during a fixed number of past generations known as *learning period*, *LP*. The algorithm stores the successes and failures of each strategy into success and failure memories, **SM** and **FM**, that store the number of successes, $ns_{k,g}$, and failures, $nf_{k,g}$, of each strategy, s_k, in the past LP generations.

Strategy selection probabilities are then in each generation, G, $G > LP$, updated by

$$p_{k,G} = \frac{S_{k,G}}{\sum_k S_{k,G}}, \quad S_{k,G} = \frac{\sum_{g=G-LP}^{G-1} ns_{k,g}}{\sum_{g=G-LP}^{G-1} ns_{k,g} + \sum_{g=G-LP}^{G-1} nf_{k,g}} + \epsilon, \tag{5}$$

where ϵ is a small constant (here, $\epsilon = 0.01$) employed to tackle cases with zero success rate [11]. Initial strategy selection probabilities are for the first *LP* generations set to be equal, i.e. $p_{k,G} = \frac{1}{K}, k \in \{1, 2, \ldots, K\}$.

Trial vector generation strategies can include arbitrary combinations of differential mutation and crossover. The strategies used in this study are summarized in Fig. 1. The strategies *DE/rand/1/bin* and *DE/rand/2/bin* have slow convergence but strong exploration capability. *DE/rand-to-best/2/bin* has fast convergence, especially for unimodal problems, but tends to get trapped in local optima and suffers from premature convergence. *DE/current-to-rand/1* is a rotation invariant type of DE that has good efficiency for rotated problems [11].

Scaling factor *F* is in SaDE for each trial vector selected randomly from normal distribution with mean 0.5 and standard deviation 0.3 [11]:

$$F_{i,G} = N(0.5, 0.3) \tag{10}$$

Scaling factors drawn from such distribution fall in 99.7 % of cases into the range [−0.4, 1.4] allowing for both, exploitation (small *F*) and exploration (large *F*) [11].

The value of crossover probability is an important problem-dependent parameter that has a major impact of algorithm performance [11]. In SaDE, it is for generation of each trial vector using a particular strategy, s_k, generated using

$$C_{k,G} = N(Cm_{k,G}, 0.1), \tag{11}$$

DE/rand/1/bin:

$$v_j^i = \begin{cases} v_j^{r1} + F(v_j^{r2} - v_j^{r3}), & \text{if } rand(0,1) < C \text{ or } j = j_{\text{rand}} \\ x_j^i, & \text{otherwise} \end{cases}, \tag{6}$$

DE/rand-to-best/2/bin:

$$v_j^i = \begin{cases} x_j^i + F(x_j^{\text{best}} - x_j^i) + F(v_j^{r1} - v_j^{r1}) + F(v_j^{r3} - v_j^{r4}), & \text{if } rand(0,1) < C \text{ or } j = j_{\text{rand}} \\ x_j^i, & \text{otherwise} \end{cases} \tag{7}$$

DE/rand/2/bin:

$$v_{i,j}^t = \begin{cases} v_j^{r1} + F(v_j^{r2} - v_j^{r3}) + F(v_j^{r4} - v_j^{r5}), & \text{if } rand(0,1) < C \text{ or } j = j_{\text{rand}} \\ x_j^i, & \text{otherwise} \end{cases}, \tag{8}$$

DE/current-to-rand/1:

$$\boldsymbol{v}^i = \boldsymbol{x}_j^i + \kappa(\boldsymbol{v}^{r2} - \boldsymbol{x}^i)F(\boldsymbol{v}^{r2} - \boldsymbol{v}^{r3}), \tag{9}$$

where $\boldsymbol{x}^{\text{best}}$ is the best solution found so far and κ is a random parameter generated for each trial vector, $\boldsymbol{v_i}$, using $\kappa = rand(0,1)$.

Fig. 1 SaDE trial vector generation strategies where $Cm_{k,G}$ is mean of the (normal) random distribution of Cs for strategy s_k in generation G based on the *Cm memory*, **CmM**, storing the crossover probabilities used in the past LP generations when creating successful trail vectors.

3.2 Gaussian Bare-Bones Differential Evolution

GBDE extends the original Bare-bones Differential Evolution algorithm [9] which was inspired by a particle swarm optimization (PSO) variant called Bare-bones PSO [8]. Bare-bones PSO, based on theoretical analysis of PSO convergence, eliminates particle velocity and samples new position of a particle, i, from Gaussian distribution with mean and standard deviation defined by the best position visited by the swarm, $\mathbf{y}$, and best position visited by the particle, $\boldsymbol{y}_i^t$ so that new particles are centered around weighted average of $\mathbf{y}$ and $\boldsymbol{y}_i^t$. The exploration/exploitation ratio is automatically adjusted from an initial focus on exploration to later focus on exploitation [16].

GBDE applies similar principles within the framework of DE. It uses a Gaussian mutation strategy defined by

$$\boldsymbol{v}^i = N(\mu, \sigma), \tag{12}$$

where N is a Gaussian random distribution with mean μ and standard deviation σ, respectively

$$\mu = \frac{\boldsymbol{x}^{\text{best}} + \boldsymbol{x}^i}{2}, \qquad \sigma = |\boldsymbol{x}^{\text{best}} - \boldsymbol{x}^i|, \tag{13}$$

where $\boldsymbol{x}^{\text{best}}$ is the best solution found so far and $\boldsymbol{x}^i$ is the target vector.

GBDE also employs DE's traditional binomial crossover. To avoid the need for manually choosing crossover probability, C, it uses the following self-adaptive strategy for its dynamic selection

$$C^i_{G+1} = \begin{cases} C^i_G, & \text{if } f_{\text{obj}}(\boldsymbol{v}^i) \leq f_{\text{obj}}(\boldsymbol{x}^i) \\ N(0.5, 0.1), & \text{otherwise} \end{cases}, \tag{14}$$

where C^i_G is crossover probability associated with ith target vector, $\boldsymbol{x}^i$, in generation G, and $N(0.5, 0.1)$ is a random value sampled from a Gaussian distribution with median 0.5 and standard deviation 0.1. The strategy attempts to change the crossover probability C^i every time it did not generate a better solution than the target vector $\boldsymbol{x}^i$. Due to its stochastic nature, Gaussian mutation prioritizes exploration over exploitation. Modified GBDE algorithm (MGBDE) [16] chooses between Gaussian mutation (12) and */DE/best* mutation, defined by

$$\boldsymbol{v}^i = \boldsymbol{x}^{\text{best}} + F(\boldsymbol{v}^{r1} - \boldsymbol{v}^{r2}), \tag{15}$$

at random.

4 Computational Experiments

A series of computational experiments was conducted to study the ability of the DE, GBBDE, and SaDE to solve the SSCFLP. All three algorithms were implemented in C++ and applied to solve a set of SSCFLP instances. In order to use them, problem specific encoding and fitness function were developed. Each candidate vector, $\mathbf{c}$, consists of n floating point values, $c_i \in [0, m-1]$. The floating point values are transformed into the discrete domain simply by truncation and correspond to the index of the facility, assigned to serve the corresponding customer. The fitness function, used in our experiments, is defined by

$$f_{\text{fit}}(\mathbf{c}) = \alpha \frac{cost(\mathbf{c})}{n \cdot \max\limits_{i\in I, j\in J}(c_{ij}) + \sum\limits_{j\in J} f_j} + \beta \frac{unsat(\mathbf{c})}{\sum\limits_{i\in I} d_i}, \tag{16}$$

where $cost(\mathbf{c})$ is the total cost of the solution specified by candidate vector $\mathbf{c}$, $unsat(\mathbf{c})$ is the total amount of unsatisfied demand, and α and β are scalar weights of fitness function terms. The first term of (16) considers the cost of the solution while the second term introduces penalty for unsatisfied demand (i.e. violation of the constraints (2), (3)).

The encoding allows a creation of invalid solutions in which the total demand exceeds the capacity of the facilities. The fitness formulae add to each infeasible solution a penalty that prevents it from surviving in the population in the long term but allows several recombinations that can utilize its potentially useful parts.

A publicly available collection of SSCFLP instances was used to compare the performance of the investigated DE variants. The collection, used recently in [6], is available from http://www-eio.upc.edu/~elena/sscplp/. It contains 57 SSCFLP instances, termed p01 to p57, with $n \in [20, 90]$ customers and $m \in [10, 30]$ facilities. In these instances, opening costs dominate the assignment costs and enable solutions for only a small number of opened facilities.

The algorithms were executed with the following fixed parameters, determined on the basis of best practices, previous experience, and extensive trial-and-error runs: the population size was set to 100, the scaling factor, F, and the crossover probability were set to 0.9, respectively. The total number of fitness function evaluations was set to 100,000,000. Fitness function parameters, α and β, were set to 1.0 and 10.0, respectively. The DE, used in the experiments, was the traditional $/DE/rand/1$ version of the algorithm.

Table 1 Final fitness of the best, average, and worst solutions of the test SS- CFLP instances obtained by DE, GBBDE, and SaDE, respectively

	DE			GBBDE			SaDE		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
p01	2150	2341.04	2505	2168	2317.89	2471	2036	2126.19	2247
p02	4674	4952.1	5270	4574	4948.29	5245	4251	4457.79	5253
p03	6222	6821.21	7176	6272	6793.44	7179	6051	6245.4	7687
p04	7588	8153.69	8448	7482	8142.54	8390	7302	7479.25	8617
p05	4664	5018.58	5462	4729	5004.85	5477	4567	4623.67	5765
p06	2387	2607.56	2842	2349	2591.39	2818	2277	2283.15	2473
p07	5397	5779.62	6466	5166	5755.31	6259	4520	4894.94	5498
p08	1244	2047.83	2421	1244	1980.54	2262	1244	2292.33	5430
p09	3129	3363.75	3634	2974	3378.79	3645	2640	2759.85	3013
p10	26208	27374.85	29500	25096	27304.02	29217	24096	24884.46	29057
p11	3960	4237.71	4488	3924	4214.83	4484	3554	3669.9	4197
p12	4197	4534.65	4900	4375	4592.5	4960	3802	3948.35	4117
p13	4307	4586.27	4908	4228	4634.9	5107	3803	3993.25	4446
p14	6955	7647.1	8570	7198	7576.06	7936	6774	7335.48	10234
p15	8939	9507.56	10196	9019	9527.79	10092	8278	8506.83	9261
p16	13193	13949.44	14781	13346	13938.4	14760	12051	12545.44	13411
p17	12011	12554.27	13910	11935	12602.29	13477	9949	10670.27	11823
p18	18014	19263.27	21098	17943	19483.52	20874	16360	17160.67	19658
p19	21018	22627.96	25329	21035	22445.08	24827	20075	20813	24480
p20	29993	32362.33	35899	30569	32588.23	35149	27898	29458.75	35970
p21	8353	9157.92	9820	8639	9175.04	9580	7848	8089.52	8712

(continued)

Table 1 (continued)

	DE			GBBDE			SaDE		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
p22	4995	5297.08	5705	4985	5337.75	5694	3725	4316.65	4829
p23	7613	8522.48	9494	7423	8467.46	10045	6946	7704.23	8503
p24	9039	10421.79	11391	9663	10553.98	11430	6707	8376	9602
p25	10758	13054. 48	16152	10876	12715.92	15295	10485	12598. 44	16023
p26	5858	6456.02	6948	5966	6442.71	7034	5080	5471.64	5805
p27	14569	17501.31	22272	13842	17362.4	20561	13914	16606.73	19058
p28	13579	15128.31	16699	13379	15127.77	17424	13394	15363.37	18124
p29	11307	13063.21	15063	11027	13098.23	15633	10305	11039.89	15443
p30	12857	13526.89	14222	13015	13553.4	14269	11435	11825.56	12053
p31	5864	6234.17	6865	5583	6147.17	6734	4887	5122.94	5763
p32	14204	16021.35	18241	13895	16087.02	17657	11972	13168.25	14833
p33	46793	49962.14	53694	46538	49835.06	53401	42008	44266.25	45989
p34	8943	10542.46	12554	8886	10474. 19	12318	8192	9379.71	10555
p35	9476	10813.77	12570	9548	10952.1	12984	7536	8731.96	10552
p36	22886	27467.5	31834	22602	27458.46	34257	23887	28724.02	31647
p37	21470	27109.92	33346	21290	26682.33	30642	18102	23258.1	28685
p38	54666	61207.08	66600	56735	62185.77	68325	51281	54910.35	57418
p39	52448	57001.79	62932	49210	56746.94	65206	45780	50487.69	53331
p40	75245	83360.19	90961	76297	83579.23	94386	69346	74203.29	78144

(continued)

Table 1 (continued)

	DE			GBBDE			SaDE		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
p41	56742	90313.04	1.18×10^5	56742	87412.83	1.08×10^5	55986	73367.19	91784
p42	11810	14274.44	17396	11808	14426.9	17490	10557	13435.92	15640
p43	8714	11498.98	14948	8474	11085.14	14282	9449	11565.21	13456
p44	43672	47124.73	50757	44140	47311.14	51855	38945	40407.15	44703
p45	26270	34819.19	40597	29983	35429.27	40662	26272	32708.15	40267
p46	74547	91714.42	1.18×10^5	81610	92076.37	1.08×10^5	71980	91395.54	1.04×10^5
p47	78730	91430.48	1.06×10^5	84408	91824.48	1.10×10^5	75783	87801.33	1.05×10^5
p48	1.19×10^5	1.48×10^5	1.86×10^5	98833	1.45×10^5	1.75×10^5	1.20×10^5	1.39×10^5	1.56×10^5
p49	1.00×10^5	1.15×10^5	1.34×10^5	1.01×10^5	1.15×10^5	1.31×10^5	97803	1.06×10^5	1.14×10^5
p50	12229	14094.96	17001	11826	14089.81	16394	9911	13228. 77	16422
p51	17681	20328.23	23643	16791	20310.31	24325	14822	17082.69	18572
p52	44333	52650.9	61426	45986	52232.54	59822	42438	48864.19	53826
p53	41517	46526.54	56025	41217	47197.56	54133	38607	41997.37	45874
p54	97988	1.21×10^5	1.43×10^5	98200	1.19×10^5	1.47×10^5	89208	1.21×10^5	1.50×10^5
p55	85357	97638.37	1.10×10^5	87637	99447.27	1.12×10^5	79041	86910.39	91589
p56	1.05×10^5	1.44×10^5	1.73×10^5	1.05×10^5	1.42×10^5	1.77×10^5	1.24×10^5	1.64×10^5	1.85×10^5
p57	99931	1.44×10^5	2.17×10^5	99887	1.45×10^5	1.84×10^5	1.39×10^5	1.73×10^5	2.04×10^5

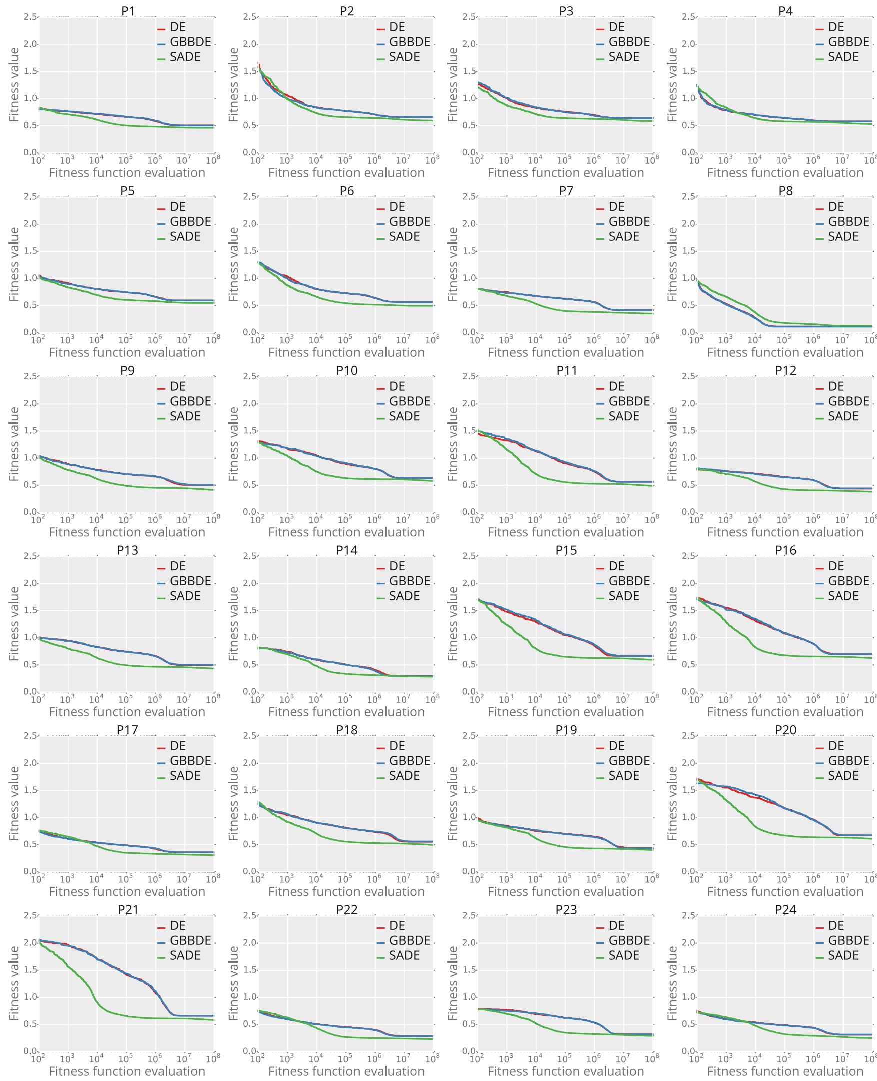


Fig. 2 Comparison of average DE, SaDE, and GBBDE runs on the first 24 test problems

Due to the stochastic nature of the algorithms, each problem instance was solved by each algorithm 50 times independently. The results of the experiments are summarized in Table 1.

The table shows that the self-adaptive optimization strategy, represented by SaDE, appears to be the most efficient entirely metaheuristic approach to the SSCFLP. SaDE has found best solutions better or on par with the best solutions found by DE and GBBDE for 49 test problem instances. The GBBDE algorithm ranked second by finding best solutions for 8 problem instances and the traditional DE was least successful with best solutions found only for 2 problem instances.

The average optimization runs of each algorithm for the first 24 problem instances are displayed in Fig. 2 (note the logarithmic scale of the x-axis on each subfigure). The images show the typical behavior of the algorithms. SaDE does not only deliver the best results but also converges faster than other algorithms in most cases.

5 Conclusions

Three variants of differential evolution were studied in context of the single source capacitated facility location problem. A simple encoding scheme, translating the combinatorial optimization problem into the continuous domain, and a novel fitness function, were defined to enable the application of real-parameter optimization methods to SSCFLP. Traditional differential evolution, self-adaptive differential evolution, and modified gaussian bare-bones differential evolution were implemented and used to solve problem instances from a test SSCFLP collection. The self-adaptive differential evolution has found better solutions for most of the test problems in the majority of test runs. Moreover, it has in the average case quickly converged to better solutions than other evaluated algorithms.

Acknowledgments This work was supported by the Czech Science Foundation under the grant no. GACR P103/16/25694Y and by SGS project, VSB-Technical University of Ostrava, under the grant no. SP2016/97.

References

1. Arostegui, Jr., M.A.A., Kadipasaoglu, S.N., Khumawala, B.M.: An empirical comparison of tabu search, simulated annealing, and genetic algorithms for facilities location problems. Int. J. Prod. Econ. **103**(2), 742–754 (2006). http://www.sciencedirect.com/science/article/pii/S0925527306000971
2. Caserta, M., Rico, E.Q.: A cross entropy-based metaheuristic algorithm for large-scale capacitated facility location problems. J. Oper. Res. Soc. **60**(10), 1439–1448 (2009)
3. Cortinhal, M.J.A., Captivo, M.E.: Metaheuristics. Genetic Algorithms for the Single Source Capacitated Location Problem, pp. 187–216. Kluwer Academic Publishers, Norwell (2004). http://dl.acm.org/citation.cfm?id=982409.982419
4. Elsayed, S.M., Sarker, R.A., Essam, D.L.: An improved self-adaptive differential evolution algorithm for optimization problems. IEEE Trans. Ind. Inf. **9**(1), 89–99 (2013). http://dx.doi.org/10.1109/TII.2012.2198658
5. Guastaroba, G., Speranza, M.: A heuristic for BILP problems: the single source capacitated facility location problem. Eur. J. Oper. Res. **238**(2), 438–450 (2014). http://www.sciencedirect.com/science/article/pii/S0377221714003166
6. Ho, S.C.: An iterated tabu search heuristic for the single source capacitated facility location problem. Appl. Soft Comput. **27**(C), 169–178 (2015). http://dx.doi.org/10.1016/j.asoc.2014.11.004

7. Holmberg, K., Rönnqvist, M., Yuan, D.: An exact algorithm for the capacitated facility location problems with single sourcing. Eur. J. Oper. Res. **113**(3), 544–559 (1999)
8. Kennedy, J.: Bare bones particle swarms. In: Swarm Intelligence Symposium. SIS'03. Proceedings of the 2003, pp. 80–87. IEEE (2003)
9. Omran, M.G.H., Engelbrecht, A.P., Salman, A.: Bare bones differential evolution. Eur. J. Oper. Res. **196**(1), 128–139 (2009). http://dblp.uni-trier.de/db/journals/eor/eor196.html#OmranES09
10. Price, K.V., Storn, R.M., Lampinen, J.A.: Differential Evolution a Practical Approach to Global Optimization. Natural Computing Series. Springer, Berlin (2005)
11. Qin, A., Huang, V., Suganthan, P.: Differential evolution algorithm with strategy adaptation for global numerical optimization. IEEE Trans. Evol. Comput. **13**(2), 398–417 (2009)
12. Storn, R.: Differential evolution design of an IIR-filter. In: Proceeding of the IEEE Conference on Evolutionary Computation ICEC, pp. 268–273. IEEE Press (1996)
13. Storn, R., Price, K.: Differential Evolution: a Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces. Technical report (1995)
14. Sun, M.: A tabu search heuristic procedure for the capacitated facility location problem. J. Heur. **18**(1), 91–118 (2012)
15. Verter, V.: Foundations of Location Analysis. Uncapacitated and Capacitated Facility Location Problems, pp. 25–37. Springer US, Boston (2011). http://dx.doi.org/10.1007/978-1-4419-7572-0_2
16. Wang, H., Rahnamayan, S., Sun, H., Omran, M.: Gaussian bare-bones differential evolution. IEEE Trans. Cybern. **43**(2), 634–647 (2013)
17. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. IEEE Trans. Evol. Comput. **1**(1), 67–82 (2002). http://dx.doi.org/10.1109/4235.585893

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 450, DOI 10.1007/978-3-319-33609-1