Advances in Intelligent Systems and Computing 521

Leszek Borzemski
Adam Grzech
Jerzy Świątek
Zofia Wilimowska *Editors*

Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I

Springer

Advances in Intelligent Systems and Computing

Volume 521

Series editor

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

About this Series

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

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

Advisory Board

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

Leszek Borzemski · Adam Grzech
Jerzy Świątek · Zofia Wilimowska
Editors

Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I

Springer

Editors
Leszek Borzemski
Department of Computer Science
Faculty of Computer Science
and Management
Wrocław University of Technology
Wrocław
Poland

Jerzy Świątek
Department of Computer Science
Faculty of Computer Science
and Management
Wrocław University of Technology
Wrocław
Poland

Adam Grzech
Department of Computer Science
Faculty of Computer Science
and Management
Wrocław University of Technology
Wrocław
Poland

Zofia Wilimowska
Department of Management Systems
Faculty of Computer Science
and Management
Wrocław University of Technology
Wrocław
Poland

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-46582-1 ISBN 978-3-319-46583-8 (eBook)
DOI 10.1007/978-3-319-46583-8

Library of Congress Control Number: 2016951674

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This four-volume set of books contains the proceedings of the 37th International Conference on Information Systems Architecture and Technology, or ISAT 2016 for short, held on September 18–20, 2016, in Karpacz, Poland. The conference was organized by the Department of Management Systems and the Department of Computer Science, Wrocław University of Science and Technology, Poland.

The International Conference on Information Systems Architecture and Technology is organized by the Wrocław University of Science and Technology from the seventies of the last century. The purpose of the ISAT conference is to discuss a state of the art of the information systems concepts and applications as well as the architectures and technologies supporting modern information systems. Contemporary organizations seem to be knowledge-based organizations, and in connection with that, information becomes the most important resource. Knowledge management is the process through which organizations generate value from their intellectual and knowledge-based assets. It is a management philosophy, which combines good practice in purposeful information management with a culture of organizational learning, in order to improve the business performance. The computers are able to collect and select the information that can make some statistics, and so on but decisions have to make managers basing on their experience and taking into consideration computer support. An improvement in decision-making process is possible to be assured by analytical process supporting. Applying some analytical techniques, such as computer simulation, expert systems, and genetic algorithms, can improve the quality of managerial information.

One of the conference aims is also to consider an impact of the knowledge, information, computing, and the communication managing technologies of the organization functionality scope as well as the enterprise information systems design, implementation, and maintenance processes taking into the account various methodological, technological, and the technical aspects. It is also devoted to the information systems concepts and applications supporting exchange of goods and services by using different business models and exploiting opportunities offered by Internet-based electronic business and commerce solutions.

ISAT is a forum for specialized disciplinary research, as well as on interdisciplinary studies that aim to present original contributions and to discuss different subjects of today's information systems planning, designing, development, and implementation. The event is addressed to the scientific community, people involved in variety of topics related to information, management, computer, and communication systems and to people involved in the development of business information systems and business computer applications.

This year, we received more than 110 papers form about 10 countries. Each paper was reviewed by at least two members of Program Committee or independent reviewers. Only 86 best papers were selected for oral presentation and publication in the 37th International Conference on Information Systems Architecture and Technology proceedings.

This book is divided into four volumes, which splits papers into areas: Managing Complex Planning Environments, Systems Analysis and Modeling, Modeling of financial and Investment decisions, Risk Management, Project Management, Logistics and Market, Artificial Intelligence, Knowledge-Based Management, Web Systems, Computer Networks and Distributed Computing, High Performance Computing, Cloud Computing, Multi-agent Systems, Internet of Things, Mobile Systems, Service-Oriented Architecture Systems, Knowledge Discovery and Data Mining, Quality of Service, and E-Business Systems.

We would like to thank the Program Committee and external reviewers, who were essential for reviewing the papers and ensuring a high standard of the ISAT 2016 conference and its proceedings. We thank the authors, presenters, and participants of ISAT 2016, without them the conference could not have taken place. Finally, we thank the organizing team for their efforts during this and previous years which have led to a successful conclusion of the conference.

Wrocław, Poland
September 2016

Leszek Borzemski
Adam Grzech
Jerzy Świątek
Zofia Wilimowska

ISAT 2016 Conference Organization

General Chair

Zofia Wilimowska, Poland

Program Co-chairs

Zofia Wilimowska, Poland
Leszek Borzemski, Poland
Adam Grzech, Poland
Jerzy Świątek, Poland

Local Organizing Committee

Zofia Wilimowska, Chair
Leszek Borzemski, Vice-Chair
Adam Grzech, Vice-Chair
Jerzy Świątek, Vice-Chair
Arkadiusz Górski, Technical Editor, Conference Secretary
Anna Czarnecka, Technical Editor, Website Administrator
Agnieszka Parkitna, Conference Secretary
Anna Kamińska, Conference Secretary
Michał Kowalski, Technical Coordinator
Ziemowit Nowak, Website Support
Mariusz Fraś, Website Support

International Program Committee

Witold Abramowicz, Poland
Dhiya Al-Jumeily, UK
Iosif Androulidakis, Greece

Patricia Anthony, New Zeland
Zbigniew Banaszak, Poland
Elena Benderskaya, Russia
Leszek Borzemski, Poland
Janos Botzheim, Japan
Patrice Boursier, France
Wojciech Cellary, Poland
Haruna Chiroma, Malaysia
Edward Chlebus, Poland
Gloria Crisan, Romania
Marilia Curado, Portugal
Zhaohong Deng, China
Małgorzata Dolińska, Poland
Ewa Dudek-Dyduch, Poland
El-Sayed El-Alfy, Saudi Arabia
Naoki Fukuta, Japan
Piotr Gawkowski, Poland
Manuel Graña, Spain
Wiesław Grudzewski, Poland
Adam Grzech, Poland
Katsuhiro Honda, Japan
Marian Hopej, Poland
Zbigniew Huzar, Poland
Natthakan Iam-On, Thailand
Biju Issac, UK
Arun Iyengar, USA
Jürgen Jasperneite, Germany
Janusz Kacprzyk, Poland
Henryk Kaproń, Poland
Yannis L. Karnavas, Greece
Ryszard Knosala, Poland
Zdzisław Kowalczuk, Poland
Binod Kumar, India
Jan Kwiatkowski, Poland
Antonio Latorre, Spain
Gang Li, Australia
José M. Merigó Lindahl, Chile
Jose M. Luna, Spain
Emilio Luque, Spain
Sofian Maabout, France
Zygmunt Mazur, Poland
Pedro Medeiros, Portugal
Toshiro Minami, Japan
Marian Molasy, Poland
Zbigniew Nahorski, Poland

Kazumi Nakamatsu, Japan
Peter Nielsen, Denmark
Tadashi Nomoto, Japan
Cezary Orłowski, Poland
Sandeep Pachpande, India
Michele Pagano, Italy
George Papakostas, Greece
Zdzisław Papir, Poland
Marek Pawlak, Poland
Jan Platoš, Czech Republic
Tomasz Popławski – Poland
Edward Radosiński, Poland
Dolores I. Rexachs, Spain
José S. Reyes, Spain
Leszek Rutkowski, Poland
Gerald Schaefer, UK
Habib Shah, Malaysia
Jeng Shyang, Taiwan
Anna Sikora, Spain
Małgorzata Sterna, Poland
Janusz Stokłosa, Poland
Remo Suppi, Spain
Edward Szczerbicki, Australia
Jerzy Świątek, Poland
Eugeniusz Toczyłowski, Poland
Elpida Tzafestas, Greece
José R. Villar, Spain
Bay Vo, Vietnam
Hongzhi Wang, China
Leon S.I. Wang, Taiwan
Jan Werewka, Poland
Thomas Wielicki, USA
Zofia Wilimowska, Poland
Bernd Wolfinger, Germany
Józef Woźniak, Poland
Roman Wyrzykowski, Poland
Jaroslav Zendulka, Czech Republic
Bernard Ženko, Slovenia

ISAT 2016 Reviewers

Patricia Anthony, New Zealand
Zbigniew Antoni Banaszak, Poland
Elena Benderskaya, Russian Federation

Grzegorz Bocewicz, Poland
Leszek Borzemski, Poland
Jerzy Brzeziński, Poland
Wojciech Cellary, Poland
Krzysztof Cetnarowicz, Poland
Haruna Chiroma, Malaysia
Witold Chmielarz, Poland
Grzegorz Chodak, Poland
Robert Ryszard Chodorek, Poland
Kazimierz Choroś, Poland
Andrzej Chydziński, Poland
Gloria Cerasela Crisan, Romania
Mariusz Czekała, Poland
Pedro D. Medeiros, Portugal
Aldona Dereń, Poland
Grzegorz Dobrowolski, Poland
Ewa Dudek-Dyduch, Poland
Mariusz Fraś, Poland
Naoki Fukuta, Japan
Krzysztof Goczyła, Poland
Arkadiusz Górski, Poland
Manuel Grana, Spain
Jerzy Grobelny, Poland
Adam Grzech, Poland
Bogumila Hnatkowska, Poland
Maciej Hojda, Poland
Zbigniew Huzar, Poland
Natthakan Iam-On, Thailand
Przemysław Ignaciuk, Poland
Jerzy Józefczyk, Poland
Krzysztof Juszczyszyn, Poland
Adam Kasperski, Poland
Włodzimierz Kasprzak, Poland
Grzegorz Kołaczek, Poland
Zdzislaw Kowalczuk, Poland
Andrzej Kozik, Poland
Dorota Kuchta, Poland
Lumír Kulhánek, Czech Republic
Halina Kwaśnicka, Poland
Jan Kwiatkowski, Poland
Wojciech Lorkiewicz, Poland
Jose M. Luna, Spain
Lech Madeyski, Poland
Zbigniew Malara, Poland
Rafał Michalski, Poland

Zbigniew Nahorski, Poland
Jerzy Nawrocki, Poland
Peter Nielsen, Denmark
Tadashi Nomoto, Japan
Andrzej Nowak, Poland
Krzysztof Nowicki, Poland
Cezary Orłowski, Poland
Donat Orski, Poland
Piotr Pacyna, Poland
Michele Pagano, Italy
Agnieszka Parkitna, Poland
Marek Pawlak, Poland
Willy Picard, Poland
Jan Platoš, Czech Republic
Łukasz Popławski, Poland
Dolores Rexachs, Spain
Radosław Rudek, Poland
Jarogniew Rykowski, Poland
José Santos, Spain
Danuta Seretna-Sałamaj, Poland
Anna Sikora, Poland
Maciej Stasiak, Poland
Malgorzata Sterna, Poland
Janusz Stokłosa, Poland
Grażyna Suchacka, Poland
Remo Suppi, Spain
Joanna Szczepańska, Poland
Edward Szczerbicki, Poland
Paweł Świątek, Poland
Jerzy Świątek, Poland
Halina Tarasiuk, Poland
Kamila Urbańska, Poland
José R. Villar, Spain
Krzysztof Walczak, Poland
Zofia Wilimowska, Poland
Marek Wilimowski, Poland
Bernd Wolfinger, Germany
Jozef Woźniak, Poland
Roman Wyrzykowski, Poland
Jaroslav Zendulka, Czech Republic
Bernard Ženko, Slovenia
Maciej Zięba, Poland

Contents

Part I
Artificial Intelligence Methods

Identifying Project Alternatives with the Use of Constraint Programming

Marcin Relich

Abstract The paper is concerned with using a constraint programming paradigm for identifying alternative variants of completing a new product development project. The model of project prototyping includes variables and constraints in the fields of product, organisation and its environment. The project prototyping problem is formulated as a constraint satisfaction problem and implemented with the use of constraint programming that enables declarative description and effective solving of large combinatorial problems. Consequently, constraint programming can be considered as an appropriate framework for developing decision-making software to support identification of alternatives for project completion. An example illustrates the applicability of the proposed approach in the context of new product development projects.

Keywords Project management · New product development · Constraint satisfaction problem · Decision support system

1 Introduction

New products development (NPD) is one of the most important activities in today's companies that are located in a highly competitive environment. Increasing competition and customers' requirements impose more frequent product launches within a research and development (R&D) budget, target time and reliability. Launching a product before competitors and customer satisfaction with appropriate reliability are prerequisites for the product success. On the other hand, ensuring high reliability, especially in the complex products, is costly and time-consuming task. Moreover, companies often develop a few new products simultaneously and have to share limited resources between the NPD projects. If the company's resources (e.g. financial, human) are not sufficient to develop further the NPD

M. Relich (✉)
Faculty of Economics and Management, University of Zielona Gora, Zielona Gora, Poland
e-mail: m.relich@wez.uz.zgora.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_1

projects according to a project schedule, then it seems to be desirable to identify alternative variants of project completion. The process of identifying project alternatives can divide into two phases. The first phase is concerned with selecting factors that impact the R&D project parameters (e.g. cost, time, product reliability), identifying the relationships and forecasting project performance. In the second phase, the identified relationships are used to find alternative variants of project completion within the specified constraints.

The critical success factors in the NPD process can be considered from the perspective of a product, organisation and its environment, and they can refer to technical success of a product, on time project, cross-functional and experienced project team, top management commitment, sufficient project resources (including the number of project team members), proactive orientation in acquiring new technologies and identifying customer needs, business/financial analysis [1–3]. The empirical studies indicate that the size of a project team and the project budget are the important structural variables with potential influences on the quality of the NPD process and project success [4]. An investigation of new product forecasting practices in industrial companies shows that among popular techniques are customer/market research, looks-like analysis, trend line analysis, moving average, scenario analysis, and multi-attribute models [5, 6]. In a less extent, companies use forecasting techniques related to nonlinear regression, expert systems, neural and fuzzy neural networks [5]. This can result from the difficulties with designing and learning structures based on heuristic algorithms, and obtaining forecasting model in the clearly interpreted form. A decision support system (DSS) may help the project manager to obtain forecasting model on the basis of a knowledge base that consists of facts and rules.

A knowledge base of DSS can be considered from the perspective of constraints (e.g. the company's resources) and the identified relationships between variables (e.g. project parameters). Specification of constraints and variables permits to formulate a problem of identifying the alternative variants of project completion in terms of a constraint satisfaction problem (CSP). CSP that generally belongs to combinatorial problems is solved with the use of constraint programming (CP) [7, 8]. CP includes consistency techniques and systematic search strategies that are crucial for improving search efficiency of solving a problem [8, 9]. Numerous studies have employed CP to solve planning and scheduling problem in the various areas. For instance, the CP approach has been used to production scheduling in order to improve resource utilisation [10], resource portfolio planning of make-to-stock products for the purpose of maximising the long-term profit [11], computationally effective project-driven manufacturing in the multi-project environment [12], or project selection and scheduling problems with time-dependent resource constraints [8]. Identification of the alternative variants of project completion requires resource allocation within the specified time period, and it can be referred to a scheduling problem. Although the use of CP to project selection and scheduling problems has been widely considered in the literature, an aspect of identifying alternative variants aimed at rescuing the risky NPD projects is still unconsidered.

The paper is organised as follows: Sect. 2 presents problem formulation of a NPD project prototyping in terms of CSP. A method for identifying alternative variants of a NPD project is shown in Sect. 3. An illustrative example of the proposed approach is presented in Sect. 4. Finally conclusions and future research are presented in Sect. 5.

2 Problem Formulation of a NPD Project Prototyping in Terms of CSP

New product development depends on factors deriving from the business environment and company resources. For example, the customers' needs are an incentive for developing a new product that is designing, manufacturing and testing according to limitations such as the R&D budget, number of employees, cost of materials, environmental and product safety regulations, available technology, etc. Moreover, price of a new product is limited by price of competitors' products. To sum up, new product development depends on many factors that can change over time. The proposed approach aims to present these factors and limitations as variables and constraints in terms of a constraint satisfaction problem, and design a decision support system that help managers in identifying alternative variants of NPD project completion.

The project prototyping problem has been formulated in terms of a constraint satisfaction problem (CSP) which structure may be described as follows [12, 13]:

$$\mathrm{CSP} = ((\mathrm{X}, \mathrm{D}), \mathrm{C})$$

where $\mathrm{X} = \{x_1, x_2, \ldots, x_n\}$—a finite set of n variables, $\mathrm{D} = \{d_1, d_2, \ldots, d_n\}$—a finite set of n discrete domains of variables, and $\mathrm{C} = \{c_1, c_2, \ldots, c_k\}$—a finite set of k constraints limiting and linking variables.

A reference model of new product development can be presented as a hierarchical structure that includes the business environment, organisation and product (Fig. 1).

In the context of the presented reference model, a set of variables X includes variables that describe the characteristics of a new product, organisation and its environment. For instance, a set of variables X consists of the number of the total elements in a product, the number of the new elements in a product, the number of project team members, the number of customer requirements designed in a product, or cost of materials to build prototypes of a product. In turn, a set of constraints C includes the limits of accessible resources in a company (e.g. the number of R&D employees, the R&D budget), the desired time of project completion, the desired product reliability, and the relationships between the variables.

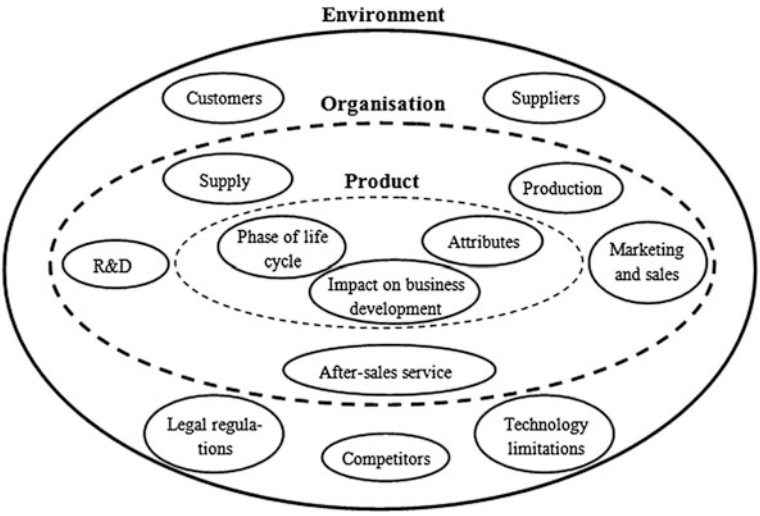

Fig. 1 Reference model of new product development

A company develop a portfolio of I new products P = {P1, …, Pi, …, PI}. In this study, project performance is evaluated according to the following criteria: the R&D budget assigned to i-th project Bi, the target time of project completion Ti and the desired product reliability Ri. In the project prototyping problem, the answers to two types of questions are sought:

1. Does project performance fulfil the project requirements?
2. What values should have variables to fulfil the project requirements?

The answer to first question requires the build of a forecasting model and its use to evaluate project performance over time. In turn, the answer to second question requires identification of a set of alternative variants (if there are any) for the NPD project completion. These two classes of questions refer to forecasting and diagnosing tasks, respectively. In this study, constraint programming technique is chosen to identify these alternative variants of project completion.

The model described in terms of CSP integrates technical parameters, available resources, expert experience, identified relationships, and user requirements in the form of knowledge base. Knowledge base is a platform for query formulation and for obtaining answers, and it comprises of facts and rules that are relevant to the system's properties and the relations between its different parts [14]. Consequently, model formulation in terms of CSP enables the design of a decision support system taking into account the available specifications and routine queries. The method of developing a decision support system for evaluating project performance and identifying the alternative variants of the NPD project completion is presented in the next section.

3 Method for Identifying Alternative Variants of a NPD Project

The proposed method refers to forecasting and diagnosing tasks according to two types of the above-presented questions. There are built forecasting models for the NPD project cost and time, and product reliability. If forecasts indicate an overrun of constraint (e.g. target cost), and there are no possibilities to increase the R&D budget, then alternative variants of the NPD project completion are sought. Figure 2 presents an example illustrating the concept of the proposed approach. The actual project performance at time t is distinguished by solid line, whereas trajectories for an original and alternative variant of project completion are distinguished by dotted lines. A set of alternative variants depends on the number and domain of decision variables (e.g. the number of elements in a product) that are considered to the modification.

Evaluation of project performance is carried out with the use of the relationships identified by forecasting models. These relationships are stored in a knowledge base as if-then rules. In this study, two forecasting models are considered. The first model bases on an econometric methodology, whereas the second model uses computational intelligence (fuzzy-neural network—FNN). FNN has the advantages of both neural networks (e.g. learning abilities, optimization abilities and connectionist structures) and fuzzy systems (e.g. if-then reasoning, simplicity of incorporating expert knowledge) [15, 16]. Moreover, FNN is able to identify the nonlinear and complex relationships (if there are any), and formulate them as if-then rules, what is an incentive to use FNN in this study.

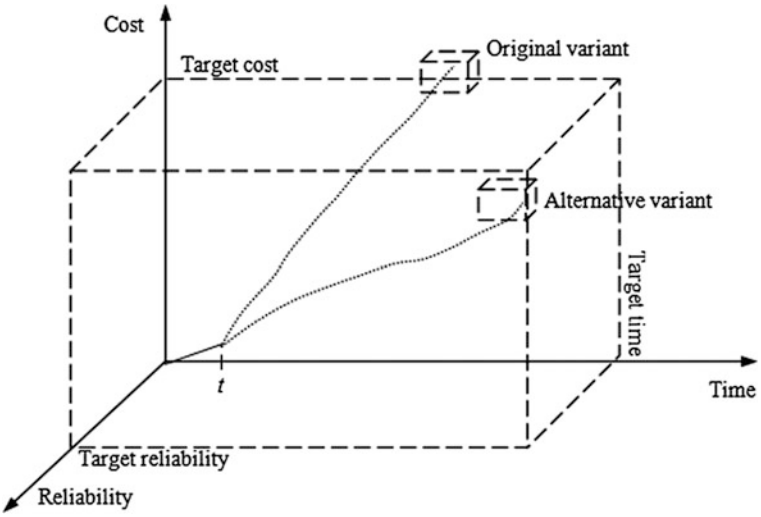

Fig. 2 Trajectories for original and alternative variant of project completion

Except the identified relationships, the knowledge base also includes rules and facts acquired by an expert from an enterprise system (e.g. financial and human resources). These rules and facts are used to evaluate the project cost, time, and reliability of a new product, and then to identify values of decision variables that ensure the desired project performance. The user interface provides the project manager a set of admissible variants of project completion according to user's preferences of the modified variables, and the evaluation of the NPD project performance.

The assumed model enables declarative approach to the problem statement, encompasses CSP framework and then it enables implementation of the considered problem in the constraint programming (CP) environment. CP is an emergent software technology for declarative description of CSP and can be considered as a pertinent framework for development of a decision support system. The CP techniques cope with CSPs with the use of paradigms such as propagate-and-search or propagate-and-distribute. CP has embedded ways to solve constraints satisfaction problems with greatly reduction of the amount of search needed [17]. Consequently, the use of CP can be significantly reduced the processing time of calculations in the case of extensive search space, for instance in scheduling problem [8, 10, 12]. CP is qualitatively different from the other programming paradigms, in terms of declarative, object-oriented, and concurrent programming. Compared to these paradigms, constraint programming is much closer to the ideal of declarative programming: to say what we want without saying how to achieve it [17]. As a result, the problem specification is closer to the original problem, obtaining solutions that are unavailable with imperative programming.

4 Illustrative Example

An example consists of two parts: the first part presents the use of a fuzzy neural network to estimating cost of product development, and the second part presents the use of CP to identify alternative variants of project completion.

4.1 Estimating Cost of Product Development

Three variables have chosen as independent variables in forecasting model of the NPD cost: number of elements in a new product (X_1), number of new elements in a product (X_2), and number of project team members (X_3). The data has been acquired from an enterprise information system and preprocessed before identifying relationships and forecasting, with the use of the principal component analysis. The data set has been divided into learning (P1–P23) and testing set (P24–P29) in order to eliminate the overtraining of FNN (too strict function adjustment to data). Cost estimation has been carried out for a forecasting model that provides the least root

mean square error (RMSE) in the testing set. To compare the forecasting quality, the RMSE for FNN, linear regression model and average has been identified.

FNN has been trained according to subtractive clustering method implemented in Matlab® software, with the following parameters: range of influence (RI) from 0.1 to 1.5, squash factor—1.25, accept ratio—0.5, reject ratio—0.15. Table 1 presents the RMSE and the number of rules for the different forecasting models of the NPD cost in the context of learning set (LS) and testing set (TS).

FNN has obtained in learning set less RMSEs than the average and linear regression model. FNN also in testing set outperforms the results of the average and linear regression model except one FNN model with parameter RI equals 0.6. This indicates on careful adjustment of parameters for learning FNN, what can be seen as the drawback of using FNN. The least RMSE in testing set has been obtained for FNN with RI equals 0.9, for which two rules have been identified. Figure 3 presents the use of FNN (RI = 0.9) for estimating the cost of NPD project with the following values of input variables: X1 = 57, X2 = 10, X3 = 3.

Cost forecast of the NPD project (75.1 monetary unit—m.u.) can be extended towards sensitivity analysis to identify the cost changes depending on the values of input variables. Figure 4 illustrates estimation of the NPD project cost for the number of elements in a new product from 50 to 60, the number of new elements in a product from 8 to 14, and 3 project members (at the left) and 4 project members (at the right).

A cost sensitivity analysis indicates the cost increase of 0.68 m.u. for each additional element in a product, and the cost increase of 4.16 m.u. for each additional new element in a product. Moreover, each additional member of project team increases the cost of new product development on the average 3.33 m.u. The presented cost estimation and sensitivity analysis is conducted for each NPD project to propose the optimal resource allocation in time span. The identified relationships between input variables and output variable (cost, time, reliability) are the basis for seeking the admissible variants for project completion.

Table 1 Comparison of forecasting models for the NPD cost

Model	RMSE in LS	RMSE in TS	Number of rules
FNN, RI = 0.1	1.339	1.902	20
FNN, RI = 0.2	1.339	2.823	11
FNN, RI = 0.3	1.339	2.381	6
FNN, RI = 0.4	1.339	2.846	6
FNN, RI = 0.5	1.390	3.038	4
FNN, RI = 0.6	1.373	8.344	4
FNN, RI = 0.7	1.405	2.390	3
FNN, RI = 0.8	1.471	1.806	2
FNN, RI = 0.9	1.487	1.783	2
FNN, RI = 1	1.491	1.797	2
FNN, RI = 1.5	1.505	1.916	2
Linear regression	2.890	3.186	1
Average	15.286	17.131	1

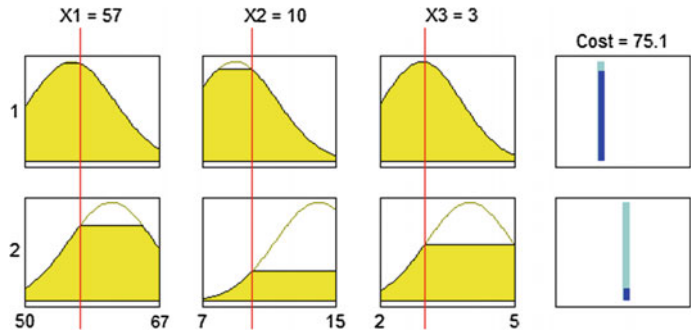

Fig. 3 Estimating the cost of NPD project with the use of FNN

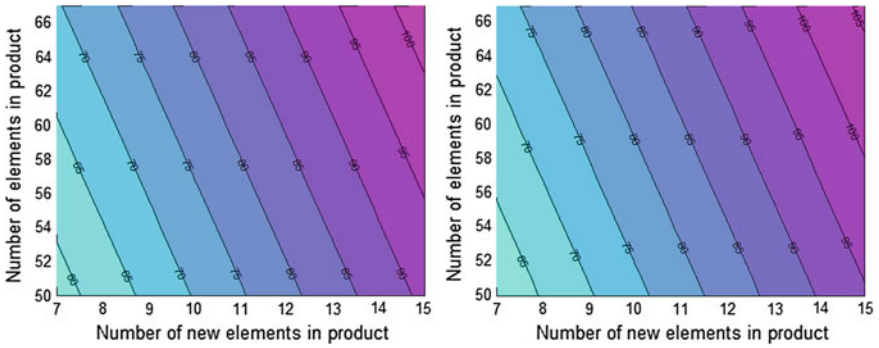

Fig. 4 Cost of NPD project depending on changes in X_1, X_2, X_3

4.2 Identifying Alternative Variants for Project Completion

Let us assume that a company develops three NPD projects {P1, P2, P3} which should be completed in time span (T) of 30 weeks and within the budget (B) of 210 m.u. The minimal product reliability is determined at 500 cycles of product use to the first technical defect. The additional constraint refers to 8 employees that can be assigned to three project teams. The cost estimation indicates the following values for the considered NDP projects: P1—75.1, P1—49.3, P1—96.7. The total planned cost of these NPD project reaches 221.1 m.u. and exceeds the project budget. In this case, the answer to the following question is sought: are there any values of input variables that fulfil project constraints? Searching the answer to this question is related to range specification of the values for input variables. A large number of admissible solutions imposes the use of techniques that enable the reduction of the amount of search needed, such as constraint programming. The considered problem has been implemented in the Oz Mozart programming environment that includes the CP paradigms.

To compare constraint programming with exhaustive search (ES) three cases have been distinguished. The cases differ in the number of admissible solutions that results from the range of the values for input variables. The first case has the following ranges of the values for input variables: P1 {X1 = 56#57, X2 = 9#10, X3 = 2#3}, P2 {X1 = 54#55, X2 = 7#8, X3 = 2#3}, P3 {X1 = 60#61, X2 = 10#11, X3 = 2#3}. The ranges of the values for input variables for the second case is as follows: P1 {X1 = 54#57, X2 = 9#10, X3 = 2#3}, P2 {X1 = 52#55, X2 = 7#8, X3 = 2#3}, P3 {X1 = 58#61, X2 = 10#11, X3 = 2#3}. In turn, the third case has the following ranges of the values for input variables: P1 {X1 = 52#57, X2 = 9#10, X3 = 2#3}, P2 {X1 = 50#55, X2 = 7#8, X3 = 2#3}, P3 {X1 = 56#61, X2 = 10#11, X3 = 2#3}.

Table 2 presents the results of seeking admissible solution for three above-described cases according to the different strategies of variable distribution. The results are compared in the context of the number of solutions, depth and time needed to find a solution. The calculations have been tested on an AMD Turion(tm) II Ultra Dual-Core M600 2.40 GHz, RAM 2 GB platform.

The results show that the use of constraint programming reduces computational time, what is especially important by a larger number of possible solutions. The user can obtain the entire set of solutions or one optimal solution (e.g. for minimal total planned cost of project portfolio). For example, in the first case presented in Table 2, the user can obtain all admissible solutions (10 instances) or one optimal solution. For the maximal reduced number of elements in a new product, the optimal solution indicates a decrease of the total planned cost of project portfolio at 207 m.u. The use of constraint programming enables the declarative description of the problem, and the strategies such as constraint propagation and variable distribution permit to reduce significantly a set of admissible solutions and the average computational time.

Table 2 Comparison of strategies for variable distribution

Case	Distribution strategy	Number of solutions	Depth	Time (s)
1	ES	512	18	0.79
	CP Naïve	10	5	0.12
	CP First-fail	10	5	0.13
	CP Split	10	5	0.10
2	ES	4096	24	13.14
	CP Naïve	308	14	1.07
	CP First-fail	308	16	0.86
	CP Split	308	13	0.82
3	ES	13824	30	124.15
	CP Naïve	2253	20	5.17
	CP First-fail	2253	22	4.92
	CP Split	2253	16	4.40

5 Conclusion

Successful development of a new product significantly impacts the company's profits and survival on a competitive market. Moreover, shorter product life cycles cause shorter time for new product development, what by the limited resources requires more effort and attention to manage the NPD projects. Therefore, it seems to be important to develop a decision support system aimed at rescuing the risky projects through identifying alternative variants of project completion. An effective implementation, development, and updating of a decision support system requires formulating a single knowledge base that reflects a reference model of NPD project prototyping. The proposed reference model encompasses the field of product, company and its environment. These fields can be described in terms of a constraint satisfaction problem that includes the sets of decision variables, their domains, and constraints, which link and limit the variables. The hierarchical and open structure of model enables solution of the decision problems with different level of specificity. The decision problems can include a query about the projected results of the proposed decisions, and about the conditions ensuring the expected results.

Characteristics of the proposed approach include the use of constraint programming to implement the constraint satisfaction problem. The results show that the use of constraint programming environment improves search efficiency of solving the considered problem, especially for a larger number of admissible solutions. Moreover, this study presents the use of fuzzy neural network to identify the relationships that are the basis for forecasting of the NPD project cost. The identified relationships are stored in a knowledge base in the form of if-then rules and used to generate a set of alternative variants of project completion. If project performance according to original specification is not possible, then the identified project alternatives can support the managers in making a decision about modification of a NPD project towards the fulfilment of constraints (e.g. the R&D budget) or abandonment of a project that indicates high risk of failure. Drawbacks of the proposed approach can be considered from the perspective of collecting enough amounts of data of the past similar NPD projects, and specifying several parameters to build and learn a fuzzy neural network. Further research can focus on reformulating the project prototyping problem towards specifying some decision variables in an imprecise form.

References

1. Ernst, H.: Success factors of new product development: a review of the empirical literature. Intl. J. Manag. Rev. **4**, 1–40 (2002)
2. Cooper, R., Edgett, S.: Maximizing productivity in product innovation. Res. Technol. Manag. **51**(2), 47–58 (2008)

3. Relich, M., Bzdyra, K.: Knowledge discovery in enterprise databases for forecasting new product success. In: Jackowski, K., et al. (eds.) IDEAL 2015. LNCS, vol. 9375, pp. 121–129. Springer (2015)
4. Ernst, H., Hoyer, W., Rubsaamen, C.: Sales, Marketing, and research-and-development cooperation across new product development stages: implications for success. J. Mark. **74**, 80–92 (2010)
5. Kahn, K.B.: An exploratory investigation of new product forecasting practices. J. Prod. Innov. Manag. **19**, 133–143 (2002)
6. Ozer, M.: A survey of new product evaluation models. J. Prod. Innov. Manag. **16**, 77–94 (1999)
7. Fruhwirth, T.W., Abdennadher, S.: Essentials of Constraint Programming. Springer, Berlin (2003)
8. Liu, S.S., Wang, C.J.: Optimizing project selection and scheduling problems with time-dependent resource constraints. Autom. Constr. **20**, 1110–1119 (2011)
9. Apt, K.R.: Principles of Constraint Programming. Cambridge University Press, Cambridge (2003)
10. Chan, W.T., Hu, H.: Constraint programming approach to precast production scheduling. J. Constr. Eng. Manag. **128**, 513–521 (2002)
11. Wang, S.M., Chen, J.C., Wang, K.J.: Resource portfolio planning of make-to-stock products using a constraint programming-based genetic algorithm. Omega **35**, 237–246 (2007)
12. Banaszak, Z., Zaremba, M., Muszyński, W.: Constraint programming for project-driven manufacturing. Int. J. Prod. Econ. **120**, 463–475 (2009)
13. Rossi, F., Van Beek, P. Walsh, T.: Handbook of Constraint Programming. Elsevier Science, Amsterdam (2006)
14. Bocewicz, G., Nielsen, I., Banaszak, Z.: Automated guided vehicles fleet match-up scheduling with production flow constraints. Eng. Appl. Artif. Intell. **30**, 49–62 (2014)
15. Azar A.T.: Adaptive Neuro-fuzzy Systems. In: Fuzzy systems. InTech (2010)
16. Relich, M.: Identifying relationships between eco-innovation and product success. In: Golinska, P., Kawa, A. (eds.) Technology Management for Sustainable Production and Logistics, pp. 173–192. Springer, Berlin (2015)
17. Van Roy, P., Haridi, S.: Concepts, Techniques and Models of Computer Programming. Massachusetts Institute of Technology, Cambridge (2004)

A New Algorithm for Online Management of Fuzzy Rules Base for Nonlinear Modeling

Krystian Łapa

Abstract In this paper a new algorithm for online management of fuzzy rules base for nonlinear modeling is proposed. The online management problem is complex due to limitations of memory and time needed for calculations. The proposed algorithm allows an online creation and management of fuzzy rules base. It is distinguished, among the others, by mechanisms of: managing of number of fuzzy rules, managing of fuzzy rules weights and possibilities of background learning. The proposed algorithm was tested on typical nonlinear modeling problems.

Keywords Neuro-fuzzy system · Evolving fuzzy system · Background learning

1 Introduction

With the development of technology the amount of information (data) that can be stored and processed increases significantly. These data exceed the reach of commonly used hardware and software tools to capture, process and analyze it in an acceptable time. The typical systems have limitations due to limit of memory for storing data and time needed for iterative learning (processing of data to obtain best possible results of the system [3]). One of the solutions for this problem is online processing of the data [7]. In this solution the systems usually process incoming data, adapt to them (by modifying of its parameters) and lose information about processed data. Systems like that are classified as evolving systems (ES). The key element of processing data online is to find compromise between complexity and stability (possibilities of life-long learning - adapting of the system to new data at any point in time [15]).

The evolving systems include, among the others, the evolving fuzzy systems (EFS) [6]. The evolving fuzzy systems are based on fuzzy systems and fuzzy

K. Łapa (✉)
Institute of Computational Intelligence, Częstochowa University of Technology, Częstochowa, Poland
e-mail: krystian.lapa@iisi.pcz.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_2

'IF… THEN…' rules [10]. The fuzzy rules allow not only to obtain good accuracy in the field of modelling, classification and control but also are interpretable [2]. Designing of EFS involves step by step into adding and modification of fuzzy rules and fuzzy sets to the system in certain situations (for example when new incoming data sample does not activate any existing fuzzy rule). These methods have their origin in algorithms used for creation of fuzzy rule base on the basis of data samples but they needed adaptation for larger data sets while maintaining a simple structure (see e.g. [8, 9, 13]).

On the other hand, the algorithms for designing systems offline allow to obtain better accuracy. Moreover they can work on populations of the systems to avoid local optima (see e.g. [5, 10]). These algorithms are very efficient on smaller data sets and have problems with iterative processing of larger data.

In this paper a new approach for online modelling with background learning and using fuzzy rules weights is introduced. The online designing of the system is based on analyzing of firing level of the rules and using the values of the fuzzy sets membership functions. The proposed background learning method uses limited number of prepared samples obtained from processing of data samples. Moreover, the fuzzy rules weights are calculated on the basis of counters of using fuzzy rules in a designing process. It is worth to mention that, the purpose of this paper is not only to achieve best accuracy of the system, but also to show influence of background learning and weights.

This paper is divided as follows: in Sect. 2 the proposed system is described, in Sect. 3 an idea of the design of the proposed system is presented, in Sect. 4 simulation results are shown, whereas in Sect. 5 the conclusions are drawn.

2 Neuro-Fuzzy System for Online Nonlinear Modelling

The proposed approach utilizes a neuro-fuzzy system of Mamdani type [10]. This system was designed to allow online building and modifying of the fuzzy rules with automatic weights selection. The fuzzy rules are based on dynamical fuzzy sets base **C** defined as follows:

$$\mathbf{C} = \left\{ \begin{matrix} A_{1,1}, \ldots, A_{1,L_1^A}, \ldots, A_{n,1}, \ldots, A_{n,L_n^A}, \\ B_{1,1}, \ldots, B_{1,L_1^B}, \ldots, B_{m,1}, \ldots, B_{m,L_m^B} \end{matrix} \right\} = \{C_1, \ldots, C_{L^C}\}, \tag{1}$$

where $A_{i,l}$ stands for input fuzzy sets, $i = 1, \ldots, n$ stands for input index, n stands for number of inputs, $l = 1, \ldots, L_i^A$ stands for index of fuzzy set, L_i^A stands for number of input fuzzy sets from base (1) related to input i. $B_{j,l}$ stands for output fuzzy sets, $j = 1, \ldots, m$ stands for output index, m stands for number of outputs, $l = 1, \ldots, L_j^B$ stands for index of fuzzy set, L_j^B stands for number of output fuzzy sets from base (1) related to output j, $L^C = \sum_{i=1}^{n} L_i^A + \sum_{j=1}^{m} L_j^B$ stands for, changing in the learning process, total number of fuzzy sets. This approach allows

to work with elastic number of fuzzy sets. Each fuzzy set $A_{i,l}$ is represented by membership function $\mu_{A_{i,l}}(x)$, while each fuzzy set $B_{j,l}$ is represented by membership function $\mu_{B_{j,k}}(y)$. In the proposed approach a Gaussian-type membership functions were used. Therefore, for the fuzzy sets the following parameters were assigned:

$$A_{i,l} = \left\{x_{i,l}^A, \sigma_{i,l}^A, c_{i,l}^A\right\}; \quad B_{j,l} = \left\{y_{j,l}^B, \sigma_{j,l}^B, c_{j,l}^B\right\}, \tag{2}$$

where $x_{i,l}^A$ and $y_{j,l}^B$ stands for centers of fuzzy sets, $\sigma_{i,l}^A$ and $\sigma_{j,l}^B$ stands for widths of fuzzy sets, $c_{i,l}^A$ and $c_{j,l}^B$ stands for counters (treat as heaviness) of using fuzzy sets. Fuzzy set is "used" when fuzzy set membership function achieve high value for specified data sample.

The fuzzy rules base contains fuzzy rules $\mathbf{R}_k$, where $k = 1, \ldots, N$ stands for fuzzy rule index, N stands for actual number of fuzzy sets. The number of fuzzy rules can change in a learning process (see Sect. 3). In the proposed approach fuzzy rules are defined as follows:

$$\mathbf{R}_k = \left\{I_{1,k}^A, \ldots, I_{n,k}^A, I_{1,k}^B, \ldots, I_{m,k}^B, c_k^R\right\}, \tag{3}$$

where each index $I_{i,k}^A$ refers to one input fuzzy set from base $\mathbf{C}$ and each index $I_{j,k}^B$ refers to one output fuzzy set from base $\mathbf{C}$, c_k^R is fuzzy rule counter (treat as heaviness) of using fuzzy rule. Fuzzy rule is "used" when fuzzy rule firing level achieves high value for specified data sample. This approach allows sharing of single fuzzy sets by many fuzzy rules, which notation is defined as:

$$R_k|w_k\text{: IF} \begin{pmatrix} x_1 \text{ is } A_{1,I_{1,k}^A} \text{ AND} \\ x_2 \text{ is } A_{2,I_{2,k}^A} \text{ AND} \\ \cdots \\ x_n \text{ is } A_{n,I_{n,k}^A} \end{pmatrix} \quad \text{THEN} \begin{pmatrix} y_1 \text{ is } B_{1,I_{1,k}^B} \text{ AND} \\ y_2 \text{ is } B_{2,I_{2,k}^B} \text{ AND} \\ \cdots \\ y_m \text{ is } B_{m,I_{m,k}^B} \end{pmatrix}, \tag{4}$$

where w_k is a weight of k-th fuzzy rule calculated on the basis of fuzzy rules counters in a following way:

$$w_k = \alpha + (1 - \alpha) \cdot \left(\frac{c_k^R - \min\limits_{l=1,\ldots,N} \left\{c_l^R\right\}}{\max\limits_{l=1,\ldots,N} \left\{c_l^R\right\} - \min\limits_{l=1,\ldots,N} \left\{c_l^R\right\} + \xi} \right), \tag{5}$$

where α stands for parameter specifying minimum value of fuzzy rule weight ($w_k \in [\alpha, 1]$), $\xi > 0$ stands for very small value which prevents division by zero. Equation (5) assigns automatically higher values of weights for rules with higher value of fuzzy rule counters. It not only allows better interpretation of the rules (giving information which rule is more important) but also increases the possibilities of obtaining higher accuracy of the system.

The firing level (activation level) of fuzzy rule $\mathbf{R}_k$ is calculated as:

$$\tau_k(\bar{\mathbf{x}}) = \underset{i=1}{\overset{n}{T}}\left\{\mu_{A_{i,l_{i,k}^A}}(\bar{x}_i)\right\} = T\left\{\mu_{A_{1,l_{1,k}^A}}(\bar{x}_1), \ldots, \mu_{A_{n,l_{n,k}^A}}(\bar{x}_n)\right\}, \tag{6}$$

where $T(\cdot)$ is any triangular t-norm [10]. In case of singleton fuzzification the interferences from k-th rule are calculated independently for each j-th output using triangular t-norm (which is an interference operator in the Mamdani type of fuzzy system):

$$\mu_{\bar{B}_{j,k}}(\bar{\mathbf{x}}, y) = \mu_{\mathbf{A}_k \to B_{j,k}}(\bar{\mathbf{x}}, y) = T\left\{\tau_k(\bar{\mathbf{x}}), \mu_{B_{l_{j,k}^B}}(y)\right\}. \tag{7}$$

The aggregation of interference of fuzzy rules is calculated as follows:

$$\mu_{\bar{B}_{j,k}}(\bar{\mathbf{x}}, y) = \underset{k=1}{\overset{N}{S^*}}\left\{\mu_{\bar{B}_{j,k}}(\bar{\mathbf{x}}, y), w_k\right\} \tag{8}$$

where $S^*(\cdot)$ is a triangular t-conorm with weights of arguments. The defuzzificated values of fuzzy system of its j-th output can be calculated with (for example) the center of area method:

$$y_j(\bar{\mathbf{x}}) = \frac{\sum_{l=1}^{L_j^B} y_{j,l}^B \cdot \mu_{B'_j}\left(\bar{\mathbf{x}}, y_{j,l}^B\right)}{\sum_{l=1}^{L_j^B} \mu_{B'_j}\left(\bar{\mathbf{x}}, y_{j,l}^B\right)}, \tag{9}$$

where $y_{j,l}^B$ are values equal to maximum (isolated) points of the functions $\mu_{B_{j,k}}(y)$ (which are centers of used in simulations Gaussian-type fuzzy sets).

3 Description of Proposed Method

The proposed method is based on three mechanisms: preliminary analysis of data samples, building of fuzzy system and background learning. The main idea of proposed method is presented on Fig. 1. The proposed mechanisms are described in detail in the further part of this section.

3.1 Preliminary Analysis of Data Samples

The purpose of preliminary analysis of data samples is to estimate initialization values of widths of fuzzy sets. Each sample $\bar{\mathbf{x}}$ consist of n input signals and m output signals: $\bar{\mathbf{x}} = \{\bar{x}_1, \ldots, \bar{x}_n, \bar{x}_{n+1}, \ldots, \bar{x}_{n+m}\} = \{\bar{x}_1, \ldots, \bar{x}_n\}$, for which a

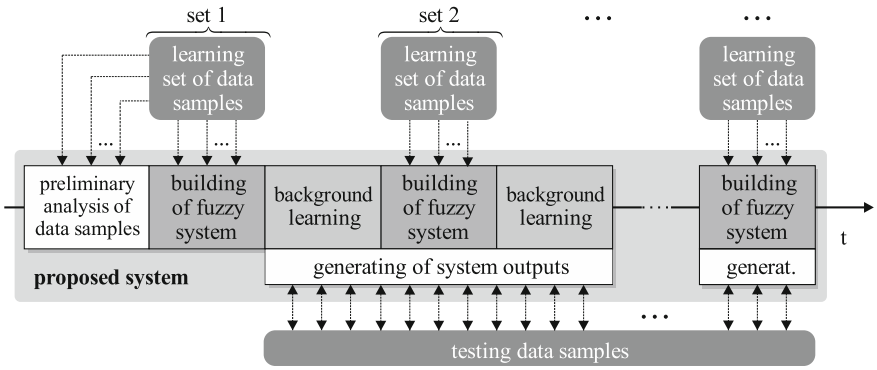

Fig. 1 The idea of proposed evolving fuzzy system

minims $\mathbf{x}^{\min}$ and maxims $\mathbf{x}^{\max}$ are determined. The minims and maxims allows to determine widths of fuzzy sets in the following way:

$$\sigma_i = \sigma^{\text{par}} \cdot \left(x_i^{\max} - x_i^{\min}\right), \tag{10}$$

where $\sigma^{\text{par}} \in [0, 1]$ stands for parameter specifying the initial width of fuzzy sets. The minimums and maximums are initially calculated only for first set of data samples, and they are updated with every further data sample.

3.2 Building of Fuzzy System

In the proposed method a new fuzzy rules are added into the system when actual data sample $\bar{\mathbf{x}}$ does not activate with (specified by parameter $\tau^{\text{akt}} \in [0, 1]$) level any of existed fuzzy rules in form of (4). This condition can be write as:

$$\tau^{\text{akt}} < \max_{k=1,\ldots,N} \left\{ \sqrt[n]{\tau_k(\bar{\mathbf{x}})} \right\}. \tag{11}$$

The use of the square root in the Eq. (11) reduces the impact of the system (9) inputs number on results of fuzzy rules activation. If the condition (11) is met the counter of the fuzzy rule c_k^R (where k stands for index of fuzzy rule with highest value of firing level) is increased and the parameters of fuzzy rule are modified as follows:

$$\begin{cases} x_{i,l}^A = \frac{c_{i,l}^A \cdot x_{i,l}^A + 1 \cdot \bar{x}_i}{c_{i,l}^A + 1}; & c_{i,l}^A = c_{i,l}^A + 1 \quad \text{for input fuzzy set } (i \le n) \\ y_{i-n,l}^B = \frac{c_{i-n,l}^B \cdot y_{i-n,l}^B + 1 \cdot \bar{x}_i}{c_{i-n,l}^B + 1}; & c_{i-n,l}^B = c_{i-n,l}^B + 1 \quad \text{for output fuzzy set } (i > n) \end{cases}, \tag{12}$$

Due to using heaviness of fuzzy sets in this modification, the insensitivity of changes slowly decreases. Thanks to that the fuzzy sets retain in a clear (interpretable) positions (they do not overlap each other). This idea is based on the idea of moving clusters from the Ward method [16].

If the condition (11) is not met a new fuzzy rule is added to the system (4). Newly created fuzzy rules can use both the existing in the base (1) and newly created fuzzy sets. To check if the fuzzy rule can use already existing fuzzy set for each input signal $\bar{x}_i$ from data sample $\bar{\mathbf{x}}$ the following condition is checked:

$$\begin{cases} \mu^{\text{akt}} > \max\limits_{l=1,\ldots,L_i^A} \left\{ \mu_{A_{h,l}}(\bar{x}_i) \right\} & \text{for input fuzzy set } (i \leq n) \\ \mu^{\text{akt}} > \max\limits_{l=1,\ldots,L_{i-n}^B} \left\{ \mu_{B_{h-n,l}}(\bar{x}_i) \right\} & \text{for output fuzzy set } (i > n) \end{cases}, \tag{13}$$

where $\mu^{\text{akt}} \in [0, 1]$ stands for threshold value specifying when existing fuzzy set might be used (it acts similar to a function parameter τ^{akt} of fuzzy rules). If the condition (13) is met, a fuzzy rules can use existing fuzzy set (with highest value of membership function), the parameters of fuzzy set are modified according to (12), and the index I_i^A or I_{i-n}^B of the fuzzy rule is set to index of this fuzzy set. In the other case, a new fuzzy set is inserted into fuzzy sets base (1) with parameters initialized as follows:

$$\begin{cases} x_{i,l}^A = \bar{x}_i; \sigma_{i,l}^A = \sigma_i; c_{i,l}^A = 1 & \text{for input fuzzy set } (i \leq n) \\ y_{j,l}^B = \bar{x}_{n+j}; \sigma_{j,l}^B = \sigma_{n+j}; c_{j,l}^B = 1 & \text{for output fuzzy set } (i > n) \end{cases}. \tag{14}$$

The building of fuzzy system is based on standard fuzzy system mechanisms: the analysis of firing (activation) of fuzzy rules and analysis of the values of fuzzy sets membership functions. The weights of fuzzy rules are calculated automatically on the basis of fuzz rules counters (heaviness). This approach is new in the literature.

3.3 Background Learning

EFS are, by default, systems which parameters cannot be tuned. It results from theoretically infinite number of incoming online data samples. However a method of background learning (tuning) of parameters of fuzzy sets is presented (as an additional option—see Fig. 1). It was achieved by creating and storing maximum of $R^{\max}$ auxiliary samples for each fuzzy rule. Each auxiliary sample of k-th rule is stored in the form of cluster as in Ward's method [16]. Each cluster is represented by centers $x_{h,k,d}^R$ and heaviness $c_{k,d}^R$ where $d = 1, \ldots, R^{\max}$ is an index of auxiliary sample and $h = 1, \ldots, n + m$.

The process of creating auxiliary samples is connected to the process of creating and modifying fuzzy rules. Each of incoming data samples becomes an auxiliary sample for those rules for with highest value of firing (activation) level:

$$x_{h,k,d}^{R} = \bar{x}_h; \quad c_{k,d}^{R} = 1. \tag{15}$$

When the number of auxiliary samples for specified fuzzy rule is higher than maximum number of data samples $R^{\max}$ then, two closest auxiliary samples are merged. The distance between auxiliary samples (with taking into account heaviness) is calculated as follows:

$$dist_{d1,d2} = \frac{c_{k,d1}^{R} \cdot c_{k,d2}^{R}}{c_{k,d1}^{R} + c_{k,d2}^{R}} \cdot \sum_{h=1}^{n+m} \left| \frac{x_{h,k,d1}^{R} - x_{h,k,d2}^{R}}{x_h^{\max} - x_h^{\min}} \right|, \tag{16}$$

where $d1$, $d2$ are indexes of two comparing auxiliary samples. Therefore, the number $R^{\max}$ cannot be very high. The merging of two closest auxiliary samples is performed as follows:

$$x_{h,k,d3}^{R} = \frac{c_{k,d1}^{R} \cdot x_{h,k,d1}^{R} + c_{k,d2}^{R} \cdot x_{h,k,d2}^{R}}{c_{k,d1}^{R} + c_{k,d2}^{R}}; \quad c_{k,d3}^{R} = c_{k,d1}^{R} + c_{k,d2}^{R}, \tag{17}$$

where $d3$ is an index of newly created auxiliary sample. For the background learning, a genetic algorithm (GA) [10] was used (however any other learning methods can be also used—see e.g. [11, 12]), which aims to minimize error obtained for all auxiliary samples in all rules (in the learning process the auxiliary samples are treated as normal learning data samples).

3.4 System Evaluation

The evaluation function for GA includes both the complexity and accuracy of the system (9). The complexity of system (9) is defined as follows:

$$CMPL = w^{\text{rule}} \cdot N + w^{\text{fset}} \cdot \left(\sum_{i=1}^{n} L_i^{\text{A}} + \sum_{j=1}^{m} L_j^{\text{B}} \right), \tag{18}$$

where $w^{\text{rule}} \in [0, 1]$ stands for weight of fuzzy rules (set experimentally to 1.0), $w^{\text{fset}} \in [0, 1]$ stands for weight of fuzzy sets (set experimentally to 0.5). The accuracy of the system (9) is determined by *RMSE*:

$$RMSE = \frac{1}{Z \cdot m} \sum_{i=1}^{Z} \sum_{j=1}^{m} \sqrt{\left(\bar{y}_j(\bar{\mathbf{x}}_z) - x_{z,n+j} \right)^2}. \tag{19}$$

4 Simulations

In this paper a four simulation cases were tested (see Table 1). In the case I the proposed mechanisms of background learning and using fuzzy rule weights were turned-off (for a comparison). In the case II the fuzzy rule weights were turned-off to compare influence of weights on the results (case III). The case IV purpose was to achieve systems with simpler structures (by changing values of system parameters σ^{par}, τ^{akt}, μ^{akt}—see Table 1).

In this paper a set of modelling problems was used (presented in Table 2). The used data sets were modified in a way to allow to test the background learning in an online creation of fuzzy system: (a) from data a 10 % randomly chosen samples are selected and delivered into fuzzy system, (b) those samples are used to create auxiliary samples, (c) the background learning mechanism is applied (in case II–IV), (d) the 100 % of data is used for testing the system (9). This test procedure was repeated 20 times. Typical EFS are tested with each simulation for each problem, and it was repeated 50 times and then results were averaged.

The simulation parameters σ^{par}, τ^{akt}, and μ^{akt} have a deciding impact on accuracy and complexity of the system and they were set experimentally according to Table 1 (for example an increase of τ^{akt} causes creating more fuzzy rules). The weight parameter α was set to 0.2. The background learning parameters were chosen experimentally: mutation probability $p_m = 0.15$, crossover probability $p_c = 0.75$, number of maximum auxiliary samples $R^{\max} = 3$. In equations that define firing level of rules (6), interference of rules (7) and aggregation of interference (8) and the product triangular norms were used.

Table 1 Considered simulation cases

Case	BL	Weights	σ^{par}	τ^{akt}	μ^{akt}	Case description
I	No	No	0.15	0.25	0.25	Typical EFS
II	Yes	No	0.15	0.25	0.25	Proposed EFS + BL
III	Yes	Yes	0.15	0.25	0.25	Proposed EFS + BL + rule weights
IV	Yes	Yes	0.18	0.20	0.20	Proposed EFS + lower complexity

BL stands for background learning

Table 2 List of used simulation problems

Problem id (#)	Problem name	Reference	Inputs	Outputs	Samples
1	Hang function	[14]	2	1	50
2	Chemical plant	[1]	3	1	70
3	Abalone	[17]	8	1	4178
4	Machine CPU	[4]	7	1	210

Table 3 Linguistic labels used for fuzzy rule weights

Weight value	Linguistic label	Definition
<0.4	*l*	Low important
$\in \langle 0.4, 0.7 \rangle$	*i*	Important
>0.7	*v*	Very important

Table 4 Summary results of *RMSE* and *CMPL* (best *CMPL* concern best *RMSE* result)

#	Case	Avg. *RMSE*	Avg. *CMPL*	Best *RMSE*	Best *CMPL*	Avg. *N*	Avg. fuzzy sets
1	I	0.3798	12.1868	0.1371	13.5000	7.33	9.70
	II	0.3215	12.7000	0.1221	16.0000	7.66	10.08
	III	**0.3040**	12.9052	0.1153	14.5000	7.81	10.19
	IV	0.3711	**11.1184**	**0.0946**	**13.0000**	**6.66**	**8.92**
2	I	0.0483	13.8392	0.0223	**13.5000**	6.98	**13.46**
	II	0.0383	14.0454	**0.0117**	15.0000	7.24	15.22
	III	**0.0379**	14.0846	0.0144	16.0000	7.25	16.53
	IV	0.0399	**12.5100**	0.0165	14.0000	**6.35**	14.89
3	I	3.0375	24.2604	1.7169	25.5000	10.30	27.91
	II	2.5806	25.7413	1.5007	27.0000	11.19	29.10
	III	**2.4931**	25.4470	1.5251	26.5000	10.97	28.95
	IV	2.6760	**21.4996**	**1.1781**	**22.5000**	**8.90**	**25.21**
4	I	95.3925	22.8983	36.8803	22.0000	9.46	26.88
	II	58.7915	23.3000	18.1972	24.0000	9.80	27.00
	III	**53.1230**	23.3774	**15.1279**	24.0000	9.84	27.08
	IV	64.9428	**20.7622**	17.3769	**21.0000**	**8.68**	**24.17**

4.1 Simulation Results

The results obtained for considered problems for all cases are shown in Table 4. The *RMSE* from all test phases is shown in Fig. 3. The examples of obtained fuzzy sets for case IV are shown in Fig. 2 and corresponding fuzzy rules are presented in Table 5 (fuzzy sets were replaced by their linguistic labels '*v.low*', '*low*', '*medium*', '*high*', '*v.high*'—see Fig. 2 and the weights values were replaced by linguistic labels '*l*', '*i*', '*v*'—see Table 3).

4.2 Simulations Conclusions

The conclusions from simulations can be summed up as follows: (a) the results obtained for the proposed approach are close (in a field of accuracy) to the results

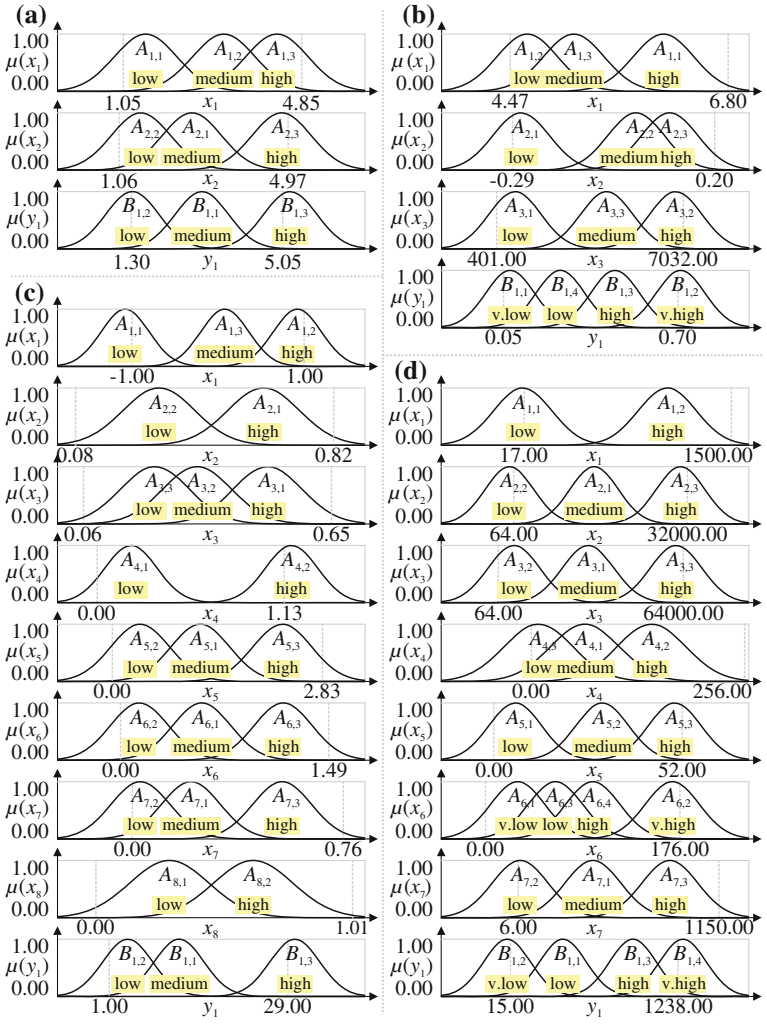

Fig. 2 The examples of fuzzy sets for considered problems: **a** #1, **b** #2, **c** #3, **d** #4. *Dotted lines* stands for minimums (*left line*) and maximums (*right line*) of learning data

obtained by other authors for offline learning and (for most cases) with use of not interpretable systems, (b) the proposed approach allowed to achieve interpretable fuzzy rules (see Fig. 2 and Table 5), (c) the background learning improved EFS by approximately 15–20 % (see Table 4), (d) the use of automatically calculated fuzzy rule weights improved EFS by additional approximately 5 %, (e) the case IV allows to obtain good accuracy with lower number of rules and lower number of fuzzy sets, (f) the method without background learning stops improving after processing $\sim$50 % of data sets (see test phase number 5 on Fig. 3).

Table 5 The example notations of fuzzy rules obtained for considered problems

#	Fuzzy rules notation
1	$\begin{cases} R_1\|v: \text{IF } (x_1 \text{ is low AND } x_2 \text{ is medium}) \text{ THEN } (y_1 \text{ is medium}) \\ R_2\|v: \text{IF } (x_1 \text{ is medium AND } x_2 \text{ is medium}) \text{ THEN } (y_1 \text{ is low}) \\ R_3\|i: \text{IF } (x_1 \text{ is high AND } x_2 \text{ is low}) \text{ THEN } (y_1 \text{ is medium}) \\ R_4\|i: \text{IF } (x_1 \text{ is low AND } x_2 \text{ is low}) \text{ THEN } (y_1 \text{ is high}) \\ R_5\|i: \text{IF } (x_1 \text{ is high AND } x_2 \text{ is high}) \text{ THEN } (y_1 \text{ is low}) \end{cases}$
2	$\begin{cases} R_1\|i: \text{IF } (x_1 \text{ is high AND } x_2 \text{ is low AND } x_3 \text{ is low}) \text{ THEN } (y_1 \text{ is v.low}) \\ R_2\|v: \text{IF } (x_1 \text{ is low AND } x_2 \text{ is medium AND } x_3 \text{ is high}) \text{ THEN } (y_1 \text{ is v.high}) \\ R_3\|i: \text{IF } (x_1 \text{ is medium AND } x_2 \text{ is low AND } x_3 \text{ is medium}) \text{ THEN } (y_1 \text{ is high}) \\ R_4\|i: \text{IF } (x_1 \text{ is high AND } x_2 \text{ is medium AND } x_3 \text{ is low}) \text{ THEN } (y_1 \text{ is v.low}) \\ R_5\|i: \text{IF } (x_1 \text{ is medium AND } x_2 \text{ is medium AND } x_3 \text{ is low}) \text{ THEN } (y_1 \text{ is low}) \\ R_6\|i: \text{IF } (x_1 \text{ is medium AND } x_2 \text{ is high AND } x_3 \text{ is medium}) \text{ THEN } (y_1 \text{ is high}) \end{cases}$
3	$\begin{cases} R_1\|v: \text{IF } \begin{pmatrix} x_1 \text{ is low AND } x_2 \text{ is high AND } x_3 \text{ is high AND } x_4 \text{ is low AND} \\ x_5 \text{is medium AND } x_6 \text{ is medium AND } x_7 \text{ is medium AND } x_8 \text{ is low} \end{pmatrix} \text{ THEN } \begin{pmatrix} y_1 \text{ is} \\ \text{medium} \end{pmatrix} \\ R_2\|v: \text{IF } \begin{pmatrix} x_1 \text{ is high AND } x_2 \text{ is high AND } x_3 \text{ is medium AND } x_4 \text{ is low AND} \\ x_5 \text{ is low AND } x_6 \text{ is low AND } x_7 \text{ is medium AND } x_8 \text{ is low} \end{pmatrix} \text{ THEN } \begin{pmatrix} y_1 \text{ is} \\ \text{medium} \end{pmatrix} \\ R_3\|v: \text{IF } \begin{pmatrix} x_1 \text{ is medium AND } x_2 \text{ is low AND } x_3 \text{ is low AND } x_4 \text{ is low AND} \\ x_5 \text{ is low AND } x_6 \text{ is low AND } x_7 \text{ is low AND } x_8 \text{ is low} \end{pmatrix} \text{ THEN } \begin{pmatrix} y_1 \text{ is} \\ \text{low} \end{pmatrix} \\ R_4\|i: \text{IF } \begin{pmatrix} x_1 \text{is low AND } x_2 \text{ is high AND } x_3 \text{ is high AND } x_4 \text{ is low AND} \\ x_5 \text{ is high AND } x_6 \text{ is high AND } x_7 \text{ is high AND } x_8 \text{ is high} \end{pmatrix} \text{ THEN } \begin{pmatrix} y_1 \text{ is} \\ \text{medium} \end{pmatrix} \\ R_5\|i: \text{IF } \begin{pmatrix} x_1 \text{ is high AND } x_2 \text{ is high AND } x_3 \text{ is high AND } x_4 \text{ is low AND} \\ x_5 \text{ is high AND } x_6 \text{ is high AND } x_7 \text{ is high AND } x_8 \text{ is high} \end{pmatrix} \text{ THEN } \begin{pmatrix} y_1 \text{ is} \\ \text{medium} \end{pmatrix} \\ R_6\|i: \text{IF } \begin{pmatrix} x_1 \text{is high AND } x_2 \text{ is high AND } x_3 \text{ is high AND } x_4 \text{ is low AND} \\ x_5 \text{is high AND } x_6 \text{ is mediumAND } x_7 \text{ is medium AND } x_8 \text{ ishigh} \end{pmatrix} \text{ THEN } \begin{pmatrix} y_1 \text{ is} \\ \text{high} \end{pmatrix} \\ R_7\|i: \text{IF } \begin{pmatrix} x_1 \text{ is high AND } x_2 \text{ is high AND } x_3 \text{ is medium AND } x_4 \text{ is high AND} \\ x_5 \text{ is low AND } x_6 \text{ is low AND } x_7 \text{ is low AND } x_8 \text{ is low} \end{pmatrix} \text{ THEN } \begin{pmatrix} y_1 \text{ is} \\ \text{medium} \end{pmatrix} \end{cases}$

(continued)

Table 5 (continued)

#	Fuzzy rules notation
4	$\begin{cases} R_1 \mid i: \text{IF} \begin{pmatrix} x_1 \text{ is low AND } x_2 \text{ is medium AND } x_3 \text{ is medium AND } x_4 \text{ is} \\ \text{medium AND } x_5 \text{ is low AND } x_6 \text{ is v.low AND } x_7 \text{ is medium} \end{pmatrix} \text{THEN} \begin{pmatrix} y_1 \text{ is} \\ \text{low} \end{pmatrix} \\ R_2 \mid v: \text{IF} \begin{pmatrix} x_1 \text{ is low AND } x_2 \text{ is low AND } x_3 \text{ is low AND } x_4 \text{ is medium} \\ \text{AND } x_5 \text{ is low AND } x_6 \text{ is v.low AND } x_7 \text{ is low} \end{pmatrix} \text{THEN} \begin{pmatrix} y_1 \text{ is} \\ \text{v.low} \end{pmatrix} \\ R_3 \mid i: \text{IF} \begin{pmatrix} x_1 \text{ is low AND } x_2 \text{ is low AND } x_3 \text{ is high AND } x_4 \text{ is medium} \\ \text{AND } x_5 \text{ is low AND } x_6 \text{ is v.high AND } x_7 \text{ is high} \end{pmatrix} \text{THEN} \begin{pmatrix} y_1 \text{ is} \\ \text{high} \end{pmatrix} \\ R_4 \mid i: \text{IF} \begin{pmatrix} x_1 \text{ is low AND } x_2 \text{ is high AND } x_3 \text{ is high AND } x_4 \text{ is high} \\ \text{AND } x_5 \text{ is medium AND } x_6 \text{ is low AND } x_7 \text{ is high} \end{pmatrix} \text{THEN} \begin{pmatrix} y_1 \text{ is} \\ \text{v.high} \end{pmatrix} \\ R_5 \mid i: \text{IF} \begin{pmatrix} x_1 \text{ is high AND } x_2 \text{ is low AND } x_3 \text{ is low AND } x_4 \text{ is low} \\ \text{AND } x_5 \text{ is low AND } x_6 \text{ is v.low AND } x_7 \text{ is low} \end{pmatrix} \text{THEN} \begin{pmatrix} y_1 \text{ is} \\ \text{v.low} \end{pmatrix} \\ R_6 \mid i: \text{IF} \begin{pmatrix} x_1 \text{ is low AND } x_2 \text{ is low AND } x_3 \text{ is medium AND } x_4 \text{ is high} \\ \text{AND } x_5 \text{ is high AND } x_6 \text{ is high AND } x_7 \text{ is medium} \end{pmatrix} \text{THEN} \begin{pmatrix} y_1 \text{ is} \\ \text{low} \end{pmatrix} \end{cases}$

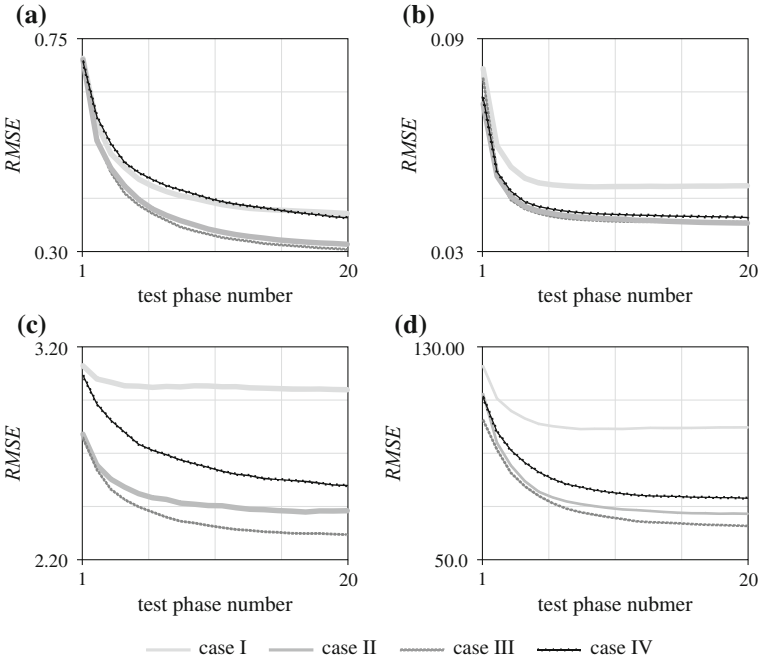

Fig. 3 RMSE for test phases for considered problems: **a** #1, **b** #2, **c** #3, **d** #4

5 Conclusions

In this paper a new algorithm for online management of fuzzy rules base for nonlinear modeling was proposed. This algorithm allows, among the others, online creating of fuzzy rules and fuzzy sets, assigning weights to fuzzy rules weights and use proposed background learning mechanism. The proposed approach allows to achieve good accuracy in comparison to standard EFS. The biggest improvement was obtained by using the background learning on the basis of auxiliary data samples. The proposed method of calculating fuzzy rules weights allows for further improvement in accuracy of the system. Moreover, the proper selection of the system parameters allows to obtain clear fuzzy rules simultaneously with low complexity of the system. The obtained results can be considered as satisfactory.

Acknowledgment The project was financed by the National Science Centre (Poland) on the basis of the decision number DEC-2012/05/B/ST7/02138.

References

1. Cpałka, K., Łapa, K., Przybył, A., Zalasiński, M.: A new method for designing neuro-fuzzy systems for nonlinear modelling with interpretability aspects. Neurocomputing **135**, 203–217 (2014)
2. Gacto, M.J., Alcalá, R., Herrera, F.: Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures. Inf. Sci. **181**(20), 4340–4360 (2011)
3. Kaisler, S., Armour, F., Espinosa, J.A., Money, W.: Big data: issues and challenges moving forward, system sciences (HICSS). In: Proceedings of the 46th Hawaii International Conference on, IEEE, pp. 995–1004 (2013)
4. Kibler, D., Aha, D.: Instance-based prediction of real-valued attributes. In Proceedings of the CSCSI (Canadian AI) Conference (1988)
5. Łapa, K., Cpałka, K.: On the application of a hybrid genetic-firework algorithm for controllers structure and parameters selection. Adv. Intel. Syst. Comput. **429**, 111–123 (2016)
6. Lughofer, E.: Evolving Fuzzy Systems—Methodologies. Advanced Concepts and Applications. Springer, Berlin (2011)
7. Lughofer, E.: On-line assurance of interpretability criteria in evolving fuzzy systems—achievements, new concepts and open issues. Inf. Sci. **251**, 22–46 (2013)
8. Lughofer, E., Hüllermeier, E.: On-line redundancy elimination in evolving fuzzy regression models using a fuzzy inclusion measure. In: Proceedings of the EUSFLAT 2011 Conference, Elsevier, Aix-Les-Bains, France, pp. 380–387 (2011)
9. Maciel, L., Lemos, A., Gomide, F., Ballini, R.: Evolving fuzzy systems for pricing fixed income options. Evolv. Syst. **3**(1), 5–18 (2012)
10. Rutkowski, L.: Computational Intelligence. Springer, Berlin (2008)
11. Rutkowski, L., Cpałka, K.: Flexible weighted neuro-fuzzy systems. In: Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02), Orchid Country Club, Singapore, November 18–22 (2002)
12. Rutkowski, L., Cpałka, K.: Neuro-fuzzy systems derived from quasi-triangular norms. In: IEEE International Conference on Fuzzy Systems, Budapest, July 26–29, vol. 2, pp. 1031–1036 (2004)
13. Soleimani, H., Lucas, K., Araabi, B.N.: Recursive Gath-Geva clustering as a basis for evolving neuro-fuzzy modelling. Evolv. Syst. **1**(1), 59–71 (2010)
14. Sugeno, M., Yasukawa, T.: A fuzzy logic based approach to qualitative modeling. IEEE Trans. Fuzzy Syst. **1**, 7–31 (1993)
15. Thrun, S.: Explanation-Based Neural Network Learning: A Lifelong Learning Approach. Kluwer Academic Publishers, Boston (1996)
16. Ward, J.H.: Hierarchical grouping to optimize an objective function. J. Am. Statist. Assoc. **58**, 236–244 (1963)
17. Waugh, S.: Extending and benchmarking Cascade-Correlation. PhD thesis, Computer Science Department, University of Tasmania (1995)

Money Laundering Analytics Based on Contextual Analysis. Application of Problem Solving Ontologies in Financial Fraud Identification and Recognition

Mariusz Chmielewski and Piotr Stąpor

Abstract Advances in automatic reasoning and the availability of semantic processing tools encourage operational specialist to extend existing link analysis methods towards contextual data awareness. In this paper we summarise a proof of concept implementation of IAFEC Ontology Toolkit for financial fraud identification based on set of problem solving ontologies. The method, algorithms and software is a contribution for IAFEC analytical tools demonstrating semantic-aware association analysis. The novelty in such approach comes from incorporating heterogeneous types of data which usually are processed by graph or network methods. The development of semantic tools, extend capabilities of graph-based approach by delivering indirect association identification as well as methods for inference path explanation. Presented material provides high level view of the method and analytical algorithms which rely on logic reasoning and semantic association identification and ranking. Developed method has been implemented as a standalone java application integrated within Protégé OWL 5.0. Such characteristic allows for further extensions and usage as a part of processing flow utilising ontology processing tools.

Keywords Financial fraud identification · Data mining · Knowledge discovery · Semantic association · Ontologies · Context-aware processing

1 Introduction

Money laundering is becoming one of key problems for governments due to obvious impact on tax evasion and the development of organized crime. Technology involvement and the globalization of money transfers make it even

M. Chmielewski (✉) · P. Stąpor
Cybernetics Faculty, Military University of Technology, Gen. Kaliskiego 2,
00-908 Warsaw, Poland
e-mail: mariusz.chmielewski@wat.edu.pl

P. Stąpor
e-mail: piotr.stapor@wat.edu.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*,
Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_3

harder to identify and recognize such practices. EU governments for many years utilize ICT tools to monitoring financial transfers in order to detect suspected transfers. However, such analysis relies mostly on manually implemented rules and expert's knowledge. Identifying new cases of frauds or frauds camouflaged by chains of organization or individuals, timespan or loosely coupled facts. For such cases there exists a need for automatic filtering and ranking tools which can act in preliminary phase of processing and pick up transfers from stream of financial data. The complexity of problems, generates also the need for research and development in the domain of automatic data retrieval, classification and inference. Several research projects have aimed at developing domain model as methods for information retrieval and inferring financial frauds such as FF POIROT [1], DOGMA [2]. This work has revised available materials and derives some assumptions shared in previous research projects [1–3] (Fig. 1).

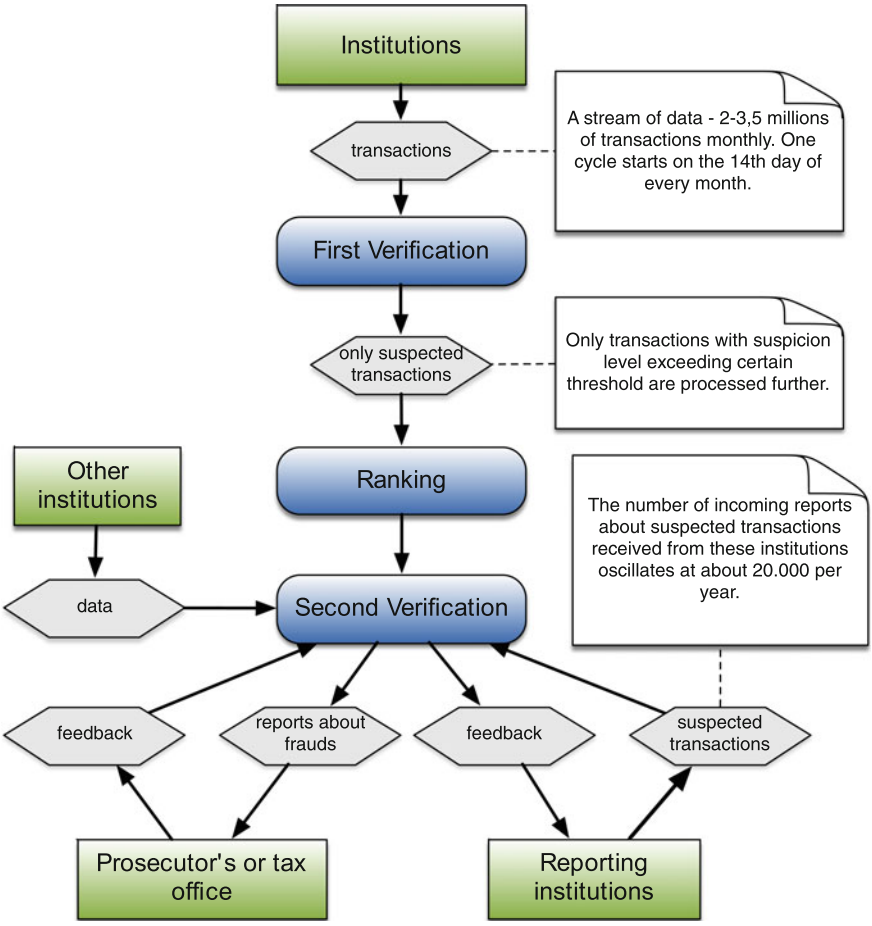

Fig. 1 Transactions' flow and fraud detection process (based on GIIF requirements)

The IAFEC project combines several analytical methods for financial fraud detection mainly connected with application of network analysis, reasoning and structural analysis. The novelty of presented approach is to apply domain modelling and problem solving ontologies in data processing chain, in order to identify hidden associations in knowledge base processed by DL, FOL and semantic graph processing reasoners. Application of semantic models gives also another advantage, context-aware relationships implementation which extend method's multi perspective capabilities.

The General Inspector of Financial Information (GIIF) is an institution established in Poland as a country's central element of national anti-money laundering and anti-terrorist financing system [4]. The article presents the approach devised to develop automatic fraud recognition system developed at Military University of Technology in cooperation and expertise provided by GIIF.

Every month GIIF collects a large volume of transactions records, ranging from 2 to 3.5 million. Every one of these has to undergo first verification, where it gets classified either as normal or suspected. Transactions tagged as belonging to the latter category are then ranked according to fraud presumption level from the first verification, the amount of money involved and importance (standard fraud, terrorist related activity, government corruption, etc.). Transactions, that score exceeding threshold value, are the ones processed further.

Second verification is more thorough than the first one. Additional data may be taken into account (such as records from Legal Register of Companies in Poland—KRS, figures collated by Central Statistical Office in Poland, information from CEPIK—Central Register of Vehicles and Drivers or from CEIDG—Central Registration and Information on Business) as well as information supplied by external institutions, which are obligated to deliver them on GIIF's demand. Two main problems at this stage are: incompatible data formats incoming from external institutions and the inability to process all suspected transactions in reasonable time. In practice, only some low-ranked transactions are left unattended.

In Poland there is a number of financial institution which may lodge transaction which raises suspicions directly to GIIF. Those records bypass the first part of process and are automatically admitted to Second verification.

Upon fraud discovery, the case is reported to either prosecutor's office or tax office, which later are expected to send feedback which may help to improve the verification process. GIIF also sends its own feedback to reporting institutions in order to improve their preliminary filtering capabilities.

2 Analytical Method

The presented approach addresses the problem of storing and analyzing a large volume of asynchronous financial transactions. For reader's convenience a schema of the entire process has been attached below.

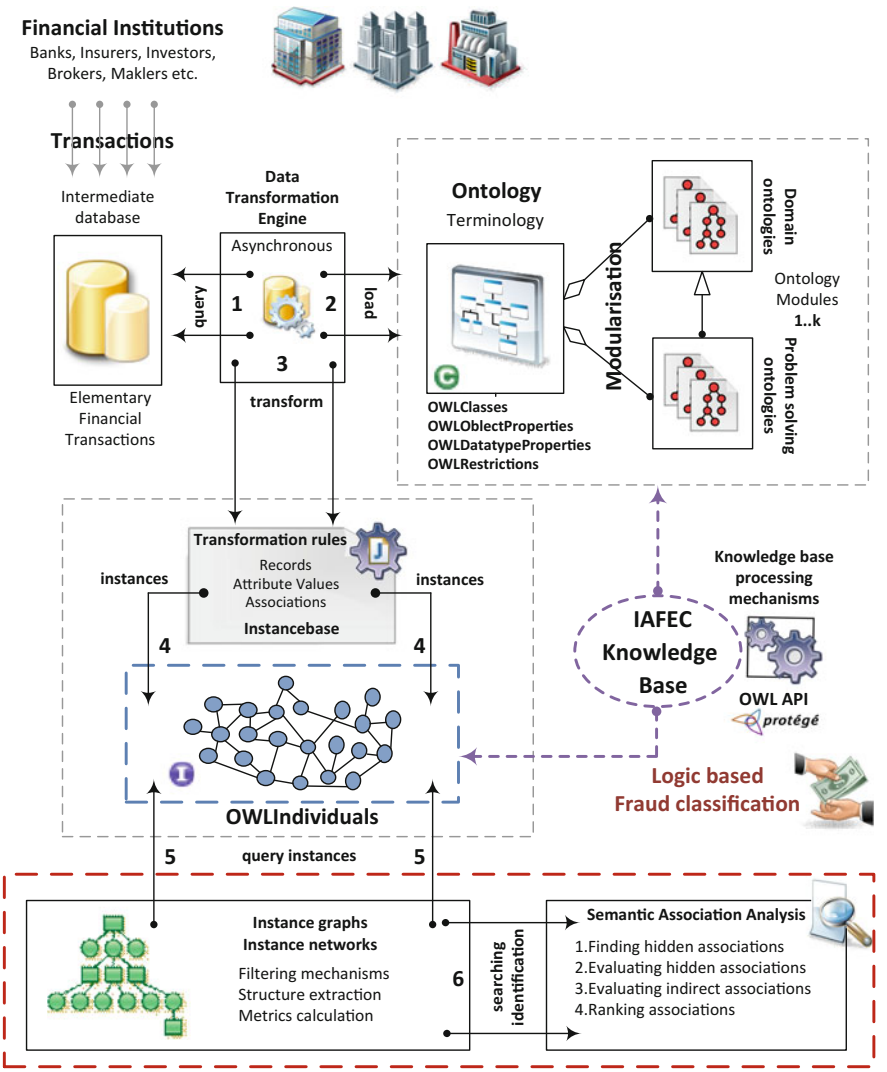

Fig. 2 The proposed solution for data acquisition, processing and reasoning

The system loads the definitions of terminology from a dedicated ontology. After that the system is ready to read data and map the transactions' records to instances placed in ontology. These individuals are then assigned to respective concepts and connected with necessary relations between themselves and instances already contained within knowledgebase. The aforementioned preprocessing is done via webservices, that contain transformation mechanisms implemented in Java.

Of course no reasoner is able to process that amount of data in sensible time. In response to this problem a following solution was devised:

– transactions are aggregated into sets during preprocessing with respect to time of realization and transaction sides (Each set is represented as an individual with datatype roles attached, pointing to numerical indicators—e.g. total amount value); instance base is transformed into multigraph and divided into connected components of which each one undergoes reasoning separately. (Lack of path between vertices indicates a lack of semantic linkage between the individuals they represent. A possibility, albeit small, that some connection between such pair could still exist is conceivable. Given circumstances—problem size and complexity, however, authors believe such approach is justified and permissible) (Fig. 2).

After preparing an instance base corresponding to a connected component the HermiT reasoner is activated and performs hyper-tableaux inference algorithm using the mechanisms of both description logic (DL). A collection of DL-safe SWRL rules is also taken into account by HermiT's engine [5]. This yields a classification of instances to appropriate concepts.

It has been shown by numerous tests, that reasoning about structure, say cycles for example, with DL and SWRL is difficult and ineffective even for a small knowledgebase. Here a structural evaluation proposed in [1] comes into play. Its inference engine analyzes structure of a given connected component and performs searches the most valuable pathways (semantic associations) for information indicating the potential cases of fraud (as defined in the ontology IAFEC). Vertices and links are then tagged with values, thus resulting in weighted multigraph creation. For further information about possible interpretations of basic structural measures—e.g. link relevance you please refer to [1–3].

A reader might also notice a clear division of ontologies into the ones containing domain description and the other intended for problem solving. The purpose of this separation will be further discussed in the next section.

3 Modelling Methodology

The process of ontology modelling has been described in many papers, but it is not common to describe the organization of ontology modules especially within large domain models. Our approach defines ground rules for organizing domain ontologies with respect to terminology and instance data. Ontology modularization helps to cluster and share ontology definitions but most of all supports memory optimization schemes while processing large instance bases.

Developed techniques [3, 6] of ontology modularization have been mainly used in Protégé OWL environment and capture following rules:

– cohesive encapsulation of domain concepts in modules, the module's theme should be logically and thematically consistent;
– separate terminology and instance data, therefore move ontology individuals to separate OWL files importing required terminologies (optimize the imports);

- store primitive concepts (OWL classes) and roles defined between primitive concepts (OWL object properties) in separate module, distinct to defined concepts—constructed using restrictions referencing primitive concepts;
- for ontology integration purposes it is useful to develop separate module storing semantic bridges, definitions of concepts mapping concepts across ontology modules using concept or role equivalence;
- distinction between DL constructors and FOL rules stored in separate modules.

Developed IAFEC ontology consists of following domain descriptions:

- Social layer description—people, social groups, all interactions with social layer and activities concerned with education, religion, interests, employment, leisure;
- Institutions and organizations their description including group members and organization profile and interests;
- Financial flow description—description of atomic financial operations performed by agents (users, automata) in financial systems, followed by annotation of participants of this process;
- Information sources—media, source description, content;
- Registry data (PESEL, GIIF, CEPiK, CEIDG, KRS).

The analytical method and software facilitates review and analysis of links (associations) in large (multi) graph structures by dividing the processed knowledge base into two separate parts: instance base containing data (facts) and terminology (represented in the form of two layers—structural and hierarchical weighted graphs) [3], containing description of modelled fragment of universe. This gives a user a clearer picture of collected information and therefore improves their analysis efficiency.

- structural layer contains relationships (properties) that exist between different types of objects—e.g. Two legal persons, between the institution financial and natural person.
- hierarchical layer allows to visualize a hierarchy of concepts—e.g. A limited partnership is a special kind of partnership, which in turn is a subtype of the company.

4 IAFEC Semantic Association Analyzer Tool

4.1 Environment Features

The project resulted in implementation of a set of tools for providing a support for analysis, detection and prevention of financial crime. The modeled fragments of reality (transactions, financial crimes) along with provided data within developed solution, are represented as multigraphs. Although automatic reasoning mechanisms provides inference by utilizing Description Logic and Semantic Web Rule

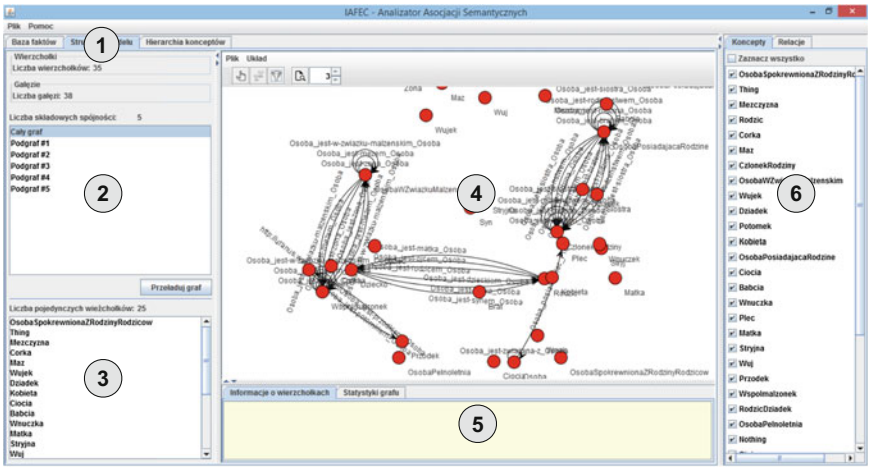

Fig. 3 Overview of functional. A family colligations ontology module loaded into Semantic Association Analyzer

Language on semantic network level, an effort has been made to introduce an extra layer of conclusion derivation—based on the theory of graphs and networks. This lead to the development of a dedicated environment called IAFEC Semantic Association Analyzer (Fig. 3).

The tool allows visualization of knowledge bases as multigraphs. In order to make presented data comprehensible and facilitate its analysis for humans, the tool splits knowledgebase into instance base and ontology. The latter is further subdivided into hierarchical and structural layers. To avoid losing important information, analyzer maintains the relationships between elements from different layers, yet renders them invisible. E.g. When a vertex representing instance is clicked, the application displays the concepts to which it has been classified by either a knowledge engineer or a reasoning engine (Table 1).

Table 1 A description of tool's functional widgets tagged with numbers in Fig. 3

No.	Description
1	The tabs allow to switch views between instance base, ontology structural and hierarchical layers
2	The list containing graph's connected components as well as an option to display whole graph
3	The list of isolated vertices
4	Ontology presented as a multigraph
5	The first tab is reserved for information (given and inferred) about selected vertex, while the latter allows user to browse a comprehensive graph's statistics
6	A list of concepts and relations with attached checkboxes allows user to select what ontology elements and associated instances they want to filter out

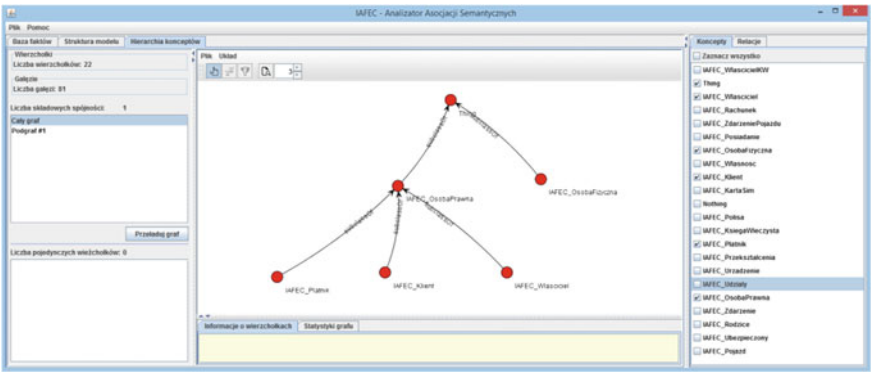

Fig. 4 A visualization of a fragment of ontology's hierarchical layer with IAFEC Semantic Association Analyzer

Given the computational complexity of algorithms used to associate facts in knowledge bases and the number of transactions inflowing to GIIF each year, the program has been designed to provide an ability to work on the weak graphs' components. This approach reduces amount of work, as searching for relationships between instances located in different connected components is usually futile due to the lack of any linkage between such pairs of individuals. This division also helps users, who may not be able to effectively explore and draw conclusions while being overwhelmed by entire available data (Fig. 4).

An example of analyzer usage to analyze and visualize hierarchical layer is provided above.

A visualization module of the analyzer provides users with means of viewing, navigating and zooming the graph structures. Its features are as follows: vertices can be selected and moved; labels can be either displayed or hidden; isolated vertices (facts or concepts unrelated with others) can be removed from visible elements list; an option is available to filter out all vertices that are located further than maximal distance (in terms of edge count) from a selected node. Finally, the module comes with a set of built-in vertices layout algorithms.

The graph structures can be exported to a GraphML files and processed further using tools that support that kind of format. Moreover, the analyzer can convert instances of datatype relationships (edges connecting individuals with literals) to attributes of vertices (Fig. 5).

4.2 Architecture

IAFEC Semantic Association Analyzer is written in Java and utilizes owl-api 3.5.0 library for performing operations involving OWL language and format. This api has

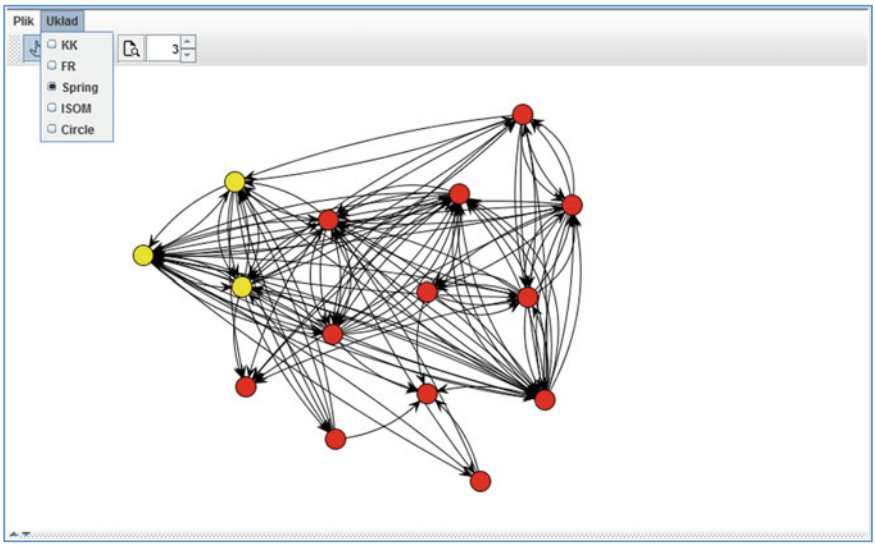

Fig. 5 The visualization of fraud scenario terminology graph (part of ontology module) provided in data module of IAFEC Semantic Association Analyzer

two advantages: firstly, it allows to connect and invoke reasoning engines through OwlReasoner interface; secondly, it is also a base library for Protégé 5.0 (a tool well-known among ontology designers) and hence its usage guarantees high level of compatibility with that environment.

The reasoner employed by the analyzer for performing an inference is *HermiT 1.3.8.4*, which is based on hyper-tableaux algorithm and supports both DL expressions and DL-safe SWRL rules [4].

For the purpose of parsing XML files, the application makes use of *jdom* library. Exporting graphs to GraphML format is obtained by employing *io* module of *jung* 2.0.1 library. *Jung* is also utilized for most functions related to graphs' visualization and manipulation. It also provides some basic algorithms for analysis of their structures.

The project utilizes some reusable components produced while developing *Terrana 2.x* environment—a project originally dedicated for detecting terrorists' attacks using semantic technologies, which later became a multi-purpose, pluggable to *Protégé*, semantic association analyzer and also an innovative tool that introduced mechanisms for structural ontology assessment [5].

The simplified structure of presented solution can be found in the UML component diagram included below (Fig. 6).

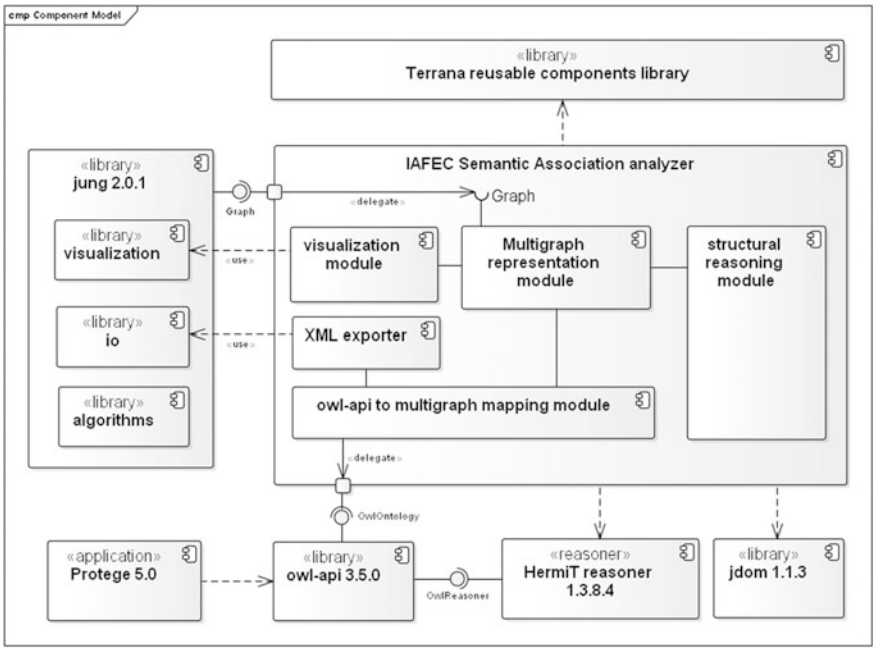

Fig. 6 The architecture component perspective describing main structure and functional responsibilities within IAFEC Semantic Association Analyzer

5 Conclusion

Presented in the paper method has implemented in form of a IAFEC Semantic Association Analyzer tool which can be used as a human-in-the-loop software and a set of predefined web services. Such construction is intentional as the environment should deliver capabilities for seamless integration within larger heterogeneous SOA environment. The framework architecture of IAFEC environment provides, a process based analytical workflow, which delivers various analytical methods. Among many GIIF and experts network based analysis, graph querying, and semantic analysis delivers effective ways for scenario based analysis as well as in-stream data selection and filtering. Especially the last variant of tool usage is useful and provides required functionality for large scale data processing. During functional tests a set of financial fraud datasets have been analyzed providing accurate results, however the analysis need to consider the limitations of in-memory ontology processing model. Most recent development of IAFEC semantic tools move towards utilizing Neo4J graph database and triple stores in order to increase the number of ontology instances. The modularization methodology simplifies memory management and supports instance data loading during the analysis which helps the knowledge engineer and delivers larger instance base processing capabilities.

Acknowledgments This work was partially supported by supported by the National Center For Research and Development research projects DOBR/0073/R/ID1/2012/03: "Advanced ICT techniques supporting data analysis processes in the domain of financial frauds" and internal grant WAT/RMN-948 "Sensor data fusion methods utilizing semantic models and artificial intelligence methods".

References

1. Basili, R., Cammisa, M., Pennacchiotti M., Zanzotto F.M.: Ontology-driven Information Retrieval in FF-Poirot (2003)
2. Jarrar, M., Meersman, R.: Ontology engineering—the DOGMA approach. In: Advances in Web Semantics I. LNCS, vol. 4891. Springer, Berlin (2008)
3. Chmielewski, M.: Ontology-based association assessment method using graph and logic reasoning techniques. Military University of Technology, Warsaw (2011)
4. Poland's Ministry of Finance. http://www.mf.gov.pl/ministerstwo-finansow/dzialalnosc/giif/system
5. Dentler, K., Cornet, R., Ten Teije A. De Keizer N.: Comparison of Reasoners for large Ontologies in the OWL 2 EL Profile. In: Semantic Web, vol. 2 (2011)
6. Chmielewski, M., Stąpor, P.: Medical data unification using ontology-based semantic model structural analysis. In: Information Systems Architecture and Technology: Proceedings of 36th International Conference on Information Systems Architecture and Technology—ISAT 2015 (2016)

ALMM-Based Methods for Optimization Makespan Flow-Shop Problem with Defects

Edyta Kucharska, Katarzyna Grobler-Dębska and Krzysztof Rączka

Abstract The paper presents a new algorithm for solving flow-shop manufacturing problem with time limits, the quality control, removing of manufacturing defects (quality lack) on an additional repair machine and re-treatment of task in technological route. Because an appearance of the defect is an unexpected event the quality control results as well as a job processing time are not known a priori. Thus, we deal with stochastic uncertainties. Our algorithm is based on algebraic-logistic meta-model (ALMM) methodology and is a combination of the searching algorithm with the special local criterion and the method of algebraic-logical models switching. The searching algorithm has been determining the deterministic problems solution on the basis of discrete process simulation until now. Switching method presents the problem by using two simple models and switching function, which specifies the rules of using these models and is used to model the removal of the manufacturing defects on an additional repair machine. The proposed approach was tested and the results of computer experiments are presented in the paper.

Keywords Flow shop scheduling problem with defects · Simulation · Optimisation · Switching · ALMM methodology

1 Introduction

The paper presents methods based on algebraic-logical meta-model for solving flow-shop manufacturing problem with time limits, the quality control, removal of the manufacturing defects (quality lack) on an additional repair machine and

E. Kucharska (✉) · K. Grobler-Dębska · K. Rączka
Department of Automatics and Biomedical Engineering, AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakow, Poland
e-mail: edyta@agh.edu.pl

K. Grobler-Dębska
e-mail: grobler@agh.edu.pl

K. Rączka
e-mail: kjr@agh.edu.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_4

re-treatment of task in technological route. The appearance of the defect is an unexpected event and would influence total job processing time. Because there are not known a priori quality control results we deal with stochastic uncertainties. To be exact, a job processing time is stochastic.

Our new approach to solve this problem proposes to combine the searching algorithm with the special local criterion and the method of algebraic-logical models switching. Our approach is based on algebraic-logistic meta-model (ALMM). So far the searching algorithm has been used to determine the deterministic problems solution on the basis of discrete process simulation. Switching method allows to present a problem by using two simple models and switching function, which specifies the rules of using these models. Combination of this method allows to solve non-deterministic problems.

Our motivation is as follow. The task of scheduling is one of the most challenging problem and the majority of real scheduling problems are known to be NP-complete or NP-hard in their general form. Moreover real manufacturing scheduling problems are also dynamic and subject to a wide range of stochastic uncertainties, such as stochastic processing time and rush order. Therefore, the production scheduling under uncertainty has indeed attracted much attention in recent years [1, 2, 4, 16, 18] However the considered flow-shop problem with defects is not equivalent to the flow-shop problem with stochastic uncertainties. The most important thing in considered problem is detection of unexpected events when schedule is executed and the possibility of including in schedule data or parameters changes which come from the occurrence of the event. The concept of switching models appears in the scientific literature. Most methods are used in financial econometric and economics models (Marcow chain, Markov switching multifractal, Regime-Switching Models) [3, 14, 19] or in sequential logic circuits. However these methods are related to totally different problems than considered flow-shop problems.

The method of algebraic-logical models switching for flow-shop problem with defects was proposed in [12, 13]. It is derived from method proposed by Dudek-Dyduch and Sękowski in [17]. Our previous research was focused on problem modeling and representation. The papers [12, 13] present general form of switching method, ALM models and definition of the switching functions and the switching algorithm for the flow-shop with defects, where only one repairing machine is required. This paper presents a combination of modeling method with optimization method to solve non-deterministic problems. The presented approach differs from the concept of the modeling method for discrete manufacturing processes with disturbances described in [7, 8], called a two-stage AL model transformation method. A parallel identical machine processes with job's due dates were there mentioned.

The paper is organized as follows. In Sect. 2 we describe the flow shop manufacturing system with time limits, where manufacturing defects are removed to re-treatment as a result of the quality control. Section 3 includes a definition of algebraic-logical meta model (ALMM) and description of its general application. In Sect. 4 we present general form of proposed approach including searching

algorithm with the special local criterion and the method of algebraic-logical models switching. In Sect. 5 the description of the proposed algorithm for considered problem is given. The last section contains conclusions.

2 Flow-Shop with Defects Problem

In the paper we consider flow shop manufacturing system with time limits, where manufacturing defects are removed to re-treatment as a result of the quality control. The dedicated machines process a set of jobs. We consider the divisible job. A job may consist of a single element up to k elements and can be divided into parts and the parts can be independently processed of each other as new jobs. In particular size of new jobs depends on the number of faulty elements. During the manufacturing process after the indicated treatment, there is a quality control performed always on the same indicated machine. A quality lack may be found for some part of the checking job in consequence of this control. Then, in manufacturing process it is necessary to use additional operations of defects removing on additional machine. After the repair operation all of technology route operations or part of them should be repeated for repaired items. Therefore, the total time of job execution is not known a priori, because time needed for performing additional operations to correct defects is unknown. The goal is to schedule jobs, so that all the jobs are completed before its due date and when its quality is accepted. It should be noted that in this paper we consider one machine with quality control and one repairing machine. The example of the problem is scheduling the performance of orders in powder coating plant. During manufacturing process, painted details are going step by step through workstation: chemical bath, dryer, powder coating booth, curing oven and packing. There is a quality control after heat treatment in curing oven. As a result of it the elements with quality luck should be exposed to the additional operation, i.e. powder coating must be removed. Then the operation from technological route must be repeated.

We use the following notation to describe considered problem. There is a set of $m+1$ dedicated machines $M = \{M_1, M_2, \ldots, M_{m+1}\}$, where first m machines constitute the production route $(M_1, M_2, \ldots, M_m)$, one $m+1$ machine is outside the production route and is named a repairing machine (on which all jobs with defects are repairing). This machine we denote as M_d. Moreover, we differentiate single machine with quality control from the production route denoted as M_{qc} and the M_r machine, on which jobs with defects are repaired. There is a store in front of each machine, where job must wait to be processed when machine is busy. There is a store for finished jobs too. The set of jobs to perform is $J = \{J_1, J_2, \ldots, J_n\}$. As we previously mentioned, one job is equivalent to a few elements. Therefore we assume that the jobs can be divided into jobs which include a smaller number of elements. Each j-th job, $1 \leq j \leq n$, is performed by m machines (all jobs follow the same route from the first machine to the m one). For each i-th operation, we have

the processing time of the operation p^{ij}, where i is a number of operation and j is a number of job. The processing time for jobs with defects is not known a priori and it is calculated based on the number of elements to repair. The processing time on the machine with quality control includes the quality control time. Moreover, due dates for some of the jobs $d(j)$ are given there. The aim is to determine the production schedule where all jobs are completed before their deadlines.

3 ALMM Approach for Multistage Decision Process

In the paper we consider discrete manufacturing process as multistage decision process. According to ALMM methodology, a problem is modeled as a multistage decision process (*MDP*) together with optimization criterion *Q.MDP* is defined as a sextuple which elements are: a set of decisions U, a set of generalized states $S = X \times T$ (X is a set of proper states and T is a subset of non-negative real numbers representing the time instants), an initial generalized state $s_0 = (x_0, t_0)$, a transition function $f(s; u)$, a set of not admissible generalized states $S_N \subset S$, a set of goal generalized states $S_G \subset S$.

The optimization task is to find an admissible decision sequence $\tilde{u}$ that optimizes criterion Q. The consecutive process states are generated as follow. A process starts with initial state s_0. Each next state depends on the previous state and the decision made at this state. The decision is chosen from different decisions, which can be made at the given state. Generation of the state sequence is terminated if the new state is a goal state, a non-admissible state, or state with an empty set of possible decisions. The sequence of consecutive process states from a given initial state to a final state (goal or non-admissible) form a process trajectory. The idea of an ALMM paradigm was proposed and developed by Dudek-Dyduch [6]. The definition of algebraic-logical meta-model of discrete deterministic process is in [5].

The ALMM approach allows one to solve discrete optimization problems by finding optimal or suboptimal solutions. Based on ALMM methodology the following general approach has been developed so far: method uses a specially designed local optimization task and the idea of semi-metric [5, 6], learning method uses the information gathered during a searching process [10], method based on learning process connected with pruning non-perspective solutions [15], substitution tasks method [9] (a solution is generated by means of sequence of dynamically created local optimization tasks so-called substitution tasks). This methodology is implemented as a part of application so-called ALMM Solver, which is currently developing [11]. In general applying a formal approach ALMM is useful for a wide class of difficult (especially NP-hard problem) multistage decision problems the parameters of which depend on the system state (e.g., the retooling time or resources depending on system state) and unexpected events during process (e.g., quality defect, lack of resource, failure modes). This methodology allows to design algorithms and methods in a formal way, on a general level.

4 Proposed Approach

To solve considered problem we propose to combine the searching algorithm with the special local criterion and the method of algebraic-logical models switching. The searching algorithm is used to determine the deterministic problems solution on the basis of discrete process simulation. Switching method is used when an occurrence of events during the process is detected and allows to present a problem by using simple models and switching function, which specifies the rules of using these models. Such combination allows to solve non-deterministic problems.

4.1 Searching Algorithm with the Special Local Criterion

The searching algorithm belongs to the class of heuristic algorithms based on partial generating and searching of states graph and using local optimization. But in basic form of those methods local optimization was only based on minimization (maximization) of the local increase of quality criterion. Our algorithm consists in generation of consecutive solutions [whole trajectories started from the initial state $s_0 = (x_0; t_0)$] in accordance with ALMM methodology. In the course of generating the trajectory, a specially designed local optimization task is used to choose decision in a particular state which takes into account much more information than only the one about the increase of the criterion. Particularly, this task may use the semi-metrics term in the state space. The ideas of using semi-metrics was presented in [5] and developed in [10, 15]. Characteristic element of the algorithm is the possibility of modification of local criterion weight coefficients as well as criterion form depending on current process state.

Local Optimization of Decision Choice The trajectory generation is connected with a choice of decision in subsequent states. Selecting the appropriate decision among all possible decisions in given state has a significant impact on ability to generate an admissible and good quality solution and on the ability to generate an admissible solution faster. One way of selecting a decision at the u from a set of possible decisions in given state $U_p(s)$ is using of the special local optimization criterion $q(u, x, t)$ consists of the following three parts: components related to the value of the global index of quality for the generated trajectory, components related to additional limitations or requirements and components responsible for the preference of certain types of decisions resulting from problem analysis. The basic form of the local criterion $q(u, x, t)$ is as follows:

$$\begin{aligned} q(u,x,t)' = {} & \Delta \mathrm{Q}(\mathrm{u},\mathrm{x},\mathrm{t}) + \widehat{\mathrm{Q}}(\mathrm{u},\mathrm{x},\mathrm{t}) \\ & + a_1\alpha_1(\mathrm{u},\mathrm{x},\mathrm{t}) + \cdots + a_i\alpha_i(\mathrm{u},\mathrm{x},\mathrm{t}) + \cdots + a_n\alpha_n(\mathrm{u},\mathrm{x},\mathrm{t}) \\ & + b_1\beta_1(\mathrm{u},\mathrm{x},\mathrm{t}) + \cdots + b_j\beta_j(\mathrm{u},\mathrm{x},\mathrm{t}) + \cdots + b_m\beta_m(\mathrm{u},\mathrm{x},\mathrm{t}) \end{aligned} \quad (1)$$

where $\Delta Q(u, x, t)$—increase of the quality index value as a result of decision u taken in the state $s = (x, t)$, $\widehat{Q}(u, x, t)$—estimation of the quality index value for the final trajectory section after the decision u has been realized, $\alpha_i(u, x, t)$—component reflecting additional limitations or additional requirements in the space of states, $i = 1, 2, \ldots, n$, $\beta_j(u, x, t)$—component responsible for the preference of certain types of decisions, $j = 1, 2, \ldots, m$ and a_i, b_j—coefficients which defines the weight of proper component.

In the case of the total quality index value Q minimizing, selected as the best decision is u^*, for which the local criterion value q is the smallest. Because of determining optimal weights of coefficients a a priori difficulty (both depend on the considered optimization problem as well as the input date for the particular optimization task), they may be calculate during simulation experiments or using learning method.

Local Optimization Criterion Modification In the simplest case the basic form of local criterion as well as weights of coefficients a can be established for one trajectory generation. But generally they can be modified in the course of the same trajectory generation too. The modification of criterion form can occur when some limitations lose sense or some additional limitation or situation takes place. The modification (increase or reduction) of the coefficients during the generation of trajectory may appear when certain limitations and preference of certain types of decisions are more/less important or inactive in given state of process. Both types of modification can occur automatically, based on previously defined rules. Regarding verification it is important, whether the modification should be done or not, and whether it has to require the minimum or possibly a small amount of calculation, for example, checking only one or a few coordinates of state. In fulfilling this requirement it seems possible to reduce the calculations the value of q which is implemented for each decision belonging to the set of possible decisions in given state.

4.2 ALM Switching Method

The switching ALM method for flow shop system with defects was proposed in [12]. Let us recall it in short. This method is an approach to formal notation of problems in which some events that would influence job processing occurs during the process execution and is derived from method presented in [17]. A detection of defect in produced element, i.e. irregular application of paint on a painted element, bad balance of springs in mattress, backfilling of the hollow headings in a mine may cause such events. It usually can not be predicted a priori when these events occur, but it is possible to predict the effects of these events, i.e. an appearance of a new job for processing, a necessity of repairing operation, a necessity of repeating an operation or several operations for a job with defects. Consequently, these events may cause data or parameters of the problem changes.

The ALM models switching method is based on the ALMM methodology and consists of the following elements: analysis of types of disturbances, determining the set of switching states, a distribution of problem for subproblems, ALM models creating for all subproblems, specifying types of switching, determination of the switching rules, the switching function definition.

It should be noted that a structure of ALM models for subproblems may be the same or different. It depends on the type of data changes. In the simplest case, we can use the same model. Furthermore for some problems, in which there are various types of data changes we can specify several models of simpler problems (subproblems of initial problem) taking into account the changes that can be predicted.

The method executing is as follows. A disturbance occurrence is detected in the nearest next process state s_k. Based on type of disturbance and current algebraic-logical model ALM_{now} the type of algebraic-logical model you want to switch ALM_{next} is specified. Then the switching function, using switching rules, calculates the initial state s_0 in ALM_{next} model based on the event occurrence and the current state of the manufacturing process represented by s_k of current ALM_{now} model. In the new state changes that are a consequence of the event are taken into account. After ALM models switching trajectory generation of process is continued. The notation of the method of ALM models switching was presented in [13].

4.3 Hybrid Algorithm

In the paper we propose to combine the searching algorithm with the special local criterion and the method of algebraic-logical models switching. The basic steps of proposed algorithm are as follows. The process starts with initial state s_0 and model ALM for problem without disturbance (denotes as ALM_A). Until any disturbance has not been detected, each next state depends on the previous state and the decision made at this state. The decision is chosen from different decisions which can be made at the given state using the adaptive local criterion. When a disturbance occurrence is detected, the switching ALM model method is used in the nearest next process state s_k. Based on type of disturbance and current algebraic-logical model $ALM_{now} = ALM_A$ the type of algebraic-logical model you want to switch $ALM_{next} = ALM_B$ is specified. Then the switching function calculates the initial state s_0 in ALM_{next} model based on the event occurrence and the current state of the manufacturing process represented by s_k of current ALM_B model. Then next states are calculated using the local criterion and each state are checked in respect of belonging to set of switching states. In the case of such belonging the further switching ALM models occurs, respectively to the basic model without disturbances ALM_A, other appropriate models which depends on identified type of disturbance or parameters for the current version of the model are recalculated. Generation of the state sequence is terminated if the new state is a goal state, a non-admissible state, or state with an empty set of possible decisions.

5 Algorithm for Flow Shop System with Defects

This section presents application of the hybrid algorithm to flow shop system with defects. Below its particular elements are presented. The algorithm generates process trajectory using the special local optimization task. A trajectory generation is interrupted when disturbance is detected. Then the switching ALM models method is used.

Local criterion For creating or modification the local optimization task a preliminary analysis of data was made. The purpose of this analysis was to determine some characteristic problem features. The analysis gives the following information: the number of task with deadline, the earliest due date and the latest deadline, maximum and minimum size of elements in task. As a result of this analysis the local criterion takes into account the following components:

$$q(u,x,t)' = \Delta Q(u,x,t) + \widehat{Q}(u,x,t) + a_1\alpha_1(u,\ x,\ t) + b_1\beta_1(u,\ x,\ t) + b_2\beta_2(u,\ x,\ t) \tag{2}$$

$\Delta Q(u,\ x,\ t)$ denotes the increase of time as a result of realizing decision u taken in the state $s = (x,t)$, $\widehat{Q}(u,\ x,\ t)$ is an estimation of the time for the final trajectory section after the decision u has been realized and is calculated as the average value of the minimum time to complete for the unfinished tasks (time to complete performing on current machine and processing time for the next machines in technological route). The estimation should take place with the lowest number of calculations, so there has been proposed relaxation which omitting temporal limitations, stoppages and breaks in task performing $\alpha(u,x,t)$ is connected with the necessity for the trajectory to omit the states of set S_N, it is defined by means of a semimetrics. Because non-admissible state omitting consists in task time requirements filling, the time reserve z_j for each not realized and not assigned job j with due date $d(j)$ is calculated. The reserve time z of j-th job determines how much time is actually to job deadline and it is equal to the difference $z_j = d(j) - t$. Because we want to obtain a state most distant from the set of non-admissible states, component $\alpha_1(u,\ x,\ t)$ is as follow (JN—denotes set of non-realized jobs):

$$\alpha_1(u,\ x,\ t) = \begin{cases} \frac{\sum_{j\in JN}\frac{1}{z_j}}{|JN|} & \text{for } z_j > 0 \text{ and } j \in JN \\ 0 & \text{for } z_j = 0 \text{ and } j \in JN \\ \infty & \text{for } z_j < 0 \text{ and } j \in JN \end{cases} \tag{3}$$

$\beta_1(u,\ x,\ t)$ and $\beta_2(u,\ x,\ t)$ are components responsible for the preference of certain types of decisions. First parameter β_1 is related to the preference decision, the result of which are performed job with the greatest number of elements in the first. The smaller the size is the greater penalty is charged. Thus $\beta_1(u,\ x,\ t)$ is

inversely proportional to the sum of the size of jobs to perform in considered decision. Second parameter β_2 is related to the preference decision, the result of which are performed job with the shortest processing time in the first. The longer time is causing a greater punishment. Thus $\beta_2(u, x, t)$ is proportional to the sum of the processing time of jobs to perform in considered decision. In the course of trajectory generation, the local optimization task may be modified. Problem analysis reveals that the moment of all tasks with deadline are already finished, it is no longer necessary to apply the component $\alpha(u, x, t)$ in the local criterion.

Switching ALM method elements Firstly, analysis of the problem and identification of occurring disturbances were made. This analysis took into account several aspects characterizing the specific problem: number of workstations in which quality control is performed, range and accuracy of quality control, actions to be taken after finding a lack of quality. The analysis showed that a repair on the additional machine and re-treatment can be modeled by switching the model of flow system (referred to as model ALM_A), on the model of flow system which takes into account the additional machine to repair deficiencies and the division of job into two parts (the model ALM_B). Switching takes place in the k-th state of the process, in which as a result of quality control defects were found. In this case, the additional repairing machine is needed and ALM_A model goes into the initial state of model ALM_B. After having repair, an additional machine is not necessary anymore, therefore there is a switching of model ALM_B into model ALM_A.

The set of switching states S_{switch} includes states in which the machine with quality control has just finished working and the result of quality control has been negative and the states in which the additional machine has just finished repair a defected job.

The basic flow shop system with deadlines and defects problem was split into two subproblems: flow-shop system with deadlines and flow shop system with deadlines and additional repairing machine. Then two algebraic-logical models were created. Both models were presented in [12]. It should be emphasized that the essential structure of the modified model has not been changed.

The following step was to specify the types of switching. Three types of switching should be executed: I—the system switches from ALM_A to ALM_B for detection of quality defect when there is no other quality defect; II—the system switches from ALM_B to $ALM_{B'}$, in which the parameters must be modified for detection of quality defect while there is another, previously detected, quality defect and it hasn't been repaired yet (it has been waiting for repair or is during the process of repair) and completion of repair previously detected quality defect and the work-in-progress store before repair machine isn't empty (there are previously detected tasks with defect); III—the system switches from ALM_B to ALM_A for completion of repair previously detected quality defect, the work-in-progress store before repair machine is empty and repair machine is not working (there is no previously detected task with defect).

The last step was determining of the switching rules and the switching function. Detailed formulas how to calculate the next state was described in [13].

6 Experiments and Results

The effectiveness of the searching algorithm with the special local criterion has been verified by simulation experiments many times and presented in papers [5, 10, 15]. In the tested local criterion (2) we set up weight values as follows: $a_1 = b_1 = b_2 = \frac{1}{3}$. We set up weights arbitrarily, because finding the optimal solution is not the subject of this paper. The purpose of the simulation experiments presented in this paper is to check how the combination of the switching method and searching algorithm works. Thus, firstly, we have tested how expensive it is to switch the one model to another model. Next, we have studied how number of defect repairs influences to all jobs execution time and the number of switching between models.

To change a number of defects randomly, we tested the proposed algorithm for the parameter of probability of defect occurrence $p \in \{0; 0.06; 0.125; 0.25; 0.5\}$. We have tested proposed algorithm also for different number of jobs and machine in technical route. Real execution time of the algorithm, models switching amount and number of generated states were calculated.

The algorithm has been implemented in C# and tested on an Intel core i5/CPU 2.3 GHz. Table 1 presents the experiments results shortly. We have tested five size of the flow-shop problem: 20 jobs and 5 machines, 20 jobs and 10 machines, 50 jobs and 5 machines, 50 jobs and 10 machines, 50 jobs and 20 machines.

For every size of problem we have tested three options of the return machine and the machine with quality control settings in technological rote. We have tested following information: makespan—total time of all jobs performing (Exec.T.), number of switching between models (Sw.Nb.), number of states in process (St. Nb.), computation time of algorithm (Time) measured in milliseconds, computation time of switching procedure (Sw.T.) measured in ticks, where time frequency in ticks per second is 2,337,919.

It can be seen that the execution time, number of states and number of switching between models is proportional to the probability of defects occurrence. The calculation time depends on the processes running on computer, but it can be seen that computation time of algorithm (measured in milliseconds) is much higher than computation time of switching procedure (measured in ticks).

This is because the time required to calculate the decision vector is much higher than the time required to calculate the switching function. For this reason, we can assume, that more economical is to switch between models than considering only one model the maximum model (with machine repair). It should be emphasized that the time of checking of defect occurrence or checking of if the repair machine has been finished repairing is calculated in computation time of switching procedure. Thus, it can be assumed that the calculation of switching between models is not more expensive than calculation of transition function within a single model.

Table 1 Results of experiments

Machines No		5	5	5	10	10	10
Jobs No		20	20	20	20	20	20
Lp of M_r		1	1	2	1	1	3
Lp of M_{qc}		2	3	3	2	3	5
p = 0	Exec.T.	2198	2198	2198	2606	2606	2606
	Sw.Nb.	0	0	0	0	0	0
	St.Nb.	82	82	82	182	182	182
	Time [ms]	170	173	173	346	334	326
	Sw.T.[t]	2911	2781	2577	3917	3327	3795
p = 0.06	Exec.T.	2198	2206	2204	2608	2610	2612
	Sw.Nb.	0	2	2	2	2	2
	St.Nb.	82	96	99	193	194	192
	Time [ms]	113	125	129	286	277	297
	Sw.T.[t]	1296	11528	11571	13660	13129	14195
p = 0.125	Exec.T.	2198	2220	2210	2616	2614	2624
	Sw.Nb.	0	3	2	2	2	2
	St.Nb.	82	103	99	194	206	210
	Time [ms]	112	144	119	267	274	315
	Sw.T.[t]	1253	2051	1462	3264	3576	4672
p = 0.25	Exec.T.	2198	2270	2218	2680	2660	2678
	Sw.Nb.	2	7	2	6	7	6
	St.Nb.	89	137	94	253	259	274
	Time [ms]	128	144	122	328	350	365
	Sw.T.[t]	1949	2627	2018	4379	4512	5479
p = 0.5	Exec.T.	2378	2583	2356	3068	3270	3362
	Sw.Nb.	11	17	10	19	19	21
	St.Nb.	230	225	168	483	496	505
	Time [ms]	261	236	181	698	672	702
	Sw.T.[t]	5003	4146	3503	10781	11065	44661

(continued)

Table 1 (continued)

Machines No		5	5	5	10	10	10	20	20	20
Jobs No		50	50	50	50	50	50	50	50	50
Lp of M_r		1	1	2	1	1	3	1	1	3
Lp of M_{qc}		2	3	3	2	3	5	2	3	5
Machines No		5	5	5	10	10	10	20	20	20
Jobs No		50	50	50	50	50	50	50	50	50
Lp of M_r		1	1	2	1	1	3	1	1	3
Lp of M_{qc}		2	3	3	2	3	5	2	3	5
p = 0	Exec.T.	5112	5112	5112	5583	5583	5583	6526	6526	6526
	Sw.Nb.	0	0	0	0	0	0	0	0	0
	St.Nb.	202	202	202	439	439	439	917	917	917
	Time [ms]	1526	1526	158	2916	2957	2914	6269	6399	6309
	Sw.T.[t]	4969	5613	5616	10370	11035	10145	22671	22517	23920
p = 0.06	Exec.T.	5138	5128	5116	5603	5645	5603	6526	6545	6658
	Sw.Nb.	4	4	2	3	4	4	4	5	3
	St.Nb.	224	230	208	477	489	483	1014	1043	1001
	Time [ms]	1854	1619	1530	3095	3197	2935	7079	6542	6658
	Sw.T.[t]	18359	17139	16809	20954	23371	21245	38471	44426	36827
p = 0.125	Exec.T.	5112	5154	5168	5633	5631	5584	6526	6558	6560
	Sw.Nb.	7	9	13	7	7	7	8	7	6
	St.Nb.	258	265	278	526	527	522	1098	1059	1051
	Time [ms]	1775	1735	1665	3065	3184	2987	7460	6975	6682
	Sw.T.[t]	6054	6469	7149	11658	13254	11709	27065	32767	26463
p = 0.25	Exec.T.	5112	5284	5398	5721	5679	5939	6466	6774	6755
	Sw.Nb.	15	14	22	13	13	16	15	15	20
	St.Nb.	307	299	336	608	583	649	1292	1309	1435
	Time [ms]	2019	1900	1789	3362	3768	3565	8458	9069	8747
	Sw.T.[t]	6792	7397	8372	14183	16128	15977	32579	41070	39577
p = 0.5	Exec.T.	6474	5582	5646	6040	6334	6422	6513	7986	7650
	Sw.Nb.	57	31	42	46	46	32	33	63	37
	St.Nb.	1534	436	460	1240	994	903	1936	2600	1970
	Time [ms]	1650	2573	2189	4957	5059	4641	11408	19846	10714
	Sw.T.[t]	44661	11365	12212	30254	28530	22423	52691	83993	57664

7 Conclusions

The paper presents approach based on ALMM to solving flow-shop problem, where manufacturing defects are removed to re-treatment as a result of the quality control. The special form of local criterion has been proposed to determine process trajectory. The simulation experiments tested the effectiveness of switching procedure for this NP-hard problem. The results show that switching between models is rather more economic than considering only one model the maximum model (with machine

repair). A comparative research between stochastic methods and the methodology of the switching method with special local criterion is planning in the future.

References

1. Almeder, C., Hartl, R.F.: A metaheuristic optimization approach for a real-world stochastic flexible flow shop problem with limited buffer. Int. J. Prod. Econ. **145**, 88–95 (2013)
2. Aytug, H., Lawley, M.A.: Executing production schedules in the face of uncertainties: a review and some future directions. Eur. J. Oper. Res. **161**(1), 86–110 (2005)
3. Billio, M., Ferrara, L., Guegan, D., Mazzi, G.L.: Evaluation of regime switching models for real-time business cycle analysis of the euro area. J. Forecast **32**, 577–586 (2013)
4. Choi, S.H., Wang, K.: Flexible flow shop scheduling with stochastic processing times: a decomposition-based approach. Comput. Ind. Eng. **63**(2), 362–373 (2012)
5. Dudek-Dyduch, E.: Learning based algorithm in scheduling. J. Intell. Manuf. **11**(2), 135–143 (2000)
6. Dudek-Dyduch E.: Algebraic logical meta-model of decision processes—new metaheuristics. In: LNAI vol. 9119, pp. 541–554. Springer, Berlin (2015)
7. Dudek-Dyduch E.: Modeling manufacturing processes with disturbances—a new method based on algebraic-logical meta-models. In: LNAI vol. 9120, pp. 353–363. Springer, Berlin (2015)
8. Dudek-Dyduch E.: Modeling manufacturing processes with disturbances—two-stage AL model transformation method. The 20th International Conference on Methods and Models in Automation and Robotics, MMAR proceedings, pp. 782–787 (2015)
9. Dudek-Dyduch E., Dutkiewicz L.: Substitution tasks method for discrete optimization. In: LNAI vol. 7895, pp. 419–430. Springer, Berlin (2013)
10. Dudek-Dyduch E., Kucharska E.: Learning method for co-operation. In: LNCS, vol. 6923, pp. 290–300. Springer, Berlin (2011)
11. Dudek-Dyduch E., Kucharska E., Dutkiewicz L., Rączka K.: ALMM solver—a tool for optimization problems. In: ICAISC 2014, pp. 328–338. Springer, Berlin (2014)
12. Grobler-Dębska K., Kucharska E., Dudek-Dyduch E.: Idea of switching algebraic-logical models in flow-shop scheduling problem with defects. The 18th International Conference on Methods and Models in Automation and Robotics, MMAR proceedings, pp. 532–537 (2013)
13. Grobler-Dębska K., Kucharska E., Jagodziński M.: ALMM-based switching method for FSS problem with defects. The 19th International Conference on Methods and Models in Automation and Robotics, MMAR proceedings, pp. 814–819 (2014)
14. Hamilton J.D.: Regime-Switching Models. Palgrave Dictionary of Economics, UK (2005)
15. Kucharska E., Dudek-Dyduch E.: Extended Learning Method for Designation of Co-Operation. In: LNCS, vol. 8615, pp. 136–157. Springer, Berlin (2014)
16. Pinedo, M.: Scheduling theory, algorithms, and systems. Prentice Hall, New Jersey (1995)
17. Sękowski H., Dudek-Dyduch E.: Knowledge based model for scheduling in failure modes. LNCS, vol. 7268, pp. 591–599. Springer, Berlin (2012)
18. Shabtay D.: The just-in-time scheduling problem in a flow-shop scheduling. Eur. J. Oper. Res. **216**, 521–532 (2012)
19. Yin G., Krishnamurthy V.: Regime switching stochastic approximation algorithm with application to adaptive discrete stochastic optimization. Ind. Appl. Math. **14**, 1187–1215 (2004)

Dynamic Visualization of Generalized One-Sided Concept Lattices and Their Reductions

Miroslav Smatana and Peter Butka

Abstract One of the approaches applied in data analysis is related to the theory of concept lattices, also known as Formal Concept Analysis (FCA), which is suitable for processing and analysis of object-attribute input data models. Concept lattice represents hierarchically organized structure of clusters of objects (concepts) based on the presence of their shared attributes. While basic FCA framework works only with binary input data tables, several approaches were introduced in order to process fuzzy attributes. The model of Generalized One-Sided Concept Lattices (GOSCL) is suitable to work with different types of attributes used in input data tables, which helped in understanding and interpretation of analysis. One of the main issues which remains is large number of concepts for visualization to user. The solution is to provide user with the reduction methods and advanced dynamic visualization of concept lattices and their reductions. In this paper we introduce and compare some of the implemented visualizations and reductions applied to concept lattices generated from input data.

Keywords Formal concept analysis · One-sided concept lattices · Dynamic visualization · Reductions

1 Introduction

There are many approaches and tools related to the identification of conceptual structures in datasets with the main goal to support users in their understanding of data and its structure in some specific way. One of already established methods is known as Formal Concept Analysis (FCA [1]) and is used for the creation of

M. Smatana (✉) · P. Butka
Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Letna 9/B, 042 00 Kosice, Slovakia
e-mail: miroslav.smatana@tuke.sk

P. Butka
e-mail: peter.butka@tuke.sk

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_5

concept lattice, i.e., hierarchically organized structure of clusters of objects (concepts) according to the shared attributes (similarity) of particular objects. Methods from this area found their application in conceptual data analysis, machine learning, data/text mining, information retrieval, software engineering or knowledge management.

Standard approach to FCA is based on 'crisp' case, where object-attribute model (or input formal context) is based on the binary relation (object has/has-not the attribute). Due to fact that most of the data are often described also by other types of attributes, several approaches were designed to work with multi-valued input contexts, e.g., method of conceptual scaling or fuzzy approaches for processing of fuzzy attributes (cf. [2–5]). In order to provide more practical data mining tool, one-sided fuzzification is used instead of full fuzzy models. It means that only one part of object-attribute model is fuzzified, e.g., objects are considered as it is usual for crisp case (classic subsets) and their attributes obtain fuzzy values (on the side of attributes we have fuzzy sets). The resulted model is usually called one-sided concept lattice and was introduced in [6]. Additionally, in [7] generalized version was introduced (Generalized One-Sided Concept Lattice—GOSCL), which is able to process data tables with different types of attributes, i.e., different truth value structures for attributes.

While generalized one-sided concept lattices in FCA helped to improve understanding and interpretation of analysis [8], one of the problems still remains important—how to provide to user the result of FCA-based analysis in some more suitable form, if number of concept lattices is large and generated structure is too complex. This becomes one of the main topics in FCA area and solutions can be seen in reduction methods (extraction of most interesting concepts or substructures) and enhanced visualization, or some of their combinations based on the dynamic visualization of output for user according to his current needs, i.e., usage of focused visualization, search for specific information within the lattice, or reduction of concept lattice. In this paper we focus on some of such methods mostly related to dynamic visualization and searches.

In the following section we provide the basic information on generalized one-sided concept lattices. Section 3 contains related work for concept lattices reduction methods and visualization of hierarchical structures. Next section shows some of our examples of dynamic visualizations of lattices and their reductions, including the method for local search of most interesting concept to some query (combination of attributes), which can be used as starting point for the following navigation of user.

2 Generalized One-Sided Concept Lattice

For the purpose of this paper we only describe some basic details regarding the theory of GOSCL (more details can be found in [7] or [9], theoretical background on generation of fuzzy concept lattices which leads to this specific case are described in [4]).

In general, input data for GOSCL are in form of object-attribute data table, i.e., 4-tuple c = (B,A,L,R) is called generalized one-sided formal context with set of objects B, set of attributes A, mapping L: $A \rightarrow CL$ used for identification of truth values structure for particular attributes (CL—class of all complete lattices, for any attribute a is L(a) structure of values for this attribute), and R represents data table (so-called generalized incidence relation), where $R(b,a) \in L(a)$ for any object $b \in B$ and attribute $a \in A$, i.e., $R(b,a)$ is value from structure $L(a)$ assigned to object b for attribute a.

In order to construct generalized one-sided concept lattice, for generalized one-sided formal context (B,A,L,R) we define a pair of mappings $\uparrow$: $2^B \rightarrow \prod_{a\in A} L(a)$ and $\downarrow$: $\prod_{a\in A} L(a) \rightarrow 2^B$ as follows:

$$\uparrow (\mathrm{X})(\mathrm{a}) = \inf_{b\in X}(\mathrm{R}(\mathrm{b},\mathrm{a})), \tag{1}$$

$$\downarrow (\mathrm{g}) = \{b \in B\colon \forall a \in A, g(a) \leq R(b,a)\}. \tag{2}$$

This pair of mappings ($\uparrow$,$\downarrow$) forms so-called Galois connection between 2^B and $\prod_{a\in A} L(a)$ and therefore is also basic for forming of resulted concept lattice. Then we can find set C(B,A,L,R) which contains all pairs (X,g), where $X \subseteq B$, $g \in \prod_{a\in A} L(a)$, satisfying $\uparrow(X) = g$ and $\downarrow(g) = X$. Here, X is usually called extent and g intent of some concept (X,g). With the definition of partial order on the set C(B,A,L,R) using formula $(X_1,g_1) \leq (X_2,g_2)$ iff $X_1 \subseteq X_2$ iff $g_1 \geq g_2$ we have complete lattice called generalized one-sided concept lattice. The model of GOSCL can be easily computed using incremental algorithm, already described in [8]. This algorithm starts with the greatest element (with intent containing all greatest elements for every attribute in the input dataset), then data table rows are processed one-by-one and all possible intents are found in every step of addition of particular objects intents to the set of all intents, which should stay closed under meets after every step of processing one object (row). At the end, final set of intents is used to create output set of concepts. More details on algorithm, its implementation and examples can be found in related papers (e.g. [10]).

At the end of this section we provide small illustrative example of generalized one-sided context and its corresponding GOSCL model. Let B = {b1,b2,b3,b4} is set of objects and A = {a_1,a_2,a_3,a_4} is set of attributes. Here, a_1 is binary attribute represented by two element chain with values 0 and 1, attribute a_2 is numeric attribute with values from real interval <0,1>, attribute a_3 is ordinal attribute represented by chain lattice (with more than two values, otherwise it is simply binary attribute) with three values from {0,1,2}, and attribute a_4 is nominal attribute represented by the lattice of type M_n (sometimes called diamond lattice). Generally, M_n can be used for representation of any nominal attribute, where we have greatest element ("i" in our case for attribute a_4), lowest element ("o" for attribute a_4), and n incomparable elements representing nominal values (in our example values "a" and "b" for attribute a_4). Example of data table is presented in Table 1 and corresponding generalized one-sided concept lattice constructed based on this data table is presented in Fig. 1.

Table 1 Illustrative example of generalized one-sided input context (data table)

Objs/Atts	A1	A2	A3	A4
b1	0	0.55	0	a
b2	0	0.3	1	o
b3	1	0.2	1	b
b4	0	0.8	2	i

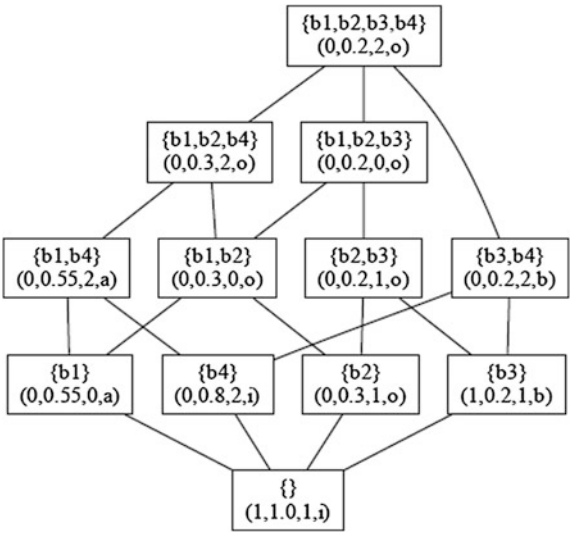

Fig. 1 Illustrative example generalized one-sided concept lattice corresponding to input data table presented in Table 1

3 Related Work

Simplified interpretation of concept lattices is important issue and there are mainly two ways for the improvement of understandability of analysis using FCA-based methods: concept lattices reductions and advanced visualization of concept lattices.

3.1 Concept Lattices Reduction Methods

Formal concept analysis was successfully applied in many applications. However, when is applied for analysis of even medium sized data the size of a concept lattice may be large, therefore some reductions are used in order to make the results more understandable and applicable. Several methods were already introduced for these purposes. Some of the techniques are threshold-based, where main goal is to find relevant concepts using some ranking method and remove the concepts under the threshold, examples of such approaches can be found in [11, 12]. In [13] authors proposed another interesting way of concept lattice reduction based on removing

links between concepts in order to get tree-based structure. Simplification of concept lattices can be also achieved by the reduction of input formal context from which concept lattice is generated. Example of such reductions based on the usage of singular value decomposition (SVD) can be found in [14, 15]. As SVD is time consuming, another approach was introduced in [16], where authors used fuzzy k-means algorithm to formal context reduction. Similar approach based on the removing of similar objects from input formal context and their replacement by one new representative object was presented in [17]. Also, it is possible to achieve reduction of concept lattice based on the merging of similar concepts into one cluster, example of such approach can be found in [18] where authors described fuzzy conceptual clustering approach.

3.2 Visualization of Hierarchical Data

Data visualization is a graphic representation of specific information, which offers an effective way for fast understanding and presentation of data. It is a key element in decision support systems, information seeking, or knowledge management. Visualization technique is defined as graphic representation for visualization of information, which must be external, systematic and clarified. Currently, a lot of visualization techniques are available, e.g., Lenger and Eppler [19] created large table of 100 visualization techniques divided by type of their usage.

There are many techniques related to visualization of hierarchical data (structures of or similar to concept lattice). Wills in [20] presented several techniques for hierarchical data visualization as node-edge layouts (standard node-edge layout for a hierarchical network, dendrogram of a hierarchical clustering), space-filling layouts (space-filling radial layout, treemap) and some techniques for interactive visualization. In [21] author presented innovative approach for hierarchical-temporal data visualization using a Tree-Ring metaphor. Another technique described in [22] is dedicated to visualization of a large set of hierarchical data on one screen using nested rectangles, where first rectangles of concepts on lower levels are built and are connected to rectangles of their parents. In [23] innovative way for visualization based on combination of traditional treemaps with arc diagram is presented. In [24] authors described the exploitation of dynamic treemaps for visualization of large set of hierarchical data within the predefined part of structure, where only nodes at actual level and their descendants are shown. In [25] dynamic Voronoi treemaps were presented, which allow us to visualize hierarchical data evolution in time.

While all previous approaches were related to visualization of hierarchies where each node has only one parent, concept lattices (in non-reduced form) are structure where nodes (concepts) can have multiple parents. Logically, Hasse diagram is frequently used for this task even if not concept lattice is visualized, e.g., in [26] authors use it for visualization and indexing of photos from social networks, in [27] Hasse diagram was applied to socioeconomic data visualization, together with the

technique for clarification of diagram using connection to self-organizing maps, and in [28] Holten presented visualization technique for such structures using radial layout.

4 Concept Lattice Visualization

Visualization is important part of data analysis using concept lattices, if we want to use it in practical data mining tasks, i.e., it is important to find fast and effective way for better understanding of concept lattices and their searching. Standard static visualization techniques are usually sufficient only for visualization of concept lattices which contains small number of concepts and links between them, understanding the data analysis results for larger lattices becomes more difficult. For that reason, we present several dynamic visualization techniques in this section on data generated by GOSCL, which were implemented in order to help visualize and search results to users. All proposed graphs were created by D3.js (http://d3js.org/), which is JavaScript library used for data visualization. All visualization techniques were tested on dataset which consist of 101 objects (each object represents some animal) and 17 attributes (15 of nominal type, 1 chain, 1 interval-based).

4.1 Dynamic Visualization

4.1.1 Dynamic Hasse Diagram

First, we started with Hasse diagram on our dataset. As mentioned before, it is natural visualization technique which accepts hierarchical data where nodes can have multiple parents. However, this technique has problem with larger data, even if we will use layer-based layout. Therefore, we made dynamic version of Hasse diagram (still organized in layers from left to right), where user can zoom to specific node, find some node details and his particular connections (see Fig. 2).

4.1.2 Double Tree

Due to problems with presentation of larger lattice, we also decided to provide tool for visualization of only some sub-part of concept lattice, which can be interesting or important for user. From the available techniques, double tree can be used for such purposes, where one middle concept from concept lattice is visualized together with the parents and children of this concept (i.e., upper and lower concepts within the structure). The edge is labeled with the changes of attributes values (or presence) between two concepts. If we click on some parent or child node, middle concept (and therefore focus of visualization) is changed to corresponding concept

Fig. 2 Dynamic Hasse diagram visualization of concept lattice

(example of technique can be found also in Fig. 5 within Sect. 4.2). This visualization technique is very transparent, however do not offer global view on whole lattice, but it can be used for exploration of concept lattice, especially if we apply function in which all already explored concepts stay visualized within the visualization desktop.

4.1.3 Zoomable Sunburst Diagram

Sunburst is an adjacency diagram. It is a variant of the node-link diagram, which rather than drawing links between nodes is drawing nodes as solid areas. Hierarchical structure should be in tree-based form, then we can have more specific nodes (children) under some general node (parent), i.e., each level of the tree hierarchy is represented by one ring (circle) with the innermost circle as the top of the hierarchy (see Fig. 3). Dynamic version of this diagram visualize hierarchy only to a certain depth from main node. As it was mentioned before, this technique can be used only on reduced concept lattice where each concept has only one parent, for this purpose we have used methods based on several tree reduction approaches described in [13].

4.1.4 Concept Browser

Also suitable way for visualization and especially exploration of large concept lattices (according to objects and their presence in concepts) is dynamic concept browser as shown in Fig. 4. The nodes on right and left represent objects and nodes

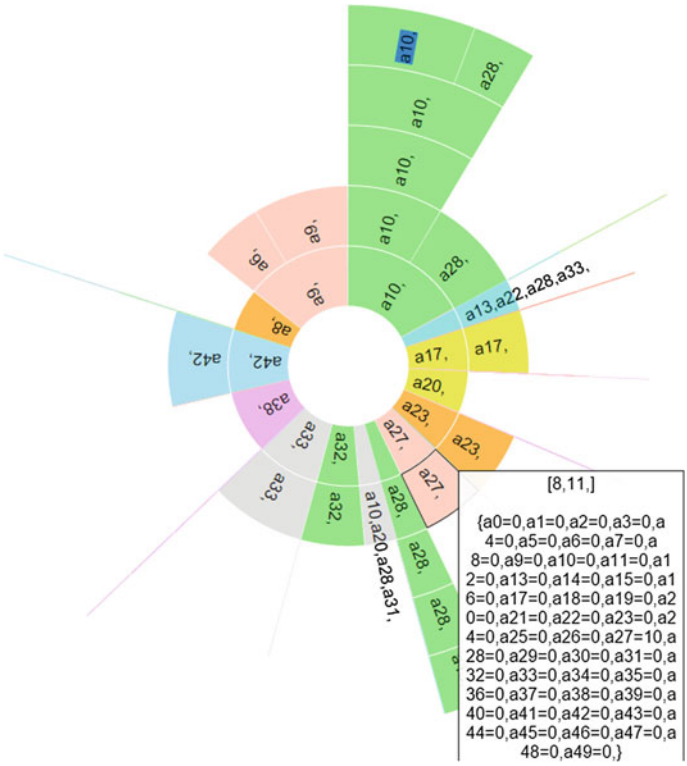

Fig. 3 Concept lattice (reduced to tree) visualization via dynamic sunburst diagram

in the middle represent concepts, edges are related to the presence of object in concept. This technique is also disarranged for large concept lattice, however it offers to user option to select only some concept and object from whole lattice, which make this technique very transparent.

4.2 Local Search in Concept Lattice and Its Exploration

As it can be seen from Sect. 4.1, it is obvious that for visualization of large concept lattice visualization techniques with local exploration features (exploration of part of lattice and navigation through structure) are more suitable and can be of high interest for user. Of course, user can select some part using previously mentioned method from the whole concept lattice, but interesting feature in case of large lattices is to provide search for most suitable concept, where exploration could start, i.e., user can use query for selection of most interesting concept and then explore its neighbors (children and parent concepts).

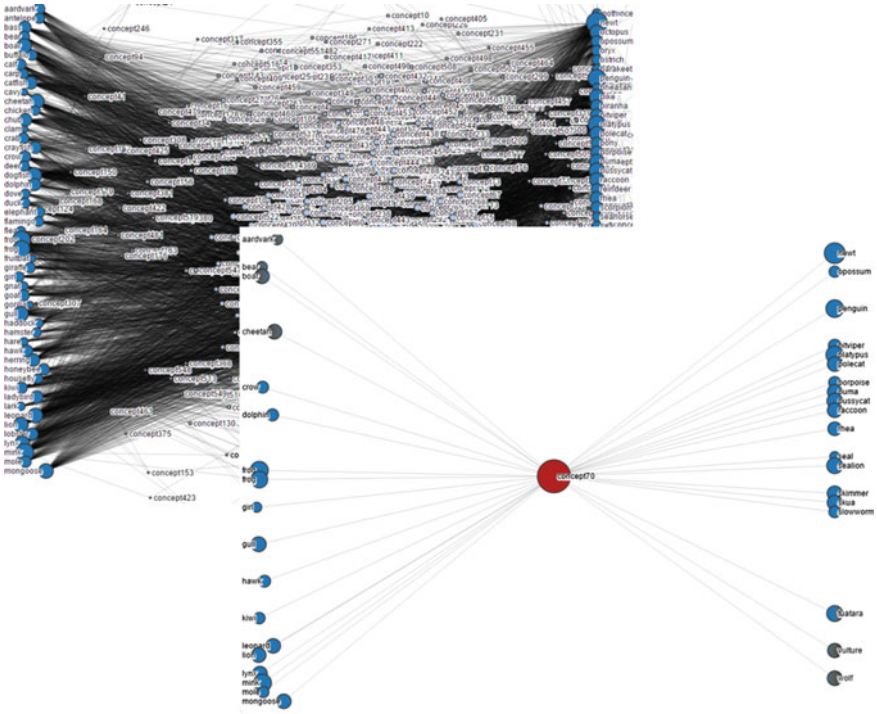

Fig. 4 Dynamic concept browser

For crisp concept lattices it is quite simple to find appropriate subpart of lattice or most similar concept to some query. We can visualize concepts with best match to query. It can be done by common used similarity metrics as Hamming or Jaccard similarity. Similarly, for the concept lattices build using fuzzy models with one type of attributes it is possible to use similarity based on Euclidean or cosine metric.

The main problem in our case is to find most suitable concept if concept lattice is generated by GOSCL algorithm, where objects are described by various types of attributes (interval, chain, nominal—diamond, general—combines chain and nominal), because we are not able to simply use some standard similarity metrics.

For that reason, we modified Levenshtein distance (http://www.levenshtein.net/) for finding the most suitable concept generated from GOSCL for presentation to user in focused visualization.

Our Levenshtein distance modification can be described by pseudocode:

1. initialize distance = 0, n = number of attributes, Q = attribute vector for query, C = attribute vector for concept
2. for i = 1:n

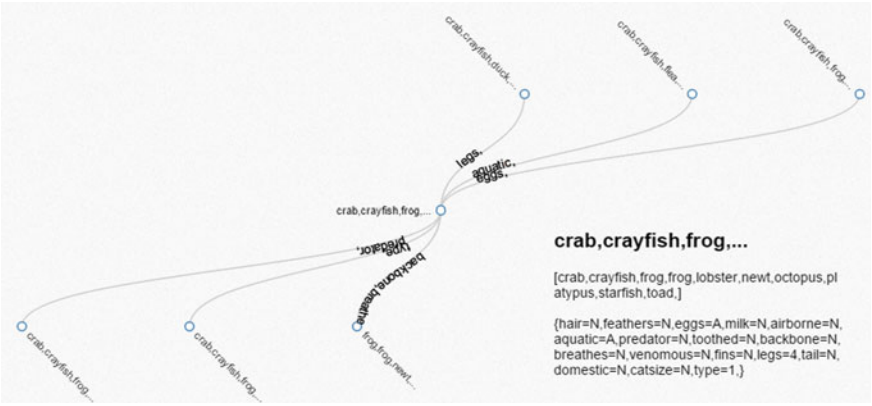

Fig. 5 Dynamic Double tree visualization of subpart of lattice acquired by the query from user

a. compare Q[i] with C[i]
b. If Q[i] is lower than C[i] then distance +=1
c. If Q[i] is incomparable with C[i] then distance +=2

3. return distance.

Dynamic visualization of subpart of concept lattice based on user query can be easily achieved by Double tree visualization technique, for example we have used concept lattices generated by GOSCL from dataset already described at the start of Sect. 4. In this case we have used user query, which tries to find concept containing animals which lays eggs, give a milk, are aquatic and has 4 or more legs. The result of this visualization is shown in Fig. 5, where we can see that our solution found suitable subpart of lattice with the middle (actual) concept that meets all conditions except for condition "giving milk". Thanks to usage of links with labels describing change of attributes to other concept we are able to easily navigate between concepts. For example, if we want to find animals which meets all conditions as in original query, but which are not aquatic, we will navigate to middle parent of actual concept.

5 Conclusion

In this paper we presented several visualization techniques and their usability with concept lattices. Dynamic visualization techniques for local subparts of lattice are very good for exploration and navigation of user within the concept lattice structure, especially if combined with some search functionality which is able to find most interesting concept for user to start the exploration. For this reason, we have also presented simple method for search in generalized one-sided concept lattices (where

input formal contexts have different types of attributes) based on the modified Levenshtein distance. This method is able to find most suitable concept to some query, which can be used as starting point for exploration and navigation by users.

Acknowledgments The work presented in this paper was supported by the Slovak VEGA grant 1/0493/16 and Slovak KEGA grant 025TUKE-4/2015.

References

1. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer Verlag, Berlin (1999)
2. Krajci, S.: A generalized concept lattice. Logic Journal of IGPL **13**(5), 543–550 (2005)
3. Medina, J., Ojeda-Aciego, M., Ruiz-Calvino, J.: Formal concept analysis via multi-adjoint concept lattices. Fuzzy Set. Syst. **160**, 130–144 (2009)
4. Pocs, J.: Note on generating fuzzy concept lattices via Galois connections. Inform. Sci. **185**(1), 128–136 (2012)
5. Antoni, L., Krajci, S., Kridlo, O., Macek, B., Piskova, L.: On heterogeneous formal contexts. Fuzzy Set. Syst. **234**, 22–33 (2014)
6. Krajci, S.: Cluster based efficient generation of fuzzy concepts. Neural Netw. World **13**(5), 521–530 (2003)
7. Butka, P., Pocs, J.: Generalization of one-sided concept lattices. Comput. Inf. **32**(2), 355–370 (2013)
8. Butka, P., Pocs, J., Pocsova, J.: Use of concept lattices for data tables with different types of attributes. J. Inf. Organ. Sci. **36**(1), 1–12 (2012)
9. Butka, P., Pocs, J., Pocsova, J.: On equivalence of conceptual scaling and generalized one-sided concept lattices. Inform. Sci. **259**, 57–70 (2014)
10. Butka, P., Pocsova, J., Pocs, J.: Design and implementation of incremental algorithm for creation of generalized one-sided concept lattices. In: Proceedings of CINTI 2012, Budapest, Hungary, pp. 373–378 (2011)
11. Butka, P., Pocs, J., Pocsová, J.: Reduction of concepts from generalized one-sided concept lattice based on subsets quality measure. Adv. Intell. Syst. Comput. **314**, 101–111 (2015)
12. Antoni, L., Krajci, S., Kridlo, O.: Randomized fuzzy formal contexts and relevance of one-sided concepts. LNAI (Subseries of LNCS) **9113**, 183–199 (2014)
13. Melo, C., Le-Grand, B., Aufaure, A.: Browsing large concept lattices through tree extraction and reduction methods. Int. J. Intell. Inf. Technol. (IJIIT) **9**(4), 16–34 (2013)
14. Gajdos, P., Moravec, P., Snasel, V.: Concept lattice generation by singular value decomposition. In: Proceedings of CLA 2004, pp. 13–22 (2004)
15. Snasel, V., Polovincak, M., Abdulla, H.: Concept lattice reduction by singular value decomposition. In: Proceedings of the SYRCoDIS 2007, Moscow, Russia (2007)
16. Kumar, ChA, Srinivas, S.: Concept lattice reduction using fuzzy K-Means clustering. Expert Syst. Appl. **37**(3), 2696–2704 (2010)
17. Dias, S., Vieira, N.: Reducing the size of concept lattices: the JBOS approach. In: Proceedings of CLA 2010, pp. 80–91 (2010)
18. Quan, T., Hui, S., Cao, T.: A fuzzy FCA-based approach to conceptual clustering for automatic generation of concept hierarchy on uncertainty data. In: Proceedings of CLA 2004, pp. 1–12 (2004)
19. Lengler, R., Eppler, M.: Towards a periodic table of visualization methods for management. In: Proceedings of the International Conference on Graphic and Visualization in Engineering (GVE 2007), Clearwater, Florida, pp. 83–88 (2007)

20. Wills, G.: Visualizing hierarchical data. In: Encyclopedia of Database Systems, pp. 3425–3432 (2009)
21. Theron, R.: Hierarchical-temporal data visualization using a tree-ring metaphor. In: Smart Graphics. Springer, Berlin, pp. 70–81 (2006)
22. Itoh, T., Yamaguchi, Y., Ikehata, Y., Kajinaga, Y.: Hierarchical data visualization using a fast rectangle-packing algorithm. IEEE Trans. Visual Comput. Graphics **10**(3), 302–313 (2004)
23. Neumann, P., Schlechtweg, S., Carpendale, S.: ArcTrees: Visualizing relations in hierarchical data. In: Proceedings of EuroVis 2005, pp. 53–60 (2005)
24. Jadeja, M., Shah, K.: Tree-map: A visualization tool for large data. In: Proceedings of 1st International Workshop on Graph Search and Beyond (GSB 2015), pp. 9–13 (2015)
25. Gotz, D.: Dynamic Voronoi Treemaps: a visualization technique for time-varying hierarchical data. IBM Research Technical Report, RC25132 (2011)
26. Crampes, M., Oliveira-Kumar, J., Ranwez, S., Villerd, J.: Visualizing social photos on a hasse diagram for eliciting relations and indexing new photos. IEEE Trans. Visual Comput. Graphics **15**(6), 985–992 (2009)
27. Fattore, M., Arcagni, A., Barberis, S.: Visualizing partially ordered sets for socioeconomic analysis. Revista Colombiana de Estadística **37**(2), 437–450 (2014)
28. Holten, D.: Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. IEEE Trans. Visual Comput. Graphics **12**(5), 741–748 (2006)

A Concept of Decision Support for Robust Resource—Constrained Scheduling Problems Using Hybrid Approach

Paweł Sitek and Jarosław Wikarek

Abstract Resource-constrained scheduling problems appear at different levels of decisions in logistics, manufacturing, computer networks, software engineering etc. They are usually characterized by many types of constraints and decision variables which often make them difficult to solve (NP-complete). In addition, these problems are often characterized by the uncertainty of resources, allocations and time. Opportunity to ask questions and get answers about the feasibility/optimality of a schedule in uncertain conditions (e.g. about available resources) is extremely important for decision-makers. This paper presents a hybrid approach to modeling and solving robust constrained scheduling problems where two environments (mathematical programming and constraint logic programming) were integrated. This integration, hybridization as well as a transformation of the problem helped reduce the combinatorial problem substantially. In order to compare the effectiveness of the proposed approach to the mathematical programming approach, illustrative example was implemented in both environments for the same data instances.

Keywords Constraint logic programming · Mathematical programming · Scheduling · Decision support · Hybrid approach · Robust scheduling

1 Introduction

Today's highly competitive environment makes it an absolute requirement on behalf of the decision makers to continuously make the best decisions in the shortest possible time. That is, there is no room for mistakes in making decisions in

P. Sitek (✉) · J. Wikarek
Control and Management Systems Section, Technical University of Kielce, Al. Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland
e-mail: sitek@tu.kielce.pl

J. Wikarek
e-mail: j.wikarek@tu.kielce.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_6

this global environment. In addition, the decision taken are most frequently subject to uncertainty associated with the resource, time, etc. Success depends on quickly allocating the organizational resources towards meeting the actual needs and wants of the customer at the lowest possible cost. Very often, the whole or parts of the decision-making problems are brought to the allocation of resources over time, which must satisfy a set of constraints. This is a generalized form of resource-constrained scheduling problem. These problems appear in logistic, manufacturing, distribution, software engineering, computer networks etc. The resource-constrained scheduling problems involve various numerical, logical and other constraints, some of which are in conflict with each other and different types of decision variables. Therefore effective and natural ways of modeling different constraint and structures is a key issue. The most efficient methods for the modeling of linear constraints are OR (Operation Research) methods, mathematical programming (MP) techniques in particular [1]. For non-linear, logical, etc. constraints, MP techniques have proved to be inefficient. For example, for modeling of allocation problems, the 0–1 constraints should be introduced. Unfortunately, the introduction of this type of constraints increases the combinatorial search space and complicates the structure of the problem. Unlike traditional approaches based on operation research, declarative constraint logic programming (CLP) provides for a natural representation of heterogeneous constraints. Constraint logic programming (CLP) is a form of constraint programming (CP), in which logic programming (LP) is extended to include concepts from constraint satisfaction problem (CSP) [2]. Constraint satisfaction problems on finite domains are typically solved using a form of search. The most widely used techniques include variants of backtracking, constraint propagation, and local search.

Constraint propagation embeds any reasoning that consists in explicitly forbidding values or combinations of values for some variables of a problem because a given subset of its constraints cannot be satisfied otherwise.

CLP approach is very effective for binary constraints (binding at most two variables). If there are more variables in the constraints the efficiency decreases dramatically. In addition, discrete optimization is not a strong suit of CP-based environments.

Based on [3, 4] and previous works on hybridization [5–7] some advantages and disadvantages of these environments have been observed. The hybrid approach of constraint logic programming and mathematical programming can help solve decision, search and optimization problems that are intractable with either of the two methods alone [8, 9].

The motivation and contribution behind this work was to apply a hybrid approach as a concept of decision support for robust resource-scheduling problems. The proposed concept allows easy modeling and effective solving of decision-making models as well as enables the decision-makers to ask all kinds of questions especially related to the uncertainty of the availability of resources and machines. This concept can be used in the planning phase to seek answers to the various types of "what if …" questions.

2 Robust Resource-Constrained Scheduling Problems

In its most general form, the resource-constrained scheduling problem [10–12] asks the following questions. What is the best way to assign the resources to the activities at specific times such that all of the constraints are satisfied (decision and search problems) and the best objective measures are made (optimization problems)?

Where are given:

- a set of activities (tasks, machine operations, development software modules, services etc.) that must be executed;
- a set of resources (machines, processors, workers, tools, materials, finances etc.) with which to perform activities;
- a set of constraints (precedence, capacity, time, availability, allocation etc.) must be satisfied;
- a set of objectives with which to evaluate a schedule's performance.

Constraints and objectives are defined during the modeling of the problem. Constraints define the "feasibility" of the schedule and their solution is sufficient for the decision and search problems. Objectives define the "optimality" of the schedule and require solutions for optimization problems.

During execution, a resource-constrained scheduling problem can be a subject to considerable uncertainty, which can be lead to numerous schedule disruptions (activities can take longer, resource requirements or availability can vary, new resource may occur, you may receive a resource failure etc.). In such cases, we should talk about robust resource-constrained problems.

The general problem includes many variations such as the flow-shop and job-shop problems, the resource-constrained project scheduling problem, and the project and multi-project scheduling problems etc.

3 A Hybrid Approach to Decision Support for Robust Resource-Constrained Scheduling Problems-Implementation Platform

The hybrid approach to modeling and solving of robust resource-constrained scheduling problems is able to bridge the gaps and eliminate the drawbacks that occur in both MP and CLP approaches which are used separately. To support this concept, a implementation platform is proposed (Fig. 1), where:

- knowledge related to the problem can be expressed as linear, integer, logical, symbolic constraints and decision variables (model CLP);
- the process of finding a solution consists of the following phases: modeling, presolving, generation and solving;
- two environments, i.e. MP and CLP are integrated and hybridized;

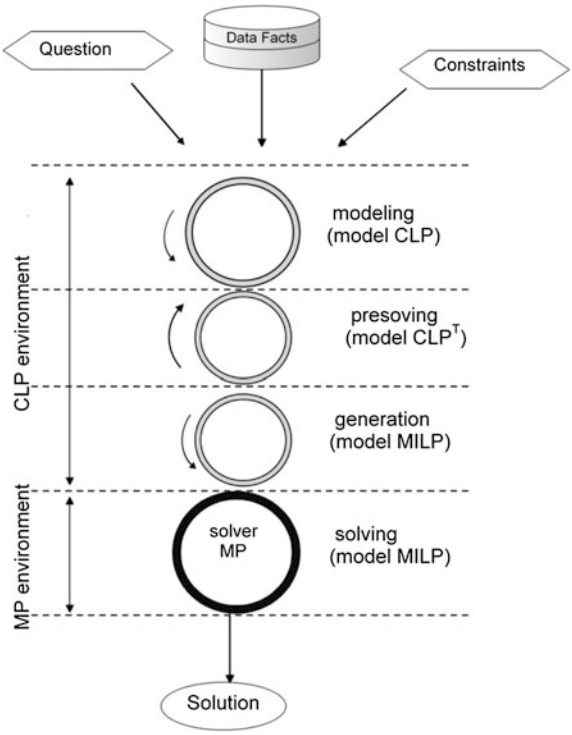

Fig. 1 The schema of implementation platform

- the transformation of the problem [13] is an integral part of this platform as a presolving method (model CLP^T);
- the final MILP (Mixed Integer Linear Programming) model is generated based on model CLP^T;
- all types of questions can be asked: general questions for decision problems: *Is it possible …*? *Is it feasible …*? and specific questions for optimization and search problems: *What is the minimum …*? *What is the number …*? *What is the configuration ... for ...*? (the list of example question in this platform for illustrative example is shown in Table 1);
- the properties of the problem, its parameters, input data are stored as sets of facts.

4 Implementation and Illustrative Example

From a variety of tools for the implementation of the CLP environment in the implementation platform, the ECLiPSe software [14] was selected. ECLiPSe is an open-source software system for the cost-effective development and deployment of constraint programming applications. Environment for the implementation of MP was LINGO by LINDO Systems. LINGO Optimization Modeling Software is a powerful tool for building and solving mathematical optimization models [15].

Table 1 Decision variables, Constraint and Questions for illustrative example

Decision variables	
V_1	The start time and the end time of each operation for activity (e.g. job)
V_2	The set of resources assigned to each operation (if this set is not fixed)
V_3	The processing times (if there are not fixed)
Constraints	
C_1	Temporal and precedence constraints—define the possible values for the start and end times of operations and the relations between the start and end time of two operations
C_2	Resource constraints—define the possible set of resources for each operation
C_3	Capacity constraints—limit the available capacity of each resource over time
Questions	
Q_1	What is the minimum C_{max} if the number of additional resources (employees) is N?
Q_{1A}	What is the minimum C_{max} if the number of additional resources (employees) is P% of N?
Q_2	Is it possible to schedule in C_{max} if the number of additional resources (employees) is N?
Q_{2A}	Is it possible to schedule in C_{max} if the number of additional resources (employees) is P% of N?
Q_3	What is the minimum C_{max} if the machine P_i is out of order in periods T_i?
Q_{3A}	Is it possible to schedule in C_{max} if the machine P_i is out of order in periods T_i?
Q_4	Is it possible to schedule orders in C_{max} if machines P1 and P2 cannot be used simultaneously?
Q_{4A}	Is it possible to schedule orders in C_{max} if additional resources S1 and S2 cannot be used simultaneously?

The illustrative example relates to scheduling in job-shop environment with additional constrained resources (employees). In a given time, only one additional resource (employee) can be assigned to a product/machine.

In order to evaluate the proposed implementation platform, a number of computational experiments were performed for the illustrative example. Variables, the main constraints and questions of the illustrative example are shown in Table 1, while the structure of the facts and their relations in Fig. 2. The experiments were performed for the instance consisting of 6 machines (P1…P6), 10 products (U1…U10) and different sets of employees (S1…S4). Each set consists of a limited number of employees with specified qualifications.

The set of facts for illustrative example is shown in Appendix 1. The experimental results for each of the questions are presented in Table 2 and the corresponding schedules for questions Q_1, Q_{1A}, in Fig. 3a, b.

This stage of the research showed a very high potential and flexibility of the platform in supporting various types of decision by questions Q_1…Q_4. The ease of modeling questions stems from declarativeness of the CLP environment.

Particularly noteworthy are the answers to the questions Q_{1A}, Q_{2A}, Q_3, Q_4, Q_{4A}. In fact they support the decision-maker under uncertainty of the resource and time. The corresponding schedules are so robust at these events (i.e. Figure 3b).

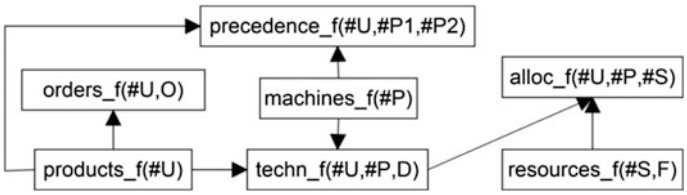

Fig. 2 Structure of facts and their relationships for illustrative example (#-key attributes, → one to many relationship)

Table 2 Results of the experiments for questions $Q_1 \ldots Q_4$ using the implementation platform

Questions	Parameters	Answer
Q_1	N = 4	C_{max} = 13
Q_{1A}	25 % (N = 1)	C_{max} = 15
Q_2	N = 4, C_{max} = 16	Yes
Q_{2A}	50 % (N = 2), C_{max} = 14	No
Q_3	M_1 is out of order in periods T = 1, T = 2, T = 3, T = 4, T = 5, T = 6	C_{max} = 15
Q_{3A}	M_1 is out of order in periods T = 1, T = 2, T = 3, T = 4, T = 5, T = 6, C_{max} = 18	Yes
Q_4	P1 and P2 cannot be used simultaneously	C_{max} = 14
Q_{4A}	Additional resources S1 and S2 cannot be used simultaneously	C_{max} = 13

In the second stage of the research, effectiveness and efficiency of the proposed platform was evaluated in relation to the MP environment. For this purpose, question $Q_1 \ldots Q_{3A}$ were implemented in two environments, MP and in the proposed implementation platform. The results are shown in Tables 2 and 3 respectively.

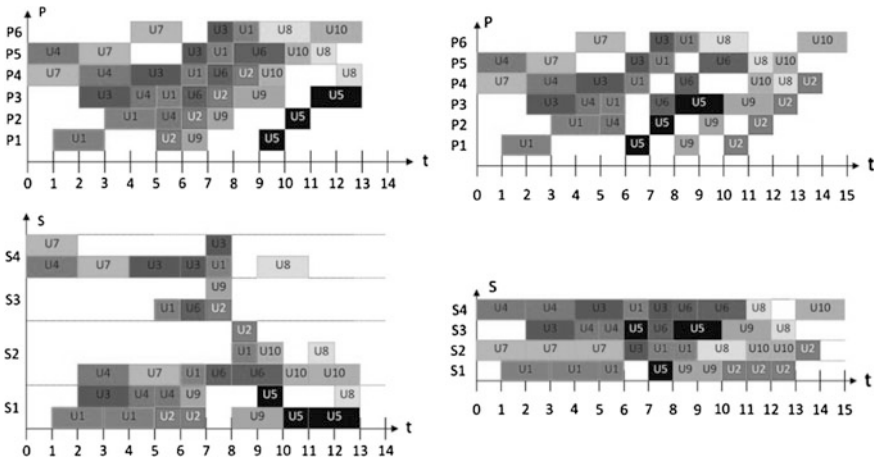

Fig. 3 **a** Schedules for machines (*top*) and employees (*bottom*) corresponding to the question Q_1 (only one employee to the machine). **b** Schedules for machines (*top*) and employees (*bottom*) corresponding to the question Q_{1A} (only one employee to the machine) where P = 25 %N (N = 1)

Table 3 Results of the experiments for questions Q1…Q4 using the implementation platform and mathematical programming (MP)—comparative analysis in terms of the number of variables, constraints and response time

Pn	MP				Implementation platform			
	V	C	T	Answer	V	C	T(s)	Answer
Q_1	9120	4976	567	$C_{max} = 13$	2736	1835	62	$C_{max} = 13$
Q_{1A}	9120	4976	900**	$C_{max} = 18$*	2736	1835	89	$C_{max} = 15$
Q_2	9120	4976	345	Yes	2736	1835	45	Yes
Q_{2A}	9120	4976	456	No	2736	1835	34	No
Q_3	8360	4976	900**	$C_{max} = 17$*	2688	1835	87	$C_{max} = 15$
Q_{3A}	8360	4976	456	Yes	2688	1835	23	Yes
Q_4	–	–	–	–	2736	1848	68	$C_{max} = 14$
Q_{4A}	–	–	–	–	2738	1845	64	$C_{max} = 13$

V The number of decision variables, *C* The number of constraints, *T* Solution time

*Feasible solution

**Calculation was stopped after 900 s

5 Conclusion

A concept of decision support for robust resource-constrained scheduling problems and the corresponding implementation platform are characterized by high flexibility and efficiency. The illustrative example given is only a small sample of the opportunities the implementation platform provides (Tables 2 and 3). Due to the use of the declarative environment, it offers an easy and convenient way to implement all sorts of questions. These may be specific, general, and logical questions, etc. They may relate to decision-making problems (answers: Yes/No), optimization problems (answers: optimum cost, optimum makespan, etc.) or a type of search problems (e.g. the parameters of the system configuration etc.). These capabilities are particularly important when making decisions under uncertainty and generating robust schedules.

The efficiency and performance of this approach have been studied using the job-shop scheduling problem with regard to the MP (Table 3).

We found that even this small example presented, it is 6–9 times faster relative to the MP for finding solutions.

Further studies will include modeling and solving of various types of problems in the area of: routing in computer networks, capacity vehicle routing [16], resource capacity project scheduling [17], multi-assortment repetitive production and other production problems in ERP [18], computer-aided machine tool selection [19], and multimodal processes [20] etc. using this programming implementation platform. It is also planned to implement the proposed implementation platform in the cloud [21].

Appendix 1: Sets of Facts for Illustrative Example

```
%machines_f (#P). - machines, #P - machine ID
machines_f(P1). machines_f(P2). machines_f(P3).
machines_f(P4). machines_f(P5). machines_f(P6).
%products_f(#U). - products, #U - product ID
products_f(U1). products_f(U2). products_f(U3).
products_f(U4). products_f(U5). products_f(U6).
products_f(U7). products_f(U8). products_f(U9).
products_f(U10).
%techn_f(#U,#P,D). - technology, #U - product ID, #P -
machine ID, D - the execution time of the product u on
the machine p
techn_f(U1,P1,2).  techn_f(U1,P2,2).  techn_f(U1,P3,1).
techn_f(U1,P4,1).  techn_f(U1,P5,1).  techn_f(U1,P6,1).
techn_f(U2,P1,1).  techn_f(U2,P2,1).  techn_f(U2,P3,1).
techn_f(U2,P4,1).  techn_f(U3,P3,2).  techn_f(U3,P4,2).
techn_f(U3,P5,1).  techn_f(U3,P6,1).  techn_f(U4,P2,1).
techn_f(U4,P3,1).  techn_f(U4,P4,2).  techn_f(U4,P5,2).
techn_f(U5,P1,1).  techn_f(U5,P2,1).  techn_f(U5,P3,2).
techn_f(U6,P3,1).  techn_f(U6,P4,1).  techn_f(U6,P5,2).
techn_f(U7,P4,2).  techn_f(U7,P5,2).  techn_f(U7,P6,2).
techn_f(U8,P4,1).  techn_f(U8,P5,1).  techn_f(U8,P6,2).
techn_f(U9,P1,1).  techn_f(U9,P2,1).  techn_f(U9,P3,2).
techn_f(U10,P4,1). techn_f(U10,P5,1). techn_f(U10,P6,2).
%resources_f (#S,F). - resources, #S additional resource
ID, F - the total number of additional resources s
resources_f(S1,20). resources_f(S2,20). re-
sources_f(S3,20).
resources_f(S4,20).
%alloc_f(#U,#P,#S) - allocation, #U - product ID, #P -
machine ID, #S additional resource ID
alloc_f(U1,P1,S1).   alloc_f(U1,P2,S1).
alloc_f(U1,P3,S1).
alloc_f(U1,P4,S2).   alloc_f(U1,P5,S2).
alloc_f(U1,P6,S2).
alloc_f(U2,P1,S1).   alloc_f(U2,P2,S1).
alloc_f(U2,P3,S1).
alloc_f(U2,P4,S2).   alloc_f(U3,P3,S1).
alloc_f(U3,P4,S2).
alloc_f(U3,P5,S2).   alloc_f(U3,P6,S2).
alloc_f(U4,P2,S1).
alloc_f(U4,P3,S1).   alloc_f(U4,P4,S2).
alloc_f(U4,P5,S2).
alloc_f(U5,P1,S1).   alloc_f(U5,P2,S1).
```

```
alloc_f(U5,P3,S1).
alloc_f(U6,P3,S1).   alloc_f(U6,P4,S2).
alloc_f(U6,P5,S2).
alloc_f(U7,P4,S2).   alloc_f(U7,P5,S2).
alloc_f(U7,P6,S2).
alloc_f(U8,P4,S1).   alloc_f(U8,P5,S2).
alloc_f(U8,P6,S2).
alloc_f(U9,P1,S1).   alloc_f(U9,P2,S1).
alloc_f(U9,P3,S1).
alloc_f(U10,P4,S2).  alloc_f(U10,P5,S2).
alloc_f(U10,P6,S2).
alloc_f(U1,P1,S3).   alloc_f(U1,P2,S3).
alloc_f(U1,P3,S3).
alloc_f(U1,P4,S4).   alloc_f(U1,P5,S4).
alloc_f(U1,P6,S4).
alloc_f(U2,P1,S3).   alloc_f(U2,P2,S3).
alloc_f(U2,P3,S3).
alloc_f(U2,P4,S4).   alloc_f(U3,P3,S3).
alloc_f(U3,P4,S4).
alloc_f(U3,P5,S4).   alloc_f(U3,P6,S4).
alloc_f(U4,P2,S3).
alloc_f(U4,P3,S3).   alloc_f(U4,P4,S4).
alloc_f(U4,P5,S4).
alloc_f(U5,P1,S3).   alloc_f(U5,P2,S3).
alloc_f(U5,P3,S3).
alloc_f(U6,P3,S3).   alloc_f(U6,P4,S4).
alloc_f(U6,P5,S4).
alloc_f(U7,P4,S4).   alloc_f(U7,P5,S4).
alloc_f(U7,P6,S4).
alloc_f(U8,P4,S3).   alloc_f(U8,P5,S4).
alloc_f(U8,P6,S4).
alloc_f(U9,P1,S3).   alloc_f(U9,P2,S3).
alloc_f(U9,P3,S3).
alloc_f(U10,P4,S4).  alloc_f(U10,P5,S4).
alloc_f(U10,P6,S4).
%precedence_f(#U,#P1,#P2). - precedence, #U - product ID,
#P1 - first machine ID, #P - second machine ID
precedence_f(U1,P1,P2).  precedence_f(U1,P2,P3).
precedence_f(U1,P3,P4).  precedence_f(U1,P4,P5).
precedence_f(U1,P5,P6).  precedence_f(U2,P1,P2).
precedence_f(U2,P2,P3).  precedence_f(U2,P3,P4).
precedence_f(U3,P3,P4).  precedence_f(U3,P4,P5).
precedence_f(U3,P5,P6).  precedence_f(U4,P5,P4).
```

```
precedence_f(U4,P4,P3).  precedence_f(U4,P3,P2).
precedence_f(U5,P1,P2).  precedence_f(U5,P2,P3).
precedence_f(U6,P3,P4).  precedence_f(U6,P4,P5).
precedence_f(U7,P4,P5).  precedence_f(U7,P5,P6).
precedence_f(U8,P6,P5).  precedence_f(U8,P5,P4).
precedence_f(U9,P1,P2).  precedence_f(U9,P2,P3).
precedence_f(U10,P4,P5). precedence_f(U10,P5,P6).
orders_f(#U,O).- orders, #U - product ID, O -  the size
of the order for product o
orders_f(U1,1). orders_f(U2,1). orders_f(U3,1).
orders_F(U4,1). orders_f(U5,1). orders_f(U6,1).
orders_f(U7,1). orders_F(U8,1). orders_f(U9,1).
orders_f(U10,1).
```

References

1. Schrijver, A.: Theory of Linear and Integer Programming. Wiley, New York (1998)
2. Rossi, F., Van Beek, P., Walsh, T.: Handbook of Constraint Programming (Foundations of Artificial Intelligence). Elsevier Science Inc, New York (2006)
3. Apt, K., Wallace, M.: Constraint Logic Programming using Eclipse. Cambridge University Press, Cambridge (2006)
4. Bocewicz, G., Banaszak, Z.: Declarative approach to cyclic steady states space refinement: periodic processes scheduling. Int. J. Adv. Manuf. Technol. **67**(1–4), 137–155 (2013)
5. Sitek, P., Wikarek, J.: A hybrid approach to the optimization of multiechelon systems. In: Mathematical Problems in Engineering, Article ID 925675. Hindawi Publishing Corporation (2014). doi:10.1155/2014/925675
6. Sitek, P., Nielsen I.E., Wikarek, J.: A hybrid multi-agent approach to the solving supply chain problems. In: Procedia Computer Science KES, pp. 1557–1566 (2014)
7. Sitek, P., Wikarek, J.: A hybrid framework for the modelling and optimisation of decision problems in sustainable supply chain management. Int. J. Prod. Res. **53**(21), 6611–6628 (2015). doi:10.1080/00207543.2015.1005762
8. Milano, M., Wallace, M.: Integrating operations research in constraint programming. Ann. Oper. Res. **175**(1), 37–76 (2010)
9. Achterberg, T., Berthold, T., Koch, T., Wolter, K.: Constraint integer programming, a new approach to integrate CP and MIP. Lect. Notes Comput. Sci. **5015**, 6–20 (2008)
10. Guyon, O., Lemaire, P., Pinson, Ă., Rivreau, D.: Solving an integrated job-shop problem with human resource constraints. Ann. Oper. Res. **213**(1), 147–171 (2014)
11. Blazewicz, J., Lenstra, J.K., Rinnooy Kan, A.H.G.: Scheduling subject to resource constraints: classification and complexity. Discrete Appl. Math. **5**, 11–24 (1983)
12. Lawrence, S.R., Morton, T.E.: Resource-constrained multi-project scheduling with tardy costs: comparing myopic, bottleneck, and resource pricing heuristics. Eur. J. Oper. Res. **64**(2), 168–187 (1993)
13. Sitek, P.: A hybrid CP/MP approach to supply chain modelling, optimization and analysis. In: Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, pp. 1345–1352. doi:10.15439/2014F89 (2014)

14. Eclipse, Eclipse—The Eclipse Foundation open source community website. Accessed May 4, www.eclipse.org (2015)
15. Lindo Systems INC, LINDO™ Software for Integer Programming, Linear Programming, Nonlinear Programming, Stochastic Programming, Global Optimization, Accessed May 4, www.lindo.com (2015)
16. Toth, P., Vigo, D.: Models, relaxations and exact approaches for the capacitated vehicle routing problem. Discrete Applied Mathematics **123**(1–3), 487–512 (2002)
17. Coelho, J., Vanhoucke, M.: Multi-mode resource-constrained project scheduling using RCPSP and SAT solvers. Eur. J. Oper. Res. **213**, 73–82 (2011)
18. Krenczyk D., Jagodzinski J.: ERP, APS and simulation systems integration to support production planning and scheduling. In: Advances in Intelligent Systems and Computing, vol. 368, pp. 451–461. Springer International Publishing (2015)
19. Gola, A., Świeć, A.: Computer-aided machine tool selection for focused flexibility manufacturing systems using economical criteria. Actual Prob. Econ. **10**(124), 383–389 (2011)
20. Bocewicz, G., Nielsen, P., Banaszak, Z., Dang, V.Q.: Cyclic steady state refinement: multimodal processes perspective. In: IFIP Advances in Information and Communication Technology, vol. 384, pp. 18–26. AICT (2012)
21. Bąk, S., Czarnecki R., Deniziak S.: Synthesis of real-time cloud applications for internet of things. Turk. J. Electr. Eng. Comput. Sci. doi:10.3906/elk-1302-178 (2013)

The Experimental Evaluation of Rules Partitioning Conception for Knowledge Base Systems

Roman Simiński

Abstract This article presents the theoretical background for implementation of *KBExplorer* software package and the summary of the empirical study focused on the evaluation of this software on large, real-word knowledge bases. *KBExplorer* package is the own, originally designed software which provides most of the expert system shell's common functions. The fundamental part of such software is the *KBExpertLib* library. This library allows to build domain expert systems using Java programming language. The first part of experiments was focused on the effectiveness of rules partition algorithm and estimation of the memory occupancy for additional data necessary for storing information about rules groups. The effectiveness evaluation of the forward and backward inference algorithms was the main goal of the second part of the experiments.

Keywords Rule knowledge base · Inference algorithms implementation · Rules partitioning strategies · Expert systems

1 Introduction

Regardless of the development of different nonknowledge-based methods, the rule representation and rules-oriented inference are still popular in the real-world applications of decision support systems [1]. Recent years have brought a renaissance of rules representation for knowledge bases. Currently, the rules are considered as standard result form of data mining methods, rules are again an important and useful material for constructing knowledge bases for different types of decision support systems [9]. This work is the part of research focused on the development of the new method and tools for building knowledge-based decision support systems. This article introduce a short description of the theoretical background for KBExplorer software package as well as the report of empirical study focused on

R. Simiński (✉)
Institute of Computer Science, Silesian University, Sosnowiec, Poland
e-mail: roman.siminski@us.edu.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_7

evaluation of realized software performed on the large, real-word knowledge bases. KBExplorer package is the own,[1] originally designed software which provides most of the expert system shell's common functions. The fundamental part of such software is the KBExpertLib library. This library allows to build domain expert systems using Java programming language.

The main goal of the experiments was the practical evaluation of two selected aspects of the rules partitions conception. The first part of experiments was focused on the effectiveness evaluation of rules partition creation, also was performed the estimation of the memory occupancy for additional data necessary for storing information about rules groups. The effectiveness evaluation of the inference algorithms was the goal of the second part of the experiments. This work is focused on the performance study of own software, mainly *KBExpertLib* library. Experimental evaluation allows the author to detect algorithmic and implementation weakness of designed software, the results will allow to formulate the schedule of the improvement works. An important question is whether the obtained results are acceptable for practical application in the expert systems area. The next step of the research will be focused on the comparative studies of *KBExpertLib* and *JESS*—both packages are dedicated for Java programming language. For this reason, the work does not refer to the other solutions, as JESS, CLIPS, DROOLS [2, 3, 5, 6], comparative study will be conducted in the future works.

The second section briefly presents theoretical background of the rules partitioning conception. The third section describes the application of partitioning conception—*KBExplorator* software family. In the Sect. 4 conducted experiment are described, the last section contains discussion of experiments results and concluding remarks.

2 Methods

2.1 Preliminary Issues

The knowledge base is a pair $KB = (RS, FS)$ where RS is a non-empty finite set of rules and FS is a finite set of facts. $RS = \{r_1, \ldots, r_n\}$, each rule $r \in RS$ will have a form of Horn's clause: r: $p_1 \wedge p_2 \wedge \ldots \wedge p_m \rightarrow c$, where m—the number of literals in the conditional part of rule r, and $m \geq 0$, p_i—i-th literal in the conditional part of rule r, $i = 1 \ldots m$, c—literal of the decisional part of rule r.

For each rule $r \in RS$ we define following the functions: $concl(r)$—the value of this function is the conclusion literal of rule r: $concl(r) = c$; $cond(r)$—the value of this function is the set of conditional literals of rule r: $\{p_1, p_2, \ldots p_m\}$, $literals(r)$—the value of this function is the set of all

[1]The prototype version of *KBExplorer* and the demo version of *KBExplorerDesktop* are availabe online: http://kbexplorer.ii.us.edu.pl.

literals of rule r: $literals(r) = cond(r) \cup \{concl(r)\}$, $csizeof(r)$—conditional size of rule r, equal to the number of conditional literals of rule r ($csizeof(r) = m$): $csizeof(r) = |cond(r)|$, $sizeof(r)$—whole size of rule r, equal to the number of conditional literals of rule r increased by the 1 for single conclusion literal, for rules in the form of Horn's clause: $sizeof(r) = csizeof(r) + 1$. We will also consider the facts as clauses without any conditional literals. The set of all such clauses f will be called set of facts and will be denoted by FS: $F = \{f: \forall_{f \in FS}\ cond(f) = \{\} \wedge f = concl(f)\}$.

Any arbitrarily created subset of rules $R \in 2^{RS}$ will be called a *group of rules*. In this work we will discuss specific subset $PR \subseteq 2^{RS}$ called partition of rules. Any partition PR is created by partitioning strategy, denoted by PS, which defines specific content of groups of rules $R \in 2^{RS}$ creating a specific partition of rules PR. We may consider many partitioning strategies for a single rule base, in this work rules partitions terminologically correspond to the mathematical definition of the partition as a division of a given set into the non-overlapping and non-empty subset. The groups of rules which create partition are pairwise disjoint and utilize all rules from RS. Each partitioning strategy PS for rules set RS generates the partition of rules $PR \subseteq 2^{RS}$: $PR = \{R_1, R_2, \ldots, R_k\}$, where: k—the number of groups of rules creating the partition PR, R_i—i-th group of rules, $R \in 2^{RS}$ and $i = 1,\ldots, k$.

2.2 Partitioning Strategies

The partition strategies for rule based knowledge bases are divided into two categories:

1. *Simple strategies*—the membership criterion decides about the membership of rule r in a particular group $R \subseteq PR$ according to the membership function mc. Simple strategy let us divide the rules by using the algorithm with time complexity not higher than $O(n \cdot k)$, where $n = |RS|$ and $k = |PR|$. Simple strategy creates final partition PR by a single search of the rules set RS and allocation of each rule r to the proper group R, according to the value of the function $mc(r, R)$.
2. *Complex strategies*—the particular algorithm decides about the membership of the rule r in some group $R \subseteq PR$, with time complexity typically higher than any simple partition strategy. Complex strategies usually do not generate final partition in a single step. Complex partitioning strategies will not be discussed in this work. An example of a complex strategy is described in the [8].

Creation of simple partition for rules set requires the definition of the membership criteria which assigns particular rule $r \in R$ to the given group of rules $R \subseteq PR$. Proposed approach assumes that the membership criteria will be defined by the mc function, which is defined individually for every simple partition strategy. The function: $RS \times PR \rightarrow [0\ldots1]$ has the value 1 if the rule $r \in RS$ with no doubt

belongs to the group $R \subseteq PR$, 0 in the opposite case. The value of the function from the range $0 < mc < 1$ means the partial membership of the rule r to the group R.

The method of determining its value and its interpretation depends on the specification of a given partition method. It is possible to achieve many different partitions of rule base using single mc function. Let us to assume that threshold value $0 \leq T \leq 1$ exists. The value of the $mc(r, R)$ function can be higher, higher or equal, equal, less, less or equal to the T value. Generally, including the symbol op indicating any operator from the set $\{>, \geq, <, \leq, =, \neq\}$, we can define simply partition of rule based knowledge base PR as follows: $PR = \left\{R: R \in 2^{RS} \wedge \forall\, r \epsilon R\, mc(r, R) \geq T\right\}$.

The algorithm of creating the partition which bases on simple strategy is presented in the pseudo-code below. The input parameters are: knowledge base RS, the function mc that defines the membership criteria and the value of the threshold T. Output data is the partition PR. Time complexity of such algorithm is $O(n{\cdot}k)$, where $n = |R|$, $k = |PR|$.

For each rule $r \in RS$ we have to check whether the goal partition PR contains the group R with rule r (the value of the mc function has to be at least T: $mc(r, R) \geq T$). If such a rule doesn't exist the given rule r becomes the seed of a new group which is added to the created partition PR.

```
Alg01:  The simple partition algorithm
Input:  RS, mc, T;
Output: PR={R1, R2,…, Rk};
procedure createPartitions( RS, var PR, mc, T )
var R, r;
begin
  PR = {};
  forall r∈RS do
    if exists R∈PR : mc( r, R )≥ T) then
      R = R ∪ r;
    else
      R = {r};
      PR = PR ∪ R;
    endif
  endfor
end
```

Partitions generated by the algorithm are unique and complete—each rule $r \in RS$ is considered, and qualified to one group R, that matches the criteria of mc. In order to obtain decision oriented [11] partitions *createPartitions* procedure should be invoked with threshold $T = 1$ and the mc function defined in the following way:

$$mc(r,R) = \begin{cases} 1 & if\ \forall r_i \in R\, concl\,(r_i) = concl(r) \\ 0 & otherwise \end{cases} \quad (1)$$

Special case of the simple strategies is the specific partition strategy called selection. The selection divides the set of rules *RS* into the two subsets *R* and *RS-R*. All rules from *R* fulfill the membership criteria for some partition strategy *PS*, and all other rules do not meet such criteria. Thus we achieve the partition *PR* = {*R*, *RS-R*}. In practical sense, selection is the operation with linear time complexity O(*n*) where *n* denotes the number of all rules in knowledge base. The selection algorithm is simple and will be omitted.

3 Application of Partitioning Strategies

The main results of the mentioned above conception of rules partitions are modified forward and backward algorithms [7, 8] as well as the proposals for new algorithms [11, 12]. The proposed modifications of the classical inference algorithms are based on information extracted from the groups of rules generated by the selected partitioning strategy [10]. The theoretical background presented above is the base for software implementation. At the current stage of development process, three main software components are implemented—*KBExplorator*, *KBExploratorDesktop* and *KBExpertLib*.

The *KBExplorator* is a web application. It allows the user to create, edit and share rule knowledge bases (Fig. 1). Each user can register in the system own account, knowledge bases created by the users are stored in the *KBExplorator* data bases and they are accessible from any standard web browser software. Stored knowledge base may be shared between registered system's users, it is also possible to download any stored knowledge base as the XML file. Actually system is only available as simple demo in polish version, the new, multi-language version is under development.

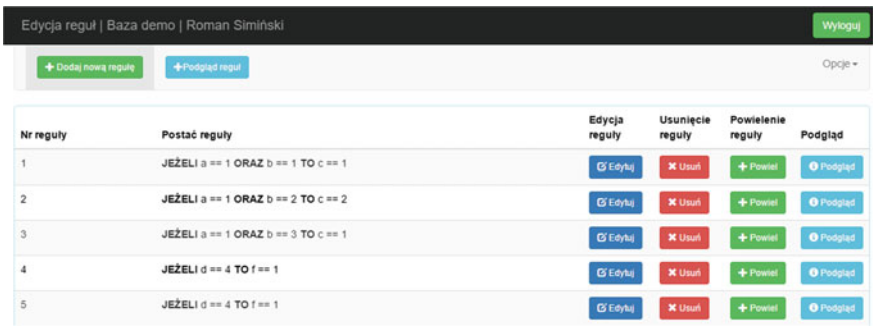

Fig. 1 The *KBExplorator*—rules editor, demo version (in Polish)

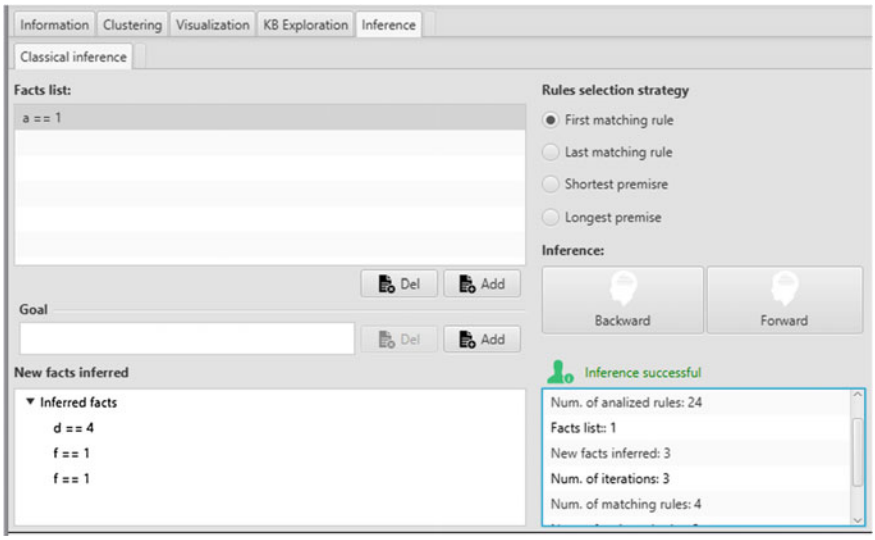

Fig. 2 The *KBExploratorDesktop*—classical forward inference results

The *KBExploratorDesktop* is the classical desktop application, which utilizes rule knowledge bases stored in the the *KBExplorator* database or saved locally in the XML files. *KBExploratorDesktop* is implemented using the *KBExpertLib*—the software library, which allows the programmers to use different kinds of inference within any software projects implemented in Java programming language. This library is able to run different kinds of forward and backward inference algorithms. The *KBExploratorDesktop* is implemented as the *JavaFX* program. In addition to mentioned before different kinds of inference (Fig. 2), the *KBExploratorDesktop* allows the user to perform rules clustering, the density-based and AHC/mAHC clustering methods are used [7]. System also provides tools for whole rule base structure view and visualization of rules groups [8].

4 Experiments

This chapter presents the experiments performed on the real-word knowledge bases. The main goal of experiments was the practical evaluation of two selected aspects of rules partitions conception. The first part of experiments was focused on the effectiveness of rules partition creation and the estimation of memory occupation for additional data necessary for storing information about rules groups. The effectiveness evaluation of inference algorithms is the goal of the second part of experiments.

Four real-word rule knowledge bases were used:

1. *eval416*—knowledge base for effectiveness evaluation of sales representatives, basic version, 416 rules.
2. *eval1199*—knowledge base for effectiveness evaluation of sales representatives, full version, 1199 rules.
3. *bud4438*—knowledge base for decision support for houses building, 4438 rules [4].
4. *bud22190*—knowledge base bud4458 duplicated five times with random modifications (because gaining access to large, real-world knowledge base is very difficult), 22,190 rules.

4.1 Partitioning Effectiveness and Memory Occupation

The main goal of first experiments was to evaluate the time efficiency of loading the rules from XML files into the internal, object oriented data structures (defined in the package *kbcore*) as well as the estimation of memory occupation for such data structures. Data structures defined in the *kbcore* package are necessary for inference, data structures defined in the *kbpartitions* package are dedicated for algorithm which use rules partitions.

All experiments were run on a typical PC desktop computer: Intel i5 2.5 GHz processor, 16 GB of RAM, classical mechanical hard disk, 64-bit Windows 10 operating system. Each experiment was repeated at least ten times, results were averaged. The Tables 1 and 2 presents final results of the experiments.

The time taken to perform knowledge base loading from XML files is satisfactory only for small bases. For knowledge base counting 22,190 rules loading time

Table 1 The average times of knowledge base loading and partioning

Knowledge base	Number of rules	Base loading time (s)	Partitioning time (ms)
eval416	416	0.312	11.34
eval1199	1119	1.487	23.11
bud4438	4438	27.39	112.87
bud22190	22,190	670.83	224.55

Table 2 Memory usage for data structures

Knowledge base	Number of rules	XML file size (KB)	Memory usage (B)	
			kbcore package	kbpartition package
eval416	416	480	95,964	22,538
eval1199	1119	1531	285,316	24,952
bud4438	4438	6797	119,7148	372,632
bud22190	22,190	27,241	4,646,348	382,016

exceeds 11 min and is not acceptable real-word usage. The time of the partitioning performance is less than 0.2 s even for largest knowledge base counting 22,190 rules. For relatively small knowledge bases (less than 4000 rules) partitioning time not exceeds 0.01 s (Table 1). Data structures necessary for storing knowledge information occupy acceptable amount of memory—even for largest knowledge data structures consume less than 5 MB of RAM. Additional data for storing information about rules partition are small and not exceed 400 KB (Table 2).

4.2 Inference Effectiveness Evaluation

In the first experiment two version of forward inference algorithm were considered—classic depth-first and modified breadth-first algorithm, implemented with use of rules partition selection. Experiment was performed on the same data and hardware configuration as experiment described in the previous section. For each base starting fact set was prepared individually and it contained facts necessary for activation all rules in knowledge base. In this way algorithms exhaustively activated all rules in knowledge base. Table 3 presents the average results of experiments.

When the forward inference is considered, even for the largest knowledge base, the average inference time was less than one second, for small knowledge bases the time taken for forward inference is near 0.2 s. For the analyzed bases the average time for breadth-first algorithm is 50 % better than time for depth-first algorithm. It is worth to notice that better results for breadth-first algorithm is caused by the properties of such algorithm. The implementation with usage of rules partition is entirely secondary and it has only technical meaning.

In the next experiment the two version of backward inference algorithms were examined. First, classical backward inference and modified algorithm described in the [11]. Backward inference is typically performed in the interactive mode—user confirms facts, thus the time efficiency is not the crucial characteristic for backward inference. The evaluation of the backward inference algorithms has been focused on comparison of two parameters:

- the number of analyzed rules, necessary to achieve inference goals,
- the number of unnecessary recursive calls of inference algorithm.

Table 3 The average times of forwards inference

Knowledge base	Rules count	Forward inference time		
		(1) Depth-first algorithm (s)	(2) Breadth-first algorithm (s)	(2) vs (1) (%)
eval416	416	0.0812	0.0415	48
eval1199	1119	0.1386	0.0782	43
bud4438	4438	0.3723	0.1869	49
bud22190	22,190	0.9345	0.6014	35

Table 4 An example results of two version of backward inference comparisions

Number of:	eval1199		bud4438	
	Classic	Partitions	Classic	Partitions
Analyzed rules	61,149	146	39,942	22,220
Goal matching rules	146	146	62	62
Activated matching rules	51	30	9	6
Recursive calls	50	29	8	5
Unnecessary recursive calls	21	0	6	3

Experiments were performed for all possible goals, for each base fact set was prepared individually to exhaustively activation all rules matched to the goal. Sample results for two selected knowledge bases and particularly selected goals are presented in the Table 4.

Analyzed knowledge bases cause the big differences between the optimistic and pessimistic case of the inference process optimization. The results of backward inference experiments depend on various factors—selected goal, internal structure of knowledge base, dependences between rules, depth of the inference chain. Thus, the expected reduction of analyzed parameters is difficult to predict. When the knowledge base may be divided into the relatively large number of groups, counting small number of the rules, it is possible to it reduce the number of analyzed rules and unnecessary recursive calls (results Table 4). When the rule base does not contain subgoals and it can't be split into the groups, the results of modified backward inference are the same as for the classic version.

5 Summary

An article describes the practical evaluation of the idea of rule knowledge bases decomposition into the groups of rules, called rules partitions. Proposed method of decomposition and featured knowledge base model were implemented in the *KBExplorator* system. In this work the *KBExploratorDesktop* was examined—it is the classical desktop application implemented in Java with usage *KBExpertLib*. Experiments were conducted on the real-word, large knowledge bases. The main goal of experiments was practical evaluation of selected aspects of rules partitions conception. The first group of experiments was focused on the effectiveness of rules partition creation and memory occupancy for additional data necessary for storing information about rules groups. The effectiveness evaluation of inference algorithms was the goal of the second group of experiments.

The results of experiments are promising. The time of the partitioning performance is less than 0.2 s even for largest knowledge base counting 22,190 rules. For relatively small knowledge bases (less than 4000 rules) partitioning time not exceeds 0.01 s. Data structures necessary for storing knowledge information occupy acceptable amount of memory—even for largest knowledge data structures

consume less than 5 MB of RAM. Additional data for storing information about rules partition are small and not exceed 400 KB. Knowledge base partitioning effectiveness and memory requirements can be evaluated as satisfactory. Unfortunately, the time taken to perform knowledge base loading from XML files is satisfactory only for small bases. For knowledge base counting 22,190 rules loading time exceeds 11 min and is not acceptable real-word usage. The loading process needs improvement—optimization of XML parsing is considered as well as different methods of physical knowledge base format (serialization and binary file representation).

When the forward inference is considered, even for the largest knowledge base, the average inference time was less than one second, for small knowledge bases the time taken for forward inference is near 0.2 s. For the analyzed bases the average time for breadth-first algorithm is 50 % better than time for depth-first algorithm. It may be considered as acceptable for most typical applications of knowledge based systems. It is worth to notice that better results for breadth-first algorithm is caused by the properties of such algorithm. The implementation with usage of rules partition is entirely secondary and it has only technical meaning. For forward inference we propose idea of creation rules partition based on similar premises and searching only representatives of them but this approach was not considered in this work (see [8, 12]). The results of backward inference experiments depend on various factors—selected goal, internal structure of knowledge base, dependences between rules, depth of the inference chain. When the knowledge base may be divided into the relatively large number of groups, counting small number of the rules, it is possible to it reduce the number of analyzed rules and unnecessary recursive calls (results Table 4). When the rule base does not contain subgoals and it can't be split into the groups, the results of modified backward inference are the same as for the classic version.

The obtained results are promising regarding the use of the proposed software on mobile devices and as the Web service. The next step of the research will be focused on the comparative studies. The comparison of *KBExpertLib* and *JESS* is planned—both packages are dedicated for Java. The implementation of *RETE* [3, 5] algorithm is also assumed, it will allow to perform a more detailed comparative experiments on real word knowledge base.

Acknowledgments This work is a part of the project "Exploration of rule knowledge bases" founded by the Polish National Science Centre (NCN: 2011/03/D/ST6/03027).

References

1. Akerkar, R., Sajja, P.: Knowledge-Based Systems. Jones & Bartlett Learning (2010)
2. Browne, P.: JBoss Drools Business Rules. Packt Publishing (2009)
3. Forgy, C.: Rete: a fast algorithm for the many patterns/many objects match problem. Artif. Intell. **19**, 17–37 (1982)

4. Jach, T., Xięski T.: Inference in expert systems using natural language processing. Beyond Databases, Architectures, and Structures, Communications in Computer and Information Science, vol. 521, pp. 288–298. Springer (2015)
5. Luc, A.: Average case complexity analysis of RETE pattern-match algorithm and average size of join in Databases. Foundations of Software Technology and Theoretical Computer Science, Lecture Notes in Computer Science, vol. 405, pp. 223–241. Springer, Berlin (1989)
6. Nalepa, G., Ligeza, A., Kaczor, K.: Overview of Knowledge Formalization with XTT2 Rules, Rule-Based Reasoning, Programming, and Applications, LNCS, vol. 6826, pp. 329–336. Springer (2011)
7. Nowak-Brzezińska, A., Simiński, R.: Knowledge mining approach for optimization of inference processes in rule knowledge bases. LNCS, vol. 7567, pp. 534–537, Springer (2012)
8. Nowak-Brzezińska, A., Simiński, R.: New inference algorithms based on rules partition. In: CS&P, The 23rd International Workshop on Concurrency, Specification and Programming (CS&P 2014). Informatik-Berichte, vol. 245, Humboldt-University, edited by Louchka Popova-Zeugmann, Chemnitz, Germany 2014, pp. 164–174
9. Polkowski, L. (ed.): Rough sets in knowledge discovery 2: applications, case studies and software systems, vol. 19. Physica (2013)
10. Simiński, R.: Multivariate approach to modularization of the rule knowledge bases. Man–Machine Interactions 4. In: 4th International Conference on Man-Machine Interactions, ICMMI 2015, Kocierz Pass, Poland, October 6–9, 2015. A. Gruca, A. Brachman, S. Kozielski, T. Czachórski, (eds.) Man-Machine Interactions 4, Series: Advances in Intelligent Systems and Computing, Volume 391, Springer International Publishing 2016, pp. 473-483
11. Simiński, R., Nowak-Brzezińska, A.: Goal-driven Inference for Web Knowledge Based System. Information Systems Architecture and Technology. In: Proceedings of 36th International Conference on Information Systems Architecture and Technology—ISAT 2015—Part IV. Wilimowska, Z., Borzemski, L., Grzech, A., Świątek, J. (eds.). Series: Advances in Intelligent Systems and Computing, vol. 432, pp. 99–109
12. Simiński, R., Wakulucz-Deja, A.: Rough Sets Inspired Extension of Forward Inference Algorithm. In: Proceedings of the 24th International Workshop CSP 2015: Concurency, Specification & Programming, Rzeszów, Poland, September 28–30 2015. Suraj, Z., Czaja, L. (eds.) ISBN 978-83-7996-181-8, vol. 2, pp. 161–172

Part II
Project Management

Effective Project Management Using Web-Based System with Integrated Knowledge Base

František Babič and Barbora Čurová

Abstract The project management represents an important approach how to manage goal-oriented collaborative activities effectively in the specified constraints as time, costs and quality. The motivation of the presented work was to design a new web-based system to support a project management with some traditional features and some new ones, e.g. a knowledge base. The target group includes small companies dealing with customer's orders and tasks distribution between the internal and external staff (e.g. a developer, a graphic designer, a SEO analyst), and use the commission system of remuneration. We proposed our solution based on a detailed analysis of the existing competitive solutions through predefined criteria and identified a list of user requirements collected in a close cooperation with potential customers. The result represents a customised web-based system offering an intuitive and user-friendly user interface with a comprehensive list of features from which the implemented knowledge base stands out. We used the user acceptance testing to verify a satisfaction with the implemented solution. The result was more than 93 % user satisfaction, a knowledge base marked as the most used and beneficial function.

Keywords Project management · Web-based system · Knowledge base

1 Motivation

The project management is the application of knowledge, skills, tools and techniques to project activities to meet or exceed stakeholder needs and expectations from a project, defined by the Project Management Body of Knowledge [3].

F. Babič (✉) · B. Čurová
Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9/B, 042 00 Košice, Slovakia
e-mail: frantisek.babic@tuke.sk

B. Čurová
e-mail: barbora.curova@tuke.sk

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_8

The effective flow of data and communication in the project team is essential to ensure the successful management of the project and we can provide it by a suitable ICT solution [1]. Our motivation to propose a new system for this purpose consisted of several factors. At the beginning, there was an idea to create a simple task manager for small companies dealing with customer's orders and tasks distribution between the internal and external staff (e.g. a developer, a graphic designer, a SEO analyst). This type of a company is relatively common in the domain of web technologies, web application or web projects and is characterised by the higher percentage of the external staff. This fact creates a need for effectively task distribution and the commission system of remuneration. We can find some possible alternatives, but in many cases these systems contain a large set of tools or functions that users will not use. In addition, the second aspect is that despite the current orientation of the existing solutions to the cloud some companies want to have their internal data under its administration. Finally, in practice, there is a lack of systems that would provide companies and their employees (especially the external ones) with sufficient modularity (proprietary functions), data supervision (a proprietary database), and the possibility to keep the remuneration and piece wage-related records directly. The aim of the proposed system is to provide the functions missing on the market as an easy to use and intuitive solution. In addition, we implemented and tested our solution in a close cooperation with a potential customer that expressed their interest in using it in the future.

We organise the paper as follows: at first, we described the motivation and relevant state of the art to identify the reason why it was a good idea to design and implement a new web-based system for supporting the project management. The next sections are devoted to the design and implementation of the proposed system containing a knowledge base first designed as an extended feature but finally evaluated by the users as the most beneficial. Description of the provided user acceptance testing and a short summary concludes the paper.

1.1 State of the Art

Systems supporting the project management are now an integrated part of a number of companies, not only in the domain of information and communication technologies. They enable a more efficient management of the project team; a goal's decomposition into particular tasks; project, quality and cost monitoring; and many other features necessary for the success and the achievement of the specified project's goals. The world market is relatively saturated with these solutions and they differ from one another in the character of the platform, offered features, security, and the level of usability. Before developing a customised solution, some existing systems available as on-line platforms were analysed, because currently the use of online solutions without the need to install a system as a desktop application represents a widespread trend. We installed and tested each of the compared

systems (TeamWork,[1] HiTask,[2] Zoho,[3] Wrike,[4] Bitrix24 system[5]) with the following criteria:

1. The system offers some basic administration for the project management.
2. The system offers some extensions to the basic administration.
3. The system can be used free of charge.
4. The system has a minimum standard developer support.
5. The manager can allocate remuneration for the tasks.
6. The manager can generate useful visual statistics.
7. The manager has access to a database containing details of their projects, tasks or team members.
8. The manager can prioritise tasks and monitor their status and progress.
9. The manager can generate customised reports.
10. The user can store a solution for occurred problems during the task realisation.

None of these systems provides a full-scale database supervision for the users, which may result in the client's distrust towards the system provider, especially when the client uses sensitive data (passwords, names, numerical identifiers, etc.). In addition, all of them provide either no or only a limited possibility for payroll management (see Table 1). Based on the collected information we can conclude that they do not meet the specified requirements. It opens a space for a new customised solution.

1.2 Knowledge Management

Knowledge management (KM) which allows reusing the previous problem solving and decision-making experiences to improve organisational processes can be helpful in the domain of the project management too [1, 5, 6]. From a process perspective, a traditional KM model can be applied in the project environment to provide an effective framework for management and transformation of the knowledge [7]. This model contains the following four processes: creation, capturing, transfer and reusing [10]. Sokhanvara et al. discussed their adoption in the project management practices [11]. Muñoz-Avila and collective of authors described a knowledge-based project-planning tool, which they used to support a proposal of new project parts based on the stored previous cases [8]. Kasvi, Vartiainen and Hailikari dealt with an identification of the critical knowledge for the project-oriented organisations [4]. One of the main findings was a need for

[1] https://www.teamwork.com/.
[2] https://hitask.com/.
[3] https://www.zoho.com/.
[4] https://www.wrike.com/.
[5] https://www.bitrix24.com/.

Table 1 Evaluation of the existing systems in terms of the specified criteria

	1	2	3	4	5	6	7	8	9	10
TeamWork	x		x	x		x		x		x
HiTask	x	x	x	x	x			x		x
Zoho	x			x				x	x	x
Wrike	x			x						x
Bitrix24	x		x	x	x			x	x	x

systematic project knowledge management if the organisation wants to distill the results and lessons from one project and deliver them into another. The study [2] was devoted to the identification of various factors that influence the success or failure of KM initiatives in project-based companies. The authors identified the absence of incentives for employees participated in the analysed KM initiatives as the most significant barrier to their success adoption. Based on this short overview we can conclude that it is important to support four basic knowledge processes during project-oriented activities, e.g. through the effectively managed knowledge base containing information about the occurred incidents during implementation, testing, deployment or maintenance of ICT solutions.

2 Web-Based Project Management System

To support the possible deployment in various conditions, we did not use any Content management system. We used only freely available technologies, frameworks, and libraries.

2.1 Proposal

At first, we created a list of relevant user requirements based on the results of the comparison analysis described above and identified real needs of the participated company (we present a proposed architecture in Fig. 1):

- User management builds on user's login and password.
- Two user interfaces, one for managers and one for employees.
- Possibility to add a new task.
- Possibility to visualise customised reports and statistics.
- Modern and professional design.

During the implementation phase, we extended this basic list with the following requirements:

- Possibility to explore and modify already created tasks.
- Possibility to generate an invoice.

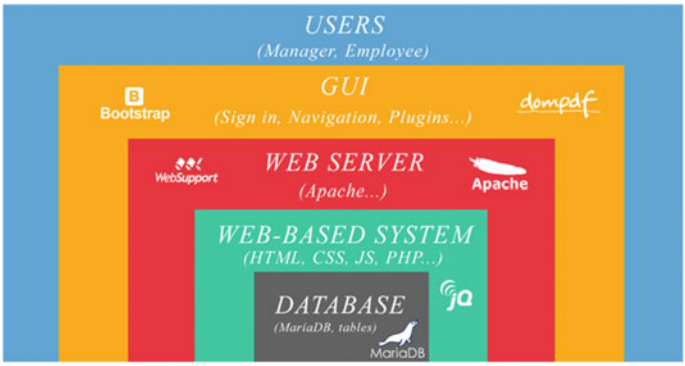

Fig. 1 Architecture of the proposed system

- Possibility to categorize the tasks.
- Possibility to save and re-use information about customers for invoice generation.
- Possibility to communicate with other members of the project team.

2.2 Final Version

This section describes all the implemented functions to meet the specified objectives and to ensure the benefits of the system's use. All functions are available to logged users; none of the sections is visible to common visitors. The system is available only in the Slovak language, but one of our plans is to extend a selection of the languages. However, this extension strongly depends on existing demand.

Adding new tasks is a basic unit of the system to support the project work. Only managers have this function assigned. Each task is described by the following parameters: name ("názov"), category ("kategória"), description, remuneration in EUR ("odmena"), responsible person ("pracovník"), deadline, priority ("priorita": high, medium, low), status (new, in progress, waiting, finished). A simple example of the task list is presented in Fig. 2.

The dashboard (Fig. 3) contains the basic notifications—latest tasks ("najnovšie úlohy"), received messages ("posledné odkazy"), tasks that are before the deadline for completion ("onedlho vyprší deadline"), graphs (e.g. "Objem platieb za mesiace"—payments per month), reports and a link to the simple communication channel.

A manager can generate a report to evaluate the work of a relevant team member (Fig. 4). A typical report contains the number of assigned tasks ("počet zadaných úloh do súčasnosti"), the number of finished tasks ("počet splnených úloh do

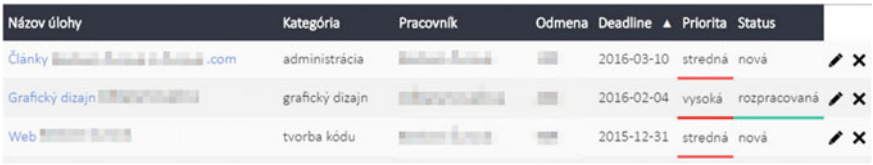

Fig. 2 The tasks list

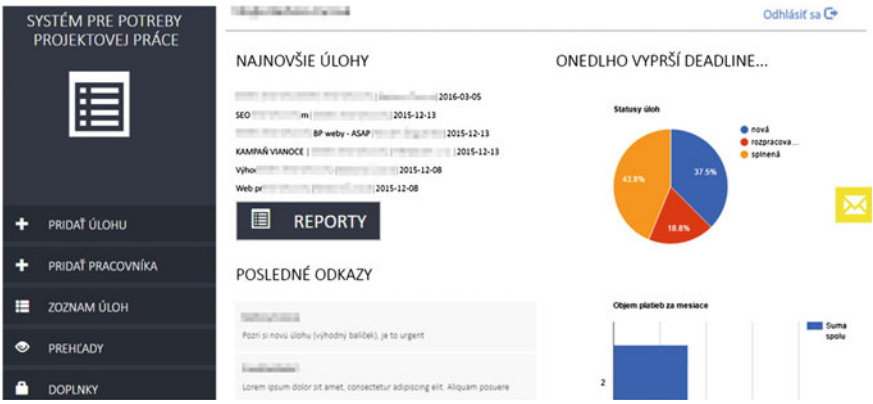

Fig. 3 The dashboard

súčasnosti"), volume of the paid payments ("objem vyplatených platieb do súčasnosti"), the number of currently open tasks ("počet otvorených úloh do súčasnosti").

Finally, we proposed a knowledge base as an advanced feature based on a detailed analysis of the typical processes in such type of companies. We aimed at an IT process representing the internal technical support, i.e. the evidence of the occurred problems and finding the most suitable solution ("hľadať prípad") for these problems from different areas (Fig. 5). The knowledge base should make this process more effective, i.e. all occurred incidents are stored in the database with a simple description, founded solution and a responsible person that solved the problem. In addition, this database can also contain some new ideas and a proposal how to make relevant activities more productive, or some knowledge extracted during related technological or organisation activities.

The effectively managed knowledge base is very important mainly due to its uniqueness for each company. Many technological problems' solutions we can find on the Internet, but such base should contain specific solutions for the specific problems identified and solved by the company staff. In our case, it includes primarily the specific solution for such information systems as e-shops, content management systems (CMS), systems for managing customer relationships (CRM), etc. We illustrated this fact by the following practical examples:

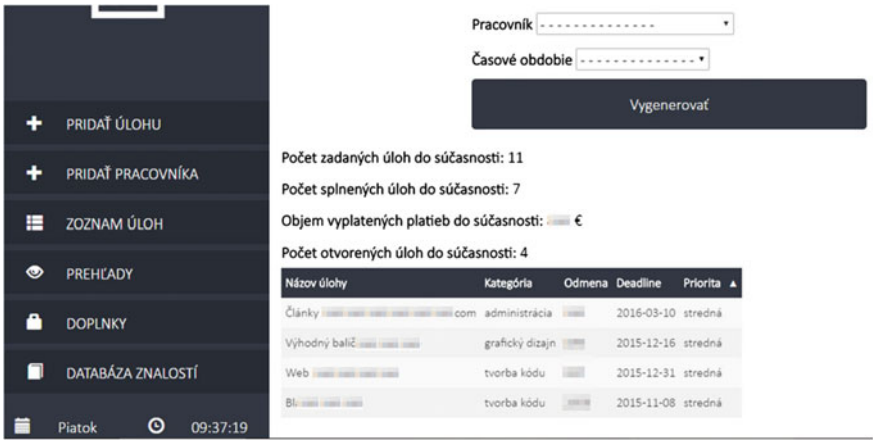

Fig. 4 Simple example of the generated report

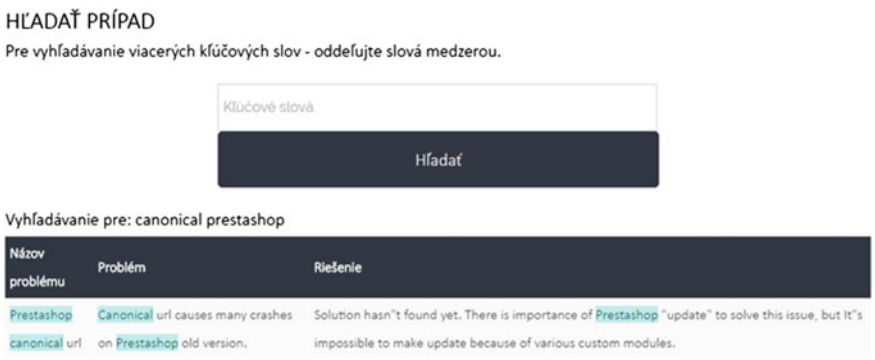

Fig. 5 Example of searching ("hľadať") in the knowledge base

- The company operates for the customer an e-shop based on a very old version of the CMS. The system was not maintained and updated for many years; therefore, it generated various conflicts between the modules. Finding the right solutions to these collisions is very difficult since support for these old versions no longer exists. Therefore, all identified problems with the suitable solutions were described and stored into the implemented knowledge base, e.g. "do not operate the canonical URL, otherwise, the system will collapse".
- The company implemented one website within the selected CMS using HTTP. It was necessary to transfer the website to a subdomain running under HTTPS. Typical tutorials recommend only rewriting the linking of databases and enabling SSL in the related CMS. However, in this case, one of the integrated modules caused a constant redirect within error. The right solution was to

rewrite some parts of the configuration file, but it took a long time to find it since the modules collisions happen in so many different combinations that the developers could not find the information on the Internet.

3 Conclusion

The paper described a case study devoted to the evaluation of the possible successful relationship between project and knowledge management in practice. We designed the proposed system based on the identified requirements from the real customer characterised as a company doing business in web technologies and utilises the services of the external staff. We implemented and deployed the system successfully in practice. We organised the user acceptance testing through a prepared questionnaire containing questions oriented to the user's satisfaction with offered features and the system itself. More than 50 employees participated in this evaluation and the overall satisfaction was more than 93 %. This number refers to the positive response of the participated users and confirms their motivation to use, complement and maintain the available knowledge base.

The system is currently available in the Slovak language but we designed it to meet the needs of a wide range of businesses. In addition to simple project management functions, it provides the knowledge base and some additional features. The module including the knowledge base and related operation services supports all mentioned four knowledge processes. The new knowledge creation represents a possibility to add a new solution or to combine several existing into the new one. In this case, all transformations from the traditional SECI model are applied [9]. The capturing represents organisation of the stored knowledge into predefined categories to ensure simple and fast searching. The transfer provides the system by itself, i.e. all the stored knowledge is available for the members of the project teams through different devices as PC, smartphones or tablets. The reusing of the stored knowledge represents the main motivation behind the proposed system, i.e. the guidance and solutions can be used each time that similar incident/problem occurs and in each iteration, it can be modified and improved.

There are many project management systems on the IT market, but we believe that this new system can be particularly beneficial mainly due to innovative elements, e.g. rewarding of the external staff and the knowledge base. For future work, we plan to design and implement a semi-automatic decision support module using the collected knowledge. This module will replace the currently implemented searching with a smarter alternative. In addition, we consider a possibility to make the implemented system compliant with an ITIL library, mainly with the parts called Service operation and Continual service improvement. The Information Technology Infrastructure Library (ITIL) is a set of practices focused on aligning IT services with the needs of the business.

Acknowledgments This publication is the result of the Project implementation: University Science Park TECHNICOM for Innovation Applications Supported by Knowledge Technology, ITMS: 26220220182, supported by the Research & Development Operational Programme funded by the ERDF (30 %); supported also by the Cultural and Educational Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic under grant No. 025TUKE-4/2015 (50 %); and by the internal faculty research project no. FEI-2015-2 (20 %).

References

1. Ahuja, V., Yang, J., Shankar, R.: Benefits of collaborative ICT adoption for building project management. Constr. Innov. **9**(3), 323–340 (2009)
2. Ajmal, M., Helo, P., Kekäle, T.: Critical factors for knowledge management in project business. J. Knowl. Manag. **14**(1), 156–168 (2010)
3. Duncan, W.R.: A Guide to the Project Management Body of Knowledge. PMI, USA (2009)
4. Kasvi, J., Vartiainen, M., Hailikari, M.: Managing knowledge and knowledge competences in projects and project organisations. Int. J. Project Manag. **21**, 571–582 (2003)
5. Koskinen, K.U.: Tacit knowledge as a promoter of project success. Eur. J. Purchasing Supply Manag. **6**(1), 41–47 (2000)
6. Liebowitz, J.: Knowledge Management Handbook. CRC Press, Boca Raton, USA (1999)
7. Morales-Arroyo, M., Chang, Y. K., De las Nievas Sánchez-Guerro, G.: The Use of KM Tools and Techniques to Reduce Coordination Problems in Project Management. Convergence of Project Management and Knowledge Management, no. 57, pp. 57–70 (2010)
8. Muñoz-Avila, H., Gupta, K., Aha, D. W., Nau, D.: Knowledge-Based Project Planning. Knowledge Management and Organizational Memories, pp. 125–134. Springer (2002)
9. Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. New York Oxford University Press, USA (1995)
10. Owen, J., Burstein, F.: Where knowledge management resides within project management. Case Studies in Knowledge Management, pp. 138–153 (2005)
11. Sokhanvara, S., Matthews, J., Yarlagadda, P.: Importance of knowledge management processes in a project-based organization: a case study of research enterprise. Proc. Eng. **97**, 1825–1830 (2014)

The Process of Verifying the Implementation of Design Patterns—Used Data Models

Rafał Wojszczyk and Włodzimierz Khadzhynov

Abstract Although the design patterns constitute the issue that has been widely discussed in the literature and used by many software developers, there is no formal control over them. The article discussed the problem of verifying the implementation of design patterns applied in object-oriented programming. Two following data models were distinguished in the process of verification: a formal representation that is an equivalent of the analysed software, and a repository of implementation of patterns containing information describing the implementation of design patterns. The proposed solution will make it possible to show implementation errors and potential problems.

Keywords Design patterns · Verifying implementation · Object-oriented programming

1 Introduction

Ward Cunningham, a well-known figure in the community of software developers, when justifying the need of software refactoring, introduced a financial metaphor called the technical debt [1, 2]: unpaid debt leads to accrued interest; the longer it remains unpaid, the more chances there are it will grow to an overwhelming and unpayable amount. When analysing this metaphor, one can conclude that unpaid debt is an unwanted situation in a company. Similarly, the software with too complicated code leads to unwelcome consequences both for the software users and

R. Wojszczyk (✉) · W. Khadzhynov
Departament of Computer Science and Management, Koszalin University of Technology, Śniadeckich 2, 75-453 Koszalin, Poland
e-mail: rafal.wojszczyk@tu.koszalin.pl

W. Khadzhynov
e-mail: hadginov@ie.tu.koszalin.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_9

the company responsible for its production. There are many methods and general recommendations aiming at improvement of every stage of a software production process, and these are: preparing the specification of system requirements identified in the preliminary analysis; subsequently, during designing, modelling of the most important parts or the whole system by means of UML diagrams is extremely popular; the implementation uses appropriate tools and good practices of programming; finally, the procedures of software testing and maintenance are used.

The good practices of programming are, among others, different kinds of patterns, including the following [3]:

- architectural patterns (e.g. MVC) linking different layers of application,
- design patterns [4] covering the class level,
- implementation patterns, also called the idioms, occurring at the level of lines of code.

One of the most popular design patterns were presented in [4], after [5], and these are: frameworks of ready mechanisms which can be used to solve typical designing and programming problems. The catalogue of patterns [4] created in 1995, was based on experiences of programming practitioners [1], and it has been used by many programmers ever since. With regard to their great popularity and a lack of formal control, there is a need of verifying the implementation of design patterns in developed and existing software products.

The article aims at presenting data models used in the process of verifying the implementation of design patterns. The result of verification is essential to attempt to assess the quality of the implementation of design patterns—it is a far-reaching goal of the conducted works. The context of the issue and the problem of verification are presented in Chap. 2. Chapters 3 and 4 include a presentation of the used data models, while Chap. 5 constitutes the conclusion.

2 Verifying the Implementation of Design Patterns

2.1 Context of the Issue

The form of description of design patterns introduced in [4] is, by definition, intended for learning to use them, and it includes: verbal description in a natural language, class diagrams (originally, it is the OMT notation), and an exemplary implementation code based on simple examples. The patterns described in that way, usually present one of the implementation variants which is not an optimal solution [1]. Therefore, a programmer, who bases his work on information presented in the literature, with every implementation of a particular pattern, should enrich the code he is generating with many factors connected, among others, with business logic, to provide better adjustment of the pattern implementation to the software context. These factors increase complexity and diversity of pattern implementation, which

result in the occurrence of different implementation variants and blurring of the boundaries of design patterns. Eventually, this leads to an increase in labour intensity during code inspection.

Design patterns are very often connected only with the stage of software designing and not with a software code [1]. It is a narrow perspective, especially, when the software is implemented by a small team working according to the Scrum methodology [6]. In the agile methodologies from which the Scrum originates, a very little emphasis is placed on system documentation and design. Frequently, decisions concerning solving of a certain problem through the pattern implementation are made swiftly during morning meetings (the so-called morning Scrum). Subsequently, implementation work can be assigned to a team member who has never encountered a particular pattern before. The effect of his work can be the implementation with seeming occurrence of the pattern, that is, software may operate correctly and pass the required tests, still, the implementation of the pattern itself would be inconsistent with requirements and would not fulfil the objective. Problems resulting from this fact would be very well noticeable with the attempt of development or modification of a particular part of software.

There are many methods and tools intended for testing and analysing software source code. These are often provided together with integrated development environment which takes care of fundamentals' conformity with language syntax and assumptions of a particular programming paradigm. The possibilities of development environments can be expanded thanks to numerous automated tools; however, these do not take design patterns into consideration. A programmer, who uses basic components, creates own software artifacts (e.g. dll libraries) including specified business logic which cannot be tested with a universal algorithm. Most of the artifacts require an individual approach and preparation of essential data and testing functions; it is often done with unit tests. A positive result of a unit test for a particular artifact does not mean proper implementation of design patterns in this artifact.

The methods related to design patterns often take up the problem of searching for instances of design patterns in software [7], [8, 9]. The measure of the mentioned studies is presenting a number of instances of design patterns [10], and it is insufficient information to meet the discussed need. Other methods [11] are mainly focused on showing structural correctness; unfortunately, this requires learning a dedicated programming language to describe implementation data. What is more, there are attempts of using well-known software metrics [12] for design patterns implementation; still, in [13] it was shown that the occurrence of patterns in software may have a negative influence on the results of metrics.

In the case of the programming practitioners, as mentioned before the team members of the Scrum groups, the choice of the proper resolution to verification the design patterns is dictated by proper utility features in order to be accepted. Such resolution should be adapted for the skills of the potential customers and should not require learning new conceptions. The formal representation of data, such as the usage of ontology in Information Technology, are relatively new, seem to be not so popular [14], difficult and uneasy to utilize in software project [15]. Well-known

UML (and similar, e.g. OMT) used for the design patterns, is found to be semi-formal representation [3] [16]. It may result from the omission in the UML diagrams some of the aspects of pattern implementation [3, 16], limited possibilities of the verification of implementation [17], also [11]: UML cannot be used to describe an infinite set of pattern instances because the language is not designed for that purpose. In [7, 8] the authors relying on the class diagrams showed that in such approach there is not possible to distinguish, for example, the strategy model from the state [7] and also identification of the Singleton pattern [8].

2.2 Verification of Implementation

Verifying the implementation of a particular design pattern means a number of actions which must be taken to show the analysed software code's conformity with rules and principles of implementation adopted for a particular pattern. The analysed software code and, more specifically, a fragment of the software source code (or intermediate code), included the implemented specific design pattern. The code of the analysed software, within the offered approach, is converted into its formal representation (code equivalent) and only this format is used in verification. The formal representation of software is the first of the discussed data models.

The second offered data model is a design patterns implementation repository which consists information describing the implementation of patterns. Partially on the basis of the solution presented in [16], the repository was proposed which assumes the description of general pattern definitions as a set of features that make up a particular pattern. Subsequently, each of the features is being described with detailed implementation data which take different variants of patterns occurrence into consideration. In the proposed approach, the implementation details were described in the reference model; however, features of patterns may be specified in different ways [18]: through an additional description in a form of rules or proper operations on the code of the analysed software.

Before the first verification process, one should complete the repository with descriptions of patterns once; subsequently, new variants of patterns can be added on the basis of the analysed software. In simplified terms, for one implemented design pattern, the verification process starts with converting the software code into the formal representation; subsequently, the verification is executed according to the occurred features. In case of using the reference model, the verification process is reduced to comparison of a fragment of code equivalent with implementation data, which are identified thanks to the description of every feature occurring in a particular design pattern.

The models described in the further study could be presented in UML notation (the class diagram). However after providing the earlier mentioned information about the usage of the UML with the design patterns, the proposed data models were developed pursuant to the Entity-Relationship diagrams, which additionally allowed to distinguish the predicted entities. ERD are the most popular diagrams

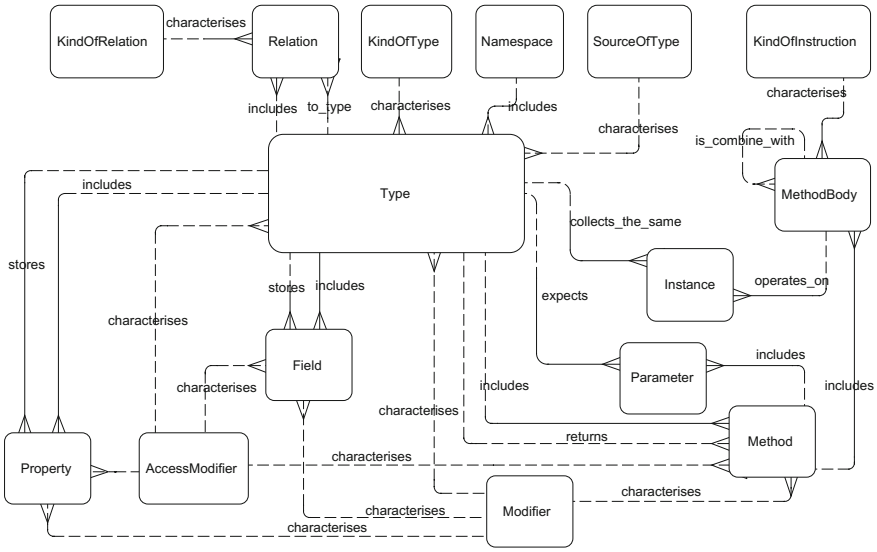

Fig. 1 Data model of formal software representation

intended for data modelling [19, 20]. The diagrams presented in the Figs. 1, 2 and 3 were prepared with the software tool Power Designer 15 using Baker's notation. Used tool allowed for accessible implementation those models as database.

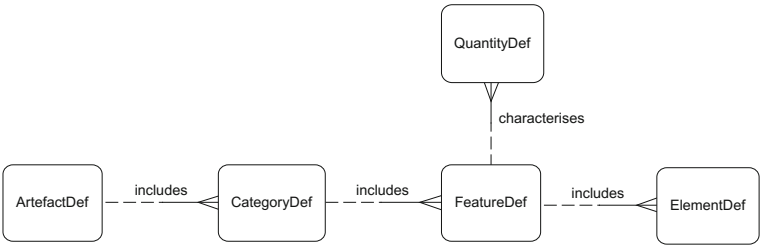

Fig. 2 Definition model

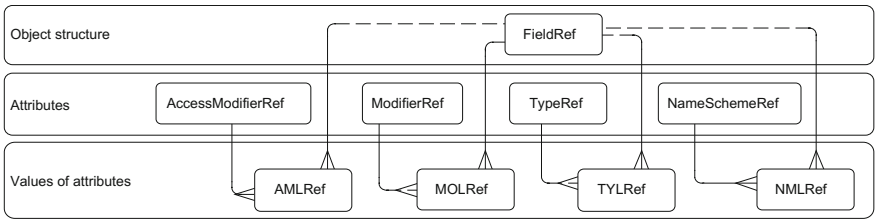

Fig. 3 Selected entities and relations describing the field in the reference model

3 Formal Representation of Software Source Code

3.1 Motivation

The most information on programming is included in the source code, which simultaneous shortcoming is a physical representation—these are properly catalogued text files which are more difficult to analyse than a structured formal representation. In some approaches [21, 22], related to analysing design patterns, a software code transformation into a formal representation was used. A software code equivalent obtained in that way provides the automated processing of data concerning the occurring implementation and, additionally, it enables one to remove useless code and unwanted information (e.g. comments written in a natural language, unit tests code). Unfortunately the existing ways of representation of the software, or the design patterns, does not fulfil the demanded requirements, what led to formulate new solutions.

3.2 Requirements

The development of the data model of the formal representation of the analysed software was accompanied with the following requirements:

- representation of software object structure, including separation of class components (separate entities for fields, properties, methods, etc.),
- representation of instances and object state changes,
- data input on the basis of managed code [23] and a possibility of enhancement with other data sources,
- execution through a relational database,
- intuitively of the most popular object-oriented languages for programmers.

Simultaneously, there were restrictions imposed reducing data specificity in relation to the source code. The offered data model did not reflect values of fields and variables (e.g. content of character strings, binary data). Data specificity was reduced also in case of specific elements of modern programming languages, e.g. anonymous types are represented as new classes; in this article it is consistent with ECMA-355 standard [23]. In consequence of the requirements and the imposed limitations, the prepared model of the formal software representation is not a direct equivalent of any programing language and could not be created automatically.

Realisation of the final from the listed requirements was combined with the model based on the OOP assumption, since every programmer using this paradigm should know it obligatory. The advantage of this method is programing language independence, which was provided for 4 different languages officially supported by. NET platform. Additionally it allows to power the model from different data sources, including using the methods of reverse engineering.

3.3 Data Model

The offered data model and software transformation algorithm were initially presented in [18]. Figure 1 shows the expanded version of the model. An element with the highest responsibility is Type entity; it is responsible for type representation (e.g. specific class) in the object-oriented programming paradigm. This entity includes:

- *Field* collection—a single element of the collection represents a field (the so-called global variable) in a particular type; it is characterised with a name, type (determines a type of data stored by the field, executed by a separate relation to *Type* entity) and dictionary entities of access modifier and modifier,
- *Property* collection—a single element of the collection represents a property (the so-called setter or getter); it is characterised by analogous fields and entities as *Filed*, and, additionally, determination of Set and Get function, that is, data record and read,
- *Method* collection—a single element of the collection represents a method (function) occurring in a certain type; it is characterised by a name, definition whether the method is a constructor, type (determined type of data returned by the method) as well as dictionary entities of access modifier and modifier; additional entities will be described later on in the article,
- *Relation* collection—a single element of the collection represents relation between a certain type and other type together with further definition of a kind of relation (e.g. inheritance, execution) through *KindOfRelation* dictionary entity,
- *KindOfType*—determines the kind of a certain type (e.g. class, interface),
- *SourceOfType*—source of origin (from the analysed software, system type or dummy statement type,
- *Namespace*—definition of name space in which a particular type is included.

Additionally, each type is characterised with a modifier and an access modifier.

Method entity includes *Parameter* collection defining parameters adopted by a certain method; each element of the collection is characterised by a name and an expected data type (separate relation to *Type* entity). *MethodBody* collection defines instructions occurring in a certain method. Composed instructions are decomposed to single instructions and characterised with a proper kind of instruction (according to OpCode [23] of the managed code, e.g. object creation) and occurrence sequence. A reflexive relation means combining particular instructions in a composed instruction. Additionally, *Instance* collection group those instructions together which refer to the same instance of a certain type. The instance of a certain type (e.g. the occurrence of a specific copy of a particular class) is characterised by instance name and unique Guid (required if instance name cannot be obtained).

Modifier dictionary entity (*Modifier*) describes modifiers (e.g. abstract, sealed, static, virtual, etc.) occurring with types and selected components of types. Access modifier entity (*AccessModifier*) determines the scope of visibility (e.g. public,

private, internal, protected) of the selected components. The content of *Modifier*, *AccessModifier*, *KindOfRelation*, *SourceOfType*, *KindOfInstruction* entities is predefined depending on a programming language represented by a data model (in this article it is C#). In the special case existence of the "empty" value in the mentioned entities is acceptable, if the programming language allows that.

4 Repository of Implementation of Design Patterns

Repository of implementation of design patterns is a treasury of information on possible ways of implementation of the analysed design patterns. Due to great implementation diversity, the decision on separation of the general definition of patterns was made. The definition of patterns is a kind of a list of contents which includes properly selected features of patterns in isolation from detailed implementation. An feature of a design pattern is one of the necessary element which characterises a certain pattern. On the basis of the example of Singleton pattern, it can be said that one of the features is instance accessing which can be implemented in a number of ways (e.g. field, property, method). The fact of the occurrence of instance accessing is an essential feature of this pattern, while the way of implementation is dependent on different factors (e.g. software code context, programming language). Similar separation appears in ISO 9126 norm defining the quality of software product. The norm is described by a tree structure which includes characteristics and in those, recursively, there are subcharacteristics. Characteristics-leaves have metrics—functions assigned to them, which define certain values on the basis of measurable software attributes. The definition of characteristics is informal—it expresses certain intent, while the metrics are formalised. In the offered approach, the definition model corresponds with the characteristics from ISO 9126, and a detailed implementation of design patterns corresponds with the metrics.

Separating the description of the definition of features from the implementation of design patterns enables one to expand the scope of verification. Depending on an analysed aspect, one can use a proper solution: structural aspects of the implementation of design patterns can be verified through comparison of the analysed software with the reference model, while aspects connected with software dynamics and behaviour may require the verification based on tracking object state changes (e.g. through execution of proper queries on software formal representation).

4.1 Definition Model

In the offered solution, the description of features was executed as a hierarchical data structure. The metamodel of this structure is presented in Fig. 2. The

hierarchy's root and, simultaneously, the most general one, is *ArtefactDef* entity which represents specific design patterns. As it has been already mentioned, the definition of each design pattern is composed of proper features, which, for convenience sake, were grouped in *CategoryDef* entity included in *ArtefactDef*. Features (collection of *FeatureDef* entity) are grouped in categories according to the criterion of assignment to the aspects of: structure composing the pattern, use, behaviour, etc. Each feature may be dependent on other feature; the type of relation is determined by *RelatedType* field (not showed in the picture) which is an enumerated type. Possible kinds of feature relations are: conjunction, disjunction and including subordinate features. A feature is characterised by a name and defined quantity *QuantityDef* entity) of a particular feature occurrence (In special case, quantity of occurrence of minimum 1 means the necessity of feature occurrence). Each feature includes detailed elements *ElementDef* entity), which enables one to decompose a feature additionally, for instance, a feature of structure creating a pattern is a specific class which is additionally decomposed to: dependence on other class, access modifier determining the range of visibility. Each element is characterised by a name and a possibility of negation (everything other than a particular element is acceptable then). In special case, there can be only one element of a particular feature and that reduces the hierarchy to the level of features. *ElementDef* entity is simultaneously a link between the definition model and the description of the detailed implementation of design patterns.

The offered data model allows for great freedom in the description of the definition of design patterns. Depending on the pattern, it is possible to describe a general feature, which implementation can be done in various ways or to specify elements of a particular feature when there is a constant implementation.

4.2 Reference Model

The reference model is one of the solutions used to describe the detailed implementation of design patterns, in other words, implementation data. The detailed implementation means supplementing the definition model of a set of fine-grained information (in other words: attributes) [24] directly connected with the software code. Fine grained information enables one to describe pattern execution in a specific programming language. Simultaneously, the description of generic implementations (in other words: template implementations) is possible. This means that the occurred data only define a general character of the described attributes without imposing any specific values, e.g. defining field's data type as any numerical type.

The offered reference model was designed for describing the implementation of a structure forming design patterns. A structure forming a pattern is usually a set of joined classes and interfaces, e.g. in the Strategy pattern, it is the interface declaring

a strategy and classes which implement this interface. According to the assumptions, the analysed software and, more specifically, data included in the formal representation, is compared to the content of the reference model. The reference model also meet requirements that were imposed on the formal representation of the software. To automate the comparison process, the reference model includes decomposition of the object structure into entities analogous to the one of the software formal representation model, that is, *TypeRef* includes *FieldRef* collection, etc. (in accordance with PascalCase, the "Ref" suffix was added to the entity of the reference model in order to distinguish entities between models and make them easier to recognize). However, data occurring in these entities were additionally decomposed to separate entities which enable one to describe different variants of fine-grained attributes. Each of the variants is marked with a proper level of matching. The level of matching means defining the extent of matching a particular attribute to a specific implementation variant. In the assumed range from "0" to "2", "2" value means the highest level of matching. The occurrence of many attributes of the same level of matching is possible.

The data, described by the reference model, can be divided into three groups:

- data forming the object structure,
- data defining attributes occurring with a particular element of the object structure,
- data representing specific values of attributes; this group also contains the definition of a level of matching.

The division for the mentioned three groups of data has been shown in the Fig. 3, which shows the extract of the referential model. In the picture only entities describing the field are visible, and some relations are not visible, that is each field can be included in a specific type (e.g. as the class component), while each type can be characterised with different attributes. The total reference model is presented in Fig. 4. Table 1 presents entities forming the object structure group in reference to attribute entities characterising them, which is true to paradigm of the object-oriented programming. Decomposition of other entities falling within the object structure group was executed similarly to the presented example of *Field* entity. *KindOfTypeRef*, *KindOfRelRef*, *ModifierRef*, *AccesModifierRef* entities are dictionaries of constant attributes, which value occurrence (for a particular element of the object structure) is stored respectively in the following entities: *KTLRef*, *KRLRef*, *MOLRef*, *AMLRef*. The meaning of these entities is similar to the one in the model of software formal representation. *NameSchemeRef*} entity is used slightly different; its values (*NMLRef* entity) enable one to describe specific names according to the schemes: starts with, ends with, includes. *TypeRef* entity plays a double role; it includes elements of object structure (type components) and occurs as a dictionary of type attribute. Eventually, the layer of entities representing specific attribute values is linked by means of association entity with *ElementDef* of the definition model.

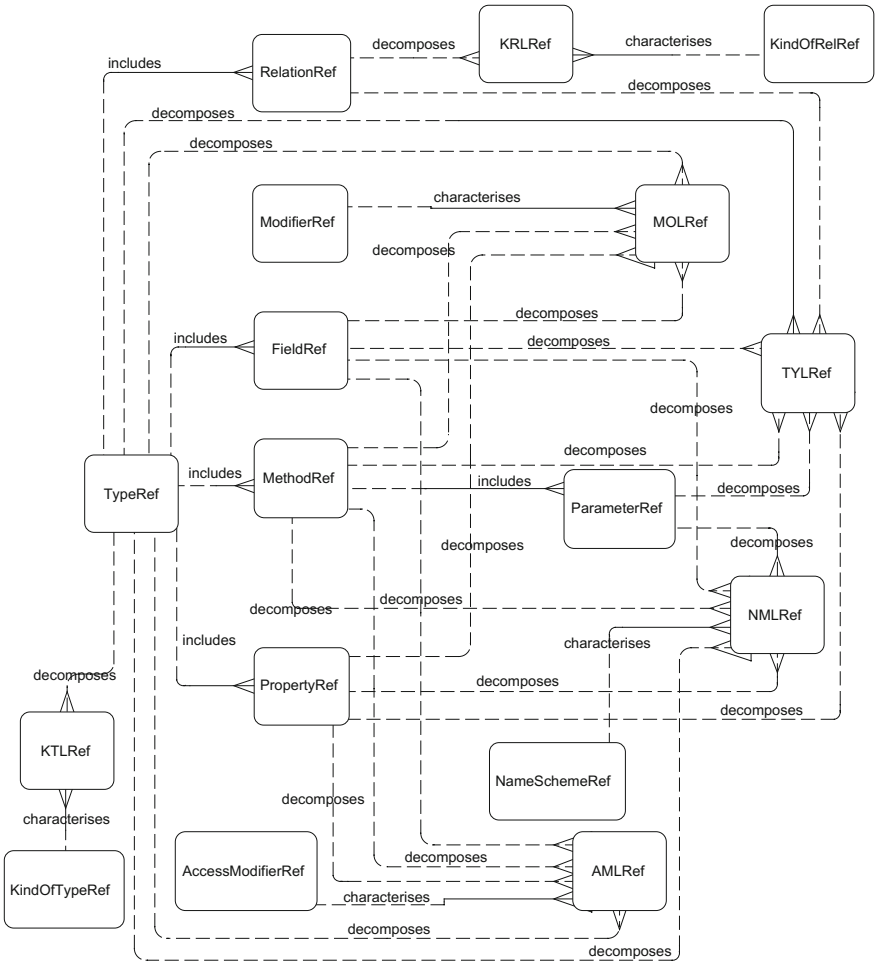

Fig. 4 Reference model

Table 1 The list of attributes occurring with entities ("+" attribute occurs in a particular entity)

	Type-Ref	*Field-Ref*	*Property-Ref*	*Method-Ref*	*Parameter-Ref*	*Relation-Ref*
KindOfTypeRef	+	−	−	−	−	−
KindOfRelationRef	−	−	−	−	−	+
ModifierRef	+	+	+	+	−	−
AccessModifierRef	+	+	+	+	−	−
NameSchemeRef	+	+	+	+	+	−
TypeRef	−	+	+	+	+	+

5 Verification of Implementation Based on the Model

The verification of the design patterns is a complex process, hence striving for automation. Even for a simple pattern, it is Singleton, the process of verification is composed of many steps which cannot be described in a concise way. The research so far, that is the examples of the verification and the implementation of design patterns with the interpretation of the results has been described in [18].

Figure 5 shows the diagram of the idea of verification based on the reference model. Figure 5a shows the example of implementation of the source code of the Singleton design pattern and the visualisation (shapes encircled with a dotted line) to make further comparison more visible. Figure 5b shows the simplified definition of the design pattern. Additionally the hierarchy division of the model definition and the groups of the data of the referential model has been marked up. In order to simplify the Fig. 5, less meaningful values of the attributes (e.g. lack of modifiers) and no matching levels (they are not allowed, e.g. access modifier of the constructor different then private) were omitted. The diagram presents only the chosen elements of the structure creating the pattern, and in the described example it has been reduced to the field affording to share the Singleton instance. Shapes with dots signify further data development [18]. Figure 5c shows the result of the comparison. Shapes from the visualisation of the formal representation of the software were

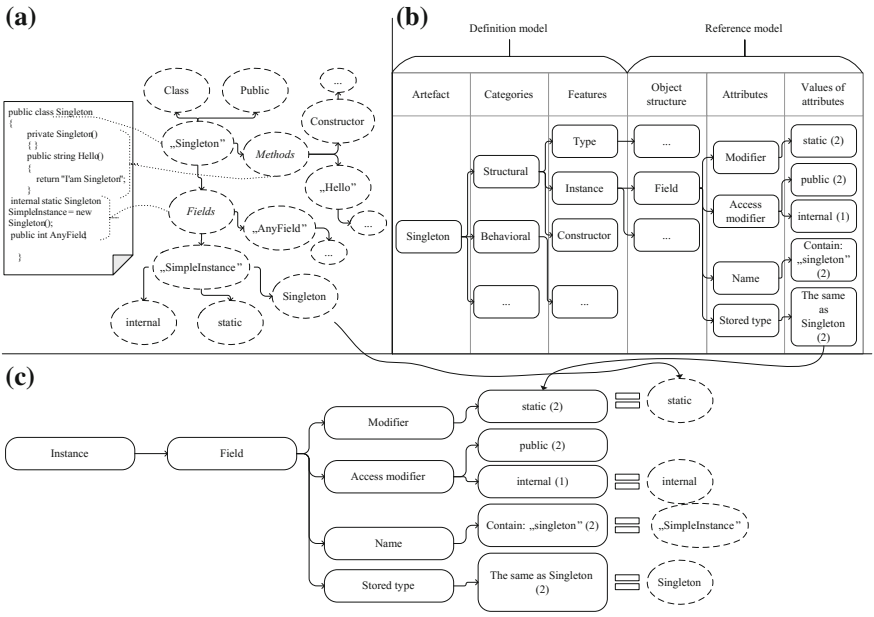

Fig. 5 The diagram describing the idea of verification based on the reference model. **a** shows the visualisation of program code, **b** shows the visualisation of definition and reference model (in brackets—the level of matching is shown), **c** shows verification result

Table 2 Consequences of the occurrence of the selected access modifiers

Access modifier	Level of matching	Consequences of occurrence
Public	2	Recommended, it makes Singleton instance accessible to every class, does not limit the use
Internal	1	Acceptable, still it limits the scope of visibility only to the library. It may cause problems when attempting integration with the analysed software
Private	0	Unacceptable, it will make an access to instance impossible, a pattern will not serve its purpose

applied on the shapes of the reference model. In the presented example the internal access modifier appeared (lover level of matching), which results are presented in the Table 2.

6 Summary

The article briefly discussed the need of verifying the implementation of design patterns and mentioned the related issues. Subsequently, it described the general assumptions of the author's own verification process and data models occurred in it. The result of verification is essential in relation to further works concerning the assessment of quality of the implementation of design patterns.

The first of the presented data models is designed for storing the equivalent of the source code of the analysed software; the second one stores definitions of design patterns. Both models and the described process were executed as a prototype tool. However, the obtained results confirmed appropriateness of the assumed solutions and these are the basis for further analyses concerning expanding the repository of implementation of patterns of a possibility of verifying behavioural aspects, e.g. based on UML diagrams, and, subsequently, offering the criteria of quality of implementation of design patterns.

References

1. Kerievsky, J.: Refactoring to Patterns. Addison-Wesley Professional, Boston (2004)
2. Fowler, M., et al.: Refactoring: Improving the Design of Existing Code. Addison-Wesley Professional, Boston (1999)
3. Pavlic, L. et al.: Improving design pattern adoption with ontology-based design pattern repository. Informatica Int. J. Comput. Inform. **33**. Ljubljana (2009)
4. Gamma, E., et al.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Professional, Boston (1994)
5. McConnell, S.: Code Complete: A Practical Handbook of Software Construction, 2nd edn. Microsoft Press (2004)

6. Martin, R.C., Martin, M.: Agile Principles, Patterns, and Practices in C#. Prentice Hall, New Jersey (2006)
7. Tsantalis, N. et al.: Design pattern detection using similarity scoring. IEEE Trans. Softw. Eng. **32**(11) 896–908 (2006)
8. Singh Rao, R., Gupta, M.: Design pattern detection by greedy algorithm using inexact graph matching. Int. J. Eng. Comput. Sci. **2**(10), 3658–3664 (2013)
9. Binun, A.: High accuracy design pattern detection. Ph.D. thesis, Rheinischen Friedrich-Wilhelms-Universitat Bonn (2012)
10. Ackerman, L., Gonzalez, C.: Patterns-Based Engineering: Successfully Delivering Solutions via Patterns. Addison-Wesley Professional, Boston (2010)
11. Blewitt, A.: Spine: language for pattern verification. Design Patterns Formalization Techniques. IGI Global, Hershey (2007)
12. Khaer, MdA, et al.: An Empirical Analysis of Software Systems for Measurement of Design Quality Level Based on Design Patterns. Computer and Information Technology, IEEE (2007)
13. Hernandez, J., et al.: Selection of metrics for predicting the appropriate application of design patterns. In: 2nd Asian Conference on Pattern Languages of Programs (2011)
14. Welty, C.: Ontology research. AI Mag. **24**(3) (2003)
15. Podsiadły-Marczykowska, T., et al.: Rule-Based Algorithm Transforming OWL Ontology Into Relational Database. Beyond Databases, Architectures, and Structures, vol. 424. Communications in Computer and Information Science, Springer, Berlin (2014)
16. Rasool, G.: Customizable feature based design pattern recognition integrating multiple techniques. Ph.D. thesis, Technische Universitat Ilmenau, Ilmenau (2010)
17. Walter, S.: Code Structure Definition and Verification. CSCUBS Proceedings, Bonn (2014)
18. Wojszczyk, R.: The model and function of quality assessment of implementation of design patterns. Appl. Comput. Sci. **11**(3). Institute of Technological Systems of Information, Lublin University of Technology, Lublin (2015)
19. Suslow, W.: Analiza i modelowanie konceptualne w inzynierii systemów oprogramowania—ujęcie humanistyczne. Wydawnictwo Uczelniane Politechniki Koszalińskiej, Koszalin, (2013)
20. Shamailh, A.A.: An experimental comparison of ER and UML class diagrams. Int. J. Hybrid Inf. Technol. **8**(2), 279–288 (2015)
21. Grzanek, K.: Realizacja systemu wyszukiwania wystąpień wzorców projektowych w oprogramowaniu przy zastosowaniu metod analizy statycznej kodu źródłowego. Ph.D. thesis, Politechnika Częstochowska, Łódz (2008)
22. Kirasić, D., Basch, D.: Ontology-Based Design Pattern Recognition. Knowledge-Based Intelligent Information and Engineering Systems, Zagreb (2008)
23. Troelsen, A.: Pro C# 2008 and the .NET 3.5 Platform. Apress, New York (2007)
24. De Lucia, A., et al.: Design pattern recovery through visual language parsing and source code analysis. J. Syst. Softw. Arch. **82**(7). Elsevier Science Inc, New York (2009)

The Scrum Pricing Model for Developing a Common Software Framework in a Multi-project Environment

Michał Turek, Jan Werewka and Dariusz Pałka

Abstract The article presents a new pricing model applied in Scrum contracting. The model will introduce an innovative approach to the promotion of software products with reference to their pricing—bounding non-profit features with the classical cost estimation. It will assume a principle of single charge for all components reused in products created for different customers. The model will be applicable mainly in multi-project Scrum environments, but will also support scaled multi-team Scrum development environments (where software development work is allocated to different Scrum teams). It will introduce a complete mechanism for cost estimation, with component pricing procedures (compatible with Scrum estimation methodologies), component reuse tracking and so on. The price offered for software to be developed will depend on its popularity among customers. So, the introduction of a new feature triggered by one customer can be beneficial for all parties taking part in the framework development, with justified share of the price. In such situations a common base for software development pricing techniques must be established. It will depend on different factors, such as: team effort, team velocities, resources and technologies used. To express these values a reference currency will be defined. The model, along with pricing mechanisms presented in the article, can be effectively used in many software products in which software evolution is a core. The pricing efficiency, commonly recognized as a feature of the model, will encourage potential customers to choose a software company that uses it.

Keywords Scrum contracting · Software development · Agile software development · Scrum · Outsourcing

M. Turek (✉) · J. Werewka · D. Pałka
AGH University of Science and Technology, Krakow, Poland
e-mail: mitu@agh.edu.pl

J. Werewka
e-mail: werewka@agh.edu.pl

D. Pałka
e-mail: dpalka@agh.edu.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_10

1 Introduction

Scrum contracting poses many challenges, especially when Scrum methodology is scaled up. In such cases many different teams work on different parts of the product, which must be consolidated—following various rules—to different software systems, i.e. software products needed by various customers. A multi-project environment [21] assumes the existence of many concurrent project development processes with a possible use of common resources. Scrum scaling considered here is based on the use of multiple Scrum teams to support one software line. Each time Scrum scaling or multi-project cooperation is considered, certain special practices must be established to share tasks or resources. Product owners assigned to particular teams must define their stories (which results in the allocation of stories in a product backlog). This must be done with care, because any kind of effort duplication or blocking caused by misallocation of user's stories ends in wasting resources. If there are many different parties interested in software development, the situation becomes even more complicated. Some functionality can be reused in many projects. But the question is who should develop it and pay for it, and, of course, who is authorized to use it. Reusable functionality can be considered common and placed in something the authors of the article call a common framework.

The method presented in the article deals with such a special case, namely, a situation in which a common software container is established to host reusable functionality. Many Scrum teams can obtain functionality from the container, if needed. They can also share newly developed functionality when a common backlog management party identifies it as potentially useful. It reflects a common commercial market situation, in which a software company owns some kind of a software framework and develops IT systems for succeeding customers. If some valuable functionality is developed for a particular customer (as a dedicated addition to his system), a software company considers its permanent addition to its reusable framework. The conditions of such reuse are individually negotiated with a customer. There could even be a special team involved in the whole process, dedicated only to work on the development of a common framework. The problem with such approach is connected with the fact that every-time negotiations are usually tedious and expensive, so an automation process for doing it should be introduced. The main aim of the study presented in the article is to define an efficient and transparent pricing methodology applied to Scrum teams working on a common software framework.

The pricing model created by the authors can be considered as a hybrid of a classical cost-based pricing model and a non-profit one. The main principle is not to charge the customer full price for something that has already been paid for by someone else. This assumption (officially known by the customer as a feature of the model) could encourage the customer to choose a particular software development company. The customer will then know that if certain components will be reused during the development of his system, he will be charged only as another

component benefiter, not the owner of the component. The price he will finally pay will thus be reduced. In other words, the contracted price will cover the amount of money spent on the actual development of the system (instead of the price the customer is willing to pay). Such approach will show a new, customer-friendly attitude to pricing procedures and (if well-known) will open a new chapter in IT service promotion in the software development market.

2 Common Framework Development by Scrum Teams

The definition of a software framework adopted in the paper states that it is a universal, reusable software environment that facilitates development of software applications, products and solutions. Numerous definitions of a software framework can be found in the literature, e.g. [11]. The framework application example [10] presents a survey of commercial frameworks for the Internet of Things. For the sake of clarity, we assume that a software platform provides the base for deploying and running applications and consists mainly of a software framework, SDK (Software Development Kit) and system software. [13] gives an example of characteristics and advantages of software platforms for unmanned systems.

In the article it is assumed that a framework consists of components which are building blocks of a system. At the source code level components may be represented by modules and at run time as units performing defined tasks and interacting with other components by using interfaces. A component-based framework is decoupled into different sets of components, each having individual features. Table 1 presents the relations described above.

The components of a CSF (Common Software Framework) at a given time can be described as a set of n components: CSF = c_1, c_2, … , c_i, … , c_n.

Development of a common software framework can be performed in many different ways, and two examples will be given below. In the first example, described in Fig. 1 in ArchiMate notation [4], there is an organization or a company which owns a CSF and has software development teams used for CSF development. The same organization can have other teams developing software products for customers and using components from a CSF in their development. In such case teams managed by a CSF PO (Product Owner) develop and maintain a CSF.

Table 1 Relations between products, features and components

		Products			Components								
		P1	P2	P3	CI	C2	C3	C4	C5	C6	C7	C8	C9
Features	F1		X		X		X		X	X	X		
	F2	X	X	X		X		X			X		X
	F3		X	X		X			X			X	X
	F4	X			X	X	X						
	F5		X	X				X					

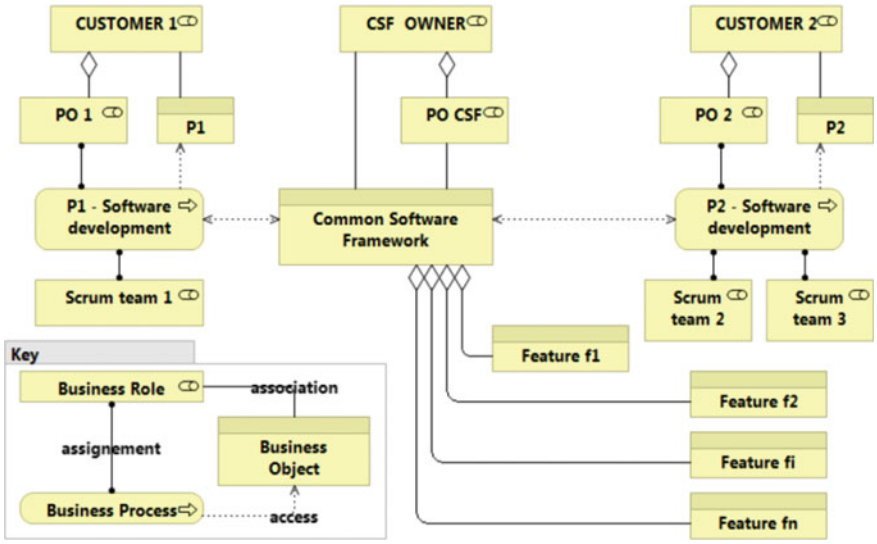

Fig. 1 CSF development performed by a dedicated Scrum team

In the second example product owners delegated by a software company will be responsible for a particular project and will provide stories for common framework backlog. The role of a CSF PO will be to analyze the ideas (stories) of other product owners and classify them as potentially useful in the future. In such cases the story can be moved from a customer Scrum team product backlog to a common framework backlog. This process was already described in details in the previous article by [21]. As a consequence, changes in the framework functionality may be implemented under control of a dedicated CSF PO, who is responsible for the system structure that supports proper reusability. In this case a customer-dedicated Scrum team will develop new features for a common framework.

Both system development techniques should lead to considerations about pricing models. Since common framework components may be built either by dedicated teams or common framework teams, special pricing techniques based on reusable component price discounting should be applied. If a common framework component is indeed reusable, a company owning it should lower its price in exchange for the reuse rights they obtain. On the other hand, if a component is finally reused, its price for another customer should be significantly lowered (in comparison to a specially designed component).

A commercial use of such an approach (where development of a CSF is price efficient and transparent) would probably require the use of a specially established monetary unit. This unit will allow for unifying cost calculations, even if the costs originate in materials used, licenses bought from a third party, etc. (Software Revenue Recognition, [18]). A highly customized pricing model should, in the first

place, be applied to projects with many similar deployments. In the next sections the authors propose a stable pricing model, which makes the automation of component pricing possible.

3 Problem Statement

The importance of developing a CSF is growing. IT companies try to develop extensive software in cooperation with other companies. There are different reasons for such collaboration, e.g. reducing the price or the lack of necessary resources. The issue considered here addresses the following areas:

- Development of a common software framework for use by different customers Software is designed for a limited set of customers (e.g. niche product)
- Cooperating companies try to reduce development costs (reusable common features)
- The CSF development should be transparent to all cooperating companies
- There is necessity for an agreed pricing model
- The pricing model should be based on stable metrics. The measure should not be dependent on time passing and geographical zones
- Software may consists of thousands of components, so automatic price calculation is important
- The price of a component should decrease if it is sold many times.

The issue described in the article has not been thoroughly investigated in the literature so far. Other articles from this area focus mainly on factors influencing costs, metrics for software costs and cost normalization.

In paper [19] a new learning-oriented and semi-automated early-stage cost estimation solution is proposed specifically designed for globally distributed software projects. 16 metrics are used in it: (1) work dispersion, (2) time zone overlap, (3) range of parallel-sequential work handover, (4) team culture, (5) client specific knowledge, (6) extent of client involvement, (7) design and technology newness, (8) allocated team size, (9) process model, (10) project effort, (11) code size, (12) development productivity, (13) defect density, (14) reuse, (15) rework, (16) project management effort. The paper proposes, in the context of globally distributed software projects, a semi-automated solution for obtaining better estimates, especially by tapping organizational memory and learning from past projects. In paper [17], the authors describe several elements which influence the estimation of the project, such as sprint points or project- and people-related factors. On the basis of an algorithmic estimation method, various factors are proposed to determine more accurate release date, costs, effort and duration of the project, especially for Scrum. Paper [22] presents the analysis of the ‘value for money’ related to global operating companies and globally distributed software engineering teams. The authors recommend the extension of the model using parameters which

directly measure customer satisfaction, such as customer feedback, the number of complaints filed, changes in the number of licenses, etc. Building metrics in software engineering is not an easy task. The literature provides some rules regarding how to define and evaluate metrics in software engineering, e.g. [14]. Effort estimation in agile software development can be done in different ways [16]. The most popular 5 size metrics are: story points, line of code (LOC), Person/Ideal hours or days, function points (FP), and use case points (UCP).

If multiple metrics are used, a common cost normalization and calibration method will be needed. In case of multiple Scrum teams involved in drawing work from a common Product Backlog, we have to make sure that normalized points will stay aligned [20]. To do so, we can choose certain known agile estimation methods, for instance:

- A centralized method, proposed by Mike Cohn [9]. To achieve normalization, all teams should start together in a joint planning poker session for an hour or so and estimate ten to twenty stories and then use them as baselines for estimating other stories.
- A semi-distributed normalization method NM, proposed by SAFe [7] using bottom-up and top-down estimations, Fix Time/Flex Scope and FixScope/Flex Time agile planning. The method is based on a proposal that each team finds a story taking 1 ideal day (1DD) Approx. 1/2 day of development and 1/2 day of testing.
- A fully distributed method [8] based on CNM (Calibrated Normalization Method). An enterprise decides on the size of one Normalized Story Point (called Normalization Basis) used by all portfolios, programs, projects, project teams in all sprints throughout the enterprise. Normalization Basis is analogous to Headquarter Currency. Normalization Basis = 40 h of effort can be used as an example. Each team estimates relative sizes of stories (Team Story Points, TSP) using any relative sizing technique, such as Planning Poker. TSPs are analogous to the local currency of each team. Each team calibrates the size of its team story point and converts the Team Story Point (TSP) numbers into normalized story point (NSP) numbers.

A fair and transparent pricing model is essential for companies collaborating on developing common software which can be used by many clients.

4 Common Feature Pricing Formulas

The idea proposed in the paper is to establish a time-lined effort pricing mechanism that will help to express real aggregated effort invested to obtain certain functionality in previous moments of time. First of all, an effort-expressing unit must be introduced. In a classical Scrum case and when we assume only men working hours as a cost generating factor, a normalized effort value (let's call it NEV) can be

established as a baseline for effort and, subsequently, a component of cost estimation. It can be based on the Scrum sprint total working hours calculation (based on ideal days and team size). For instance, it will multiply 1 week ∗ 5 days ∗ 8 h ∗ 5 team members, equaling 200 h. In this case, the total scrum sprint cost will equal NEV ∗ cost per working hour. This total cost must be used as a reference point for other possible costs (originating from materials used, licenses bought etc.). Let's call it BESCoin—a Baseline effort Scrum Coin.

$$\text{BESCoin} = \text{NEV} * \text{cost per working hour} \tag{1}$$

Effort must be expressed as a commonly identified value that could be related to currency. For better granularity and quicker Scrum task estimation, it should refer to scrum stories estimation, not only to Scrum sprint as a whole unit (considering, of course, changing Scrum team velocity[1] in Scrum sprint). We will call it a Scrum coin (SCoin). Basically, a Scrum Coin will help to evaluate the cost of men-hours used to develop a particular functionality in a product. But effort expressed in these Scrum coins could also be evaluated in other ways, depending on a scrum contracting method used during development and even non-scrum. A Scrum coin can be mapped to actual work costs, expressed in a currency and priced in a particular moment of time. An initial Scrum coin value will have to be established as a fixed value when the project starts with the use of reference team work measurement, as shown above. During the Scrum process, cost estimation will require a formula bounding SCoin with sprint team Velocity. Because Scrum team velocity grows in each Scrum sprint and a relation between BESCoin and SCoin should also be fixed, we have to define the Normalization Factor (NF) for referring the cost to Baseline effort Scrum Coin. And then:

$$\text{SCoin} = \text{BESCoin} * \text{NF}/\text{V} \tag{2}$$

where V is a team velocity expressed in story points. Having NF, we can express a number of story points assigned to one SCoin in a particular Scrum sprint. So, in fact SCoin will express real costs of a portion of effort expressed in a currency and established regardless of scrum team characteristics (the cost of Sprint, team velocity, etc.). These values will reduce each other in formula (1). We can go through an example here: Let's assume a team of 8 members having done a piece of functionality during a Scrum Sprint. A Scrum story describing this functionality was estimated for 20 Story points. Additionally, team velocity during the Sprint was 80. In this case only a portion of team effort was dedicated to develop this story. A negotiated sprint cost (meaning Baseline effort Scrum Coin) was 10,000 Euro at

[1]The term 'team velocity' used in Scrum methodology means how much functionality will be delivered by a team in a sprint. The velocity depends, among others, on the team efficiency and team members availability.

that time. Thanks to NF, in each Sprint we can calibrate the SCoin value referring to 10 amount of Story points, let's say, as an equivalent of 10 Story points. Preparing calculations, we have: V = 80, BESCoin = 10,000€, NF = 10. Subsequently, we can calculate SCoin value as SCoin = BESCoin ∗ NF/V = 10,000 ∗ 10/80 = 1250€. That is the SCoin price. All effort costs will be stored in a functionality register (expressed in SCoins paid for development in the past, that is Sprint by Sprint). For the current example it will be adding just 2 SCoins (20 story point/SCoin price). What is important, SCoin can also express costs of any other nature, if needed. Having those two values (SCoin price and the amount of SCoins paid), we can calculate the exact price paid for feature development or modification any time. If we assume that negotiated Sprint cost consists only of men-hour costs in a Sprint, the SCoin formula (1) can be extended to the following form:

$$\text{SCoin} = \text{cost per working hour} * T_{\text{number}} * h_{\text{day}} * S_{\text{days}} * \text{NF/V} \tag{3}$$

where T_{number} is a number of persons in a Scrum team, h_{day} is a number of contracted working hours a day, and S_{days} defines the Sprint length (a number of working days in a Sprint). Since the common framework functionality now marked with SCoin amounts will usually be accessed by API interfaces implemented between particular system components, SCoin amounts can now be assigned to these interfaces. Subsequently, each interface produced will now have a list of time-stamped SCoin amount values attached. If somebody uses some API interface to support a new project, a suitable license cost for its usage can easily be calculated. A company can also freely determine historical cost values on a timeline (which, as we know, depends on such factors as material pricing or working team salary changes). To calculate the effort spent to obtain existing functionality, a full product change history can now be tracked. So, the company can charge the customer for the total effort spent (for instance, also including rollbacked functionality, etc.) or not, depending on its policy. All necessary factors, such as historical team velocity changes or API methods used, are already taken into consideration. The final price for a given functionality will be a sum of historically evaluated costs of its ingredients.

5 An Automated Pricing Model Based on Component Relations in the Framework

If a customer orders additional functionality that enhances a common framework, the price calculated for him will naturally depend on both common and additionally created functionality scope. Due to large numbers of components, it is recommended to determine the costs in an automatic way. This section will introduce a solution developed for automated cost calculation under such circumstances.

In a typical scenario, a product (Product A) developed for a given customer (Customer A) contains the following elements:

- The code created especially for Product A (with the use of Scrum software development methodology)
- Elements (such as libraries, modules, components, etc.) which constitute a common framework developed by a company for many customers and/or products
- Other libraries, modules, components, etc. owned by a company, e.g. bought with a license which allows for multiple applications or created by a methodology other than Scrum development methodology.

Because, on the one hand, the final product (Product A) contains both elements which were not specially developed for this product within Scrum sprints and the elements developed earlier (included in a common framework), and, on the other hand, some elements developed specially for Product A can be incorporated into a CF and reused in products for the same customer or other customers, it is necessary to have such a price setting model which will take into account relations between elements in the product. A natural way of separating various parts of a software system is to define its components. A general definition of a component used in software engineering states that it is a manageable, reusable, and swappable piece of software. In this paper we adopt a more restrictive meaning of a component used in UML [6], namely: "a component represents a modular part of a system that encapsulates its contents and whose manifestation is replaceable within its environment. A component defines its behavior in terms of provided and required interfaces. As such, a component serves as a type whose conformance is defined by these provided and required interfaces." Figure 2 shows an example of

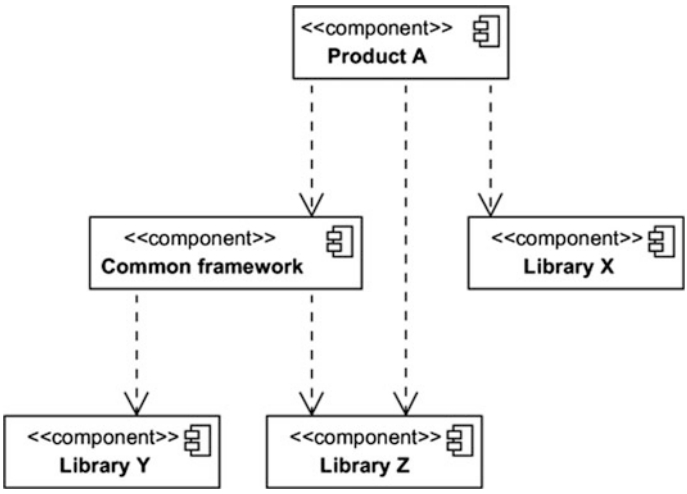

Fig. 2 An example of dependencies between system elements

dependencies between high-level components in the system of Product A. Component "Library" represents libraries owned by Software Company developing Product A, but not developed using Scrum methodology. Figure 4 shows that some components (in this example "Library Z") may be used in the final product as a result of more than one dependence—in this case "Library Z" is directly used by component "Product A" and indirectly through "Common Framework".

In many cases the structure of component dependencies constitutes a directed acyclic graph (DAG). However, in some situations a dependency graph can have cycles (sometimes cycles in dependency graphs indicate errors in the design and can be eliminated during the refactoring process). This happens when component A provides some interfaces for Component B and simultaneously (directly or indirectly) requires other interfaces from Component B. Such a situation is shown in Fig. 3. In case of cyclic dependencies between components in the method of managing costs presented in this article, it is necessary to create a composition of all components in a dependency cycle and treat them as a single component in the price setting process, which is indicated in Fig. 3 as "Composite component".

Having a dependency graph and knowing the values of particular components, it is possible to calculate the total cost of a product. It includes two basic parts:

- The cost (value) of means spent (e.g. the total number of men-hours) directly in the production of Product A. This part can be calculated with the use of standard price settings models used in Scrum and described above.
- The cost of components used to produce Product A (e.g. components from a common framework, libraries of other companies, etc.).

A dependency graph $G_D = (V, E)$ for product P_X is formed by a collection of vertices V and a collection of directed edges E. In our case, vertex $value_i \in V$ represents component c_i (used to produce P_X), so we will write c_i instead of $value_i$

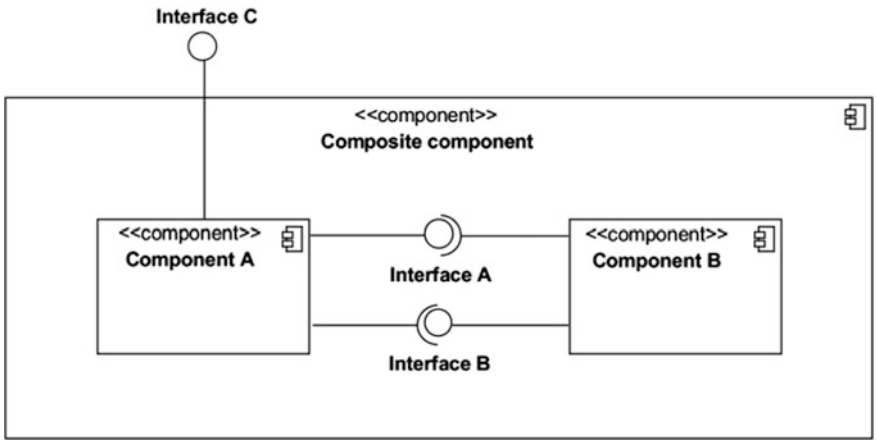

Fig. 3 A composite component created on the basis of components with cyclic dependencies

later. The set V contains all components used directly or indirectly to produce P_X. We can define the attribute cost for components from V and associate value cost (c_i) for vertex c_i, which is the function from V to R. In such case, the total cost of product P_X denoted as $cost_{TOTAL}$ can be expressed by the formula:

$$cost_{TOTAL} = \Sigma_V cost(c_i) \quad (4)$$

Note that V also contains PX, so the sum above contains the cost of means spent directly in production PX. In the approach proposed above we charge a given component c_i only once regardless of how many components from V depend on c_i. It reflects the idea that a customer pays for a given component only once, no matter how it is used in product PX.

6 Creation of Dependency Graph

If, for a given product, we want to use a method of calculating costs based on usage of components, it is necessary to create a component dependency graph for deployment of a given product. An important decision to be made while creating a dependency graph is connected with establishing the granulation of components, that is establishing the boundaries of particular components. The smaller the components, the more precise the ability of the target product to trace actually used fragments of the code (components). However, smaller components must account for a greater number of dependencies and controlling actual connections in a given deployment. Tracking dependencies at a low level requires taking into account particular implementations of a given component—such a situation takes place when there are numerous alternative implementations of one component in a common framework. On the other hand, selecting coarse granulation of components decreases labor connected with following their usage, e.g. if components in a project created in the Java environment and using the Maven building tool are the same as Maven artifacts, it is possible to automatically obtain knowledge on the links used in a built project.

However, too coarse granulation can lead to erroneous calculations of components used. Figure 4 shows the same dependency graph as presented in Fig. 2, but its component "Common Framework" with all dependencies was aggregated to one component "CF with dependencies" which leads to erroneous calculations of components used in Product A: Library Z is calculated twice.

Building tools for Java which allow for automatic generation of dependency graph include e.g.: Gradle [5], Apache Ivy [12] and, already mentioned above, Apache Maven with maven-dependency-plugin [2]. Figure 5 shows an example of

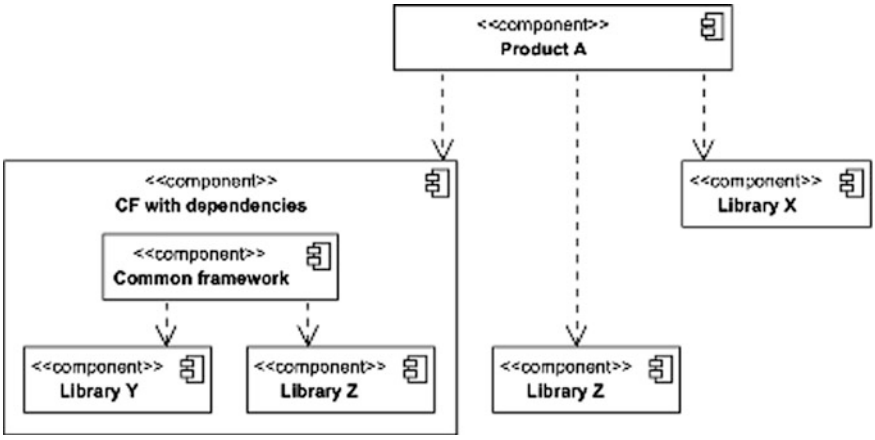

Fig. 4 A dependency graph with coarse granulation of components

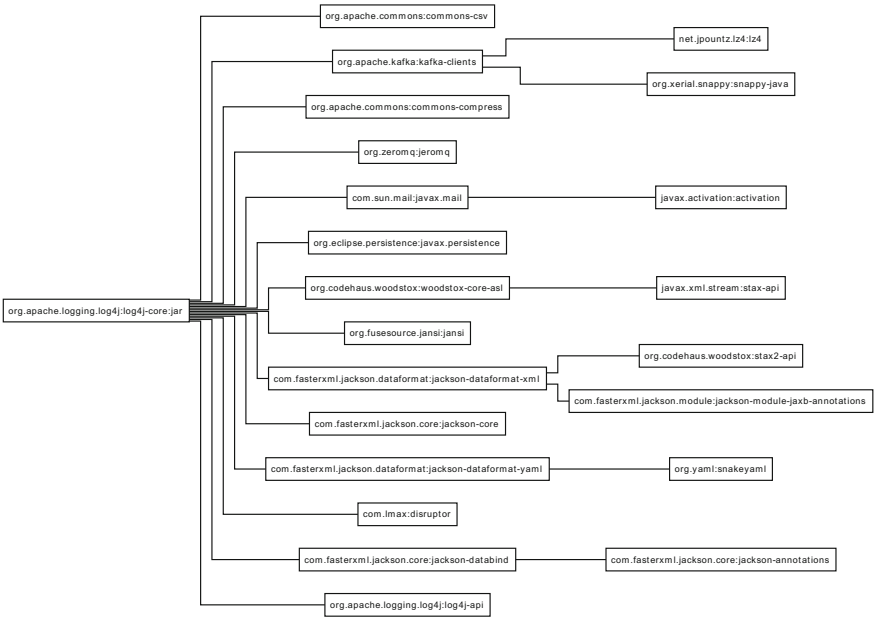

Fig. 5 An example of automatically generated dependency graph for software components

a dependency graph for log4j-core module of Apache Log4j library [1], which is one of top three libraries most often used by open source Java projects on GitHub. The graph was automatically generated using maven-dependency-plugin.

7 Aspects of Pricing Model Validation

Before it is deployed in a real company, the proposed solution should be validated by being compared to other similar solutions or by consulting experts in the field. In other branches similar solutions are known as cooperative or participative development. Agriculture is an interesting branch in which certain similarities to the proposed method can be found. The document FAO UN ACD [3] presents the development of agricultural cooperatives, which includes values and principles of cooperatives, conditions needed to create a cooperative, conditions for continued success of cooperative organizations and potential advantages of cooperative organizations. In [15] a business model of canvas of energy cooperatives is presented with the intention to provide insight into possible opportunities for a co-operative organization in the energy sector. To check the adequacy of the proposed solution, 23 postgraduate students of IT project management studies have been asked under which condition they would accept a possibility of common framework development and which pricing model they prefer. The students have on average 3–5 years of working experience in IT companies. They have knowledge on classical and agile project management.

Question A asked about situations in which they would opt for a common development software subsystem. The following answers were possible: (A.1) Lack of funds for development, (A.2) Lack of sufficient human resources, (A.3) Lack of appropriate competencies for development of a subsystem, (A.4) Improving competitiveness through the involvement of others in development of a subsystem, (A.5) Acceleration of development (slow development of the system will be obsolete).

Question B asked which pricing settlement of the joint development of a software subsystem was the best. The following answers were possible: (B.1) Fixed price specified in the contract, (B.2) The costs of leased human resources (outsourcing), (B.3) Sharing the cost of components as part of a common variable, (B.4) Settlement of the delivered product depends on the used components of the discount for often-sold components.

Question C asked about the best proposal to measure the cost of common development of a software subsystem in the long run. The following answers were possible: (C.1) The complexity of the created software, (C.2) Men-days for software development, (C.3) Scrum Coin cost of a standard Scrum team (5, sprint 1 week), (C.4) Standardized stories, (C.5) The cost calculated in a particular currency (e.g. Euro) and its conversion to other currencies.

The preferences were calculated as an average of order choice. Figure 6 shows that sharing development of common software framework development was preferred when there was a lack of, respectively: human resources (A.2), appropriate competences (A.3), and funds (A.1). In the joint development of common framework sharing development costs (B.3) was the most frequent option, which was followed by sharing the costs depending on components usage (B.4). The most popular pricing was based on men-days (C.1), then Scrum coins, and then

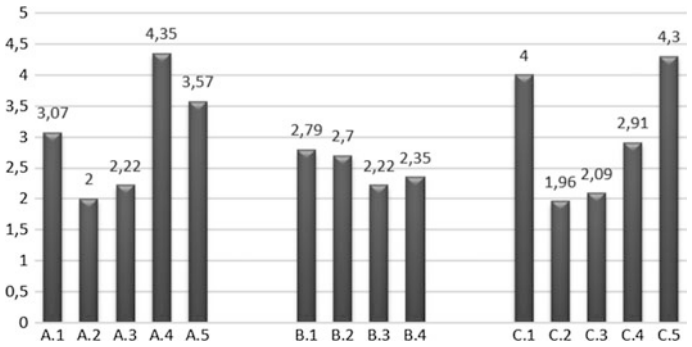

Fig. 6 The results of the survey—average preference values

standardized Scrum stories. What is interesting, the pricing model based on particular currency was not considered a good proposal.

The results of the survey indicate that the proposed pricing model seems to be an interesting option in developing a common software framework in a multi-project environment. There is a need for further investigation of business models of other organizations to validate the approach presented in the paper. Before its implementation, additional adjustments of the model and broader validation should be conducted.

8 Conclusions

The pricing methodology described in the article offers tools for establishing a new line of practices for Scrum methodology which will allow for accurate cost estimation and software license distribution for product backlog elements when a constantly developed common software framework is used.

Scum methodology is iterative and customer feedback regarding currently implemented functionality is periodically acquired by development teams. Customer's decisions can now be additionally supported by continuous licensing cost estimation. Actual and accurate common framework licensing costs can be presented each time a planning meeting is held. Cost calculation itself does not require considerable effort, so it can be used as a standard clue for a customer during a decision making process. In practice, each time a potentially shippable product increment is presented, a customer gets information about currently estimated licensing costs.

It seems that a full component dependency tracing process can easily be applied during development of integration tests in Scrum methodology. Having an exact trace of method invocations which is generated during a testing process, a knowledge base describing communications between components can be built. Method invocations from a common framework will, of course, be the center of

attention. Assuming that the proper component size granularity is kept and full knowledge of component dependencies is possessed, a secondary knowledge of component usage costs can automatically be established, with a minimum additional effort involved (focused only on the process supervision actions).

References

1. Log4j—Log4j 2 Guide—Apache Log4j 2. http://logging.apache.org/log4j/2.x/
2. Maven—Welcome to Apache Maven. https://maven.apache.org/
3. Agricultural Cooperative Development. A manual for trainers. Food and Agriculture Organization of the United Nations. http://www.fao.org/3/a-x0475e.pdf (1998)
4. The Open Group. ArchiMate 2.1 Specification. http://pubs.opengroup.org/architecture/archimate2-doc/toc:html (2013)
5. Gradle l Modern Open-Source Enterprise Build Automation. http://gradle.org/ (2015)
6. OMG Unified Modeling Language (OMG UML), Version 2.5 (2015)
7. Scaled Agile Framework—SAFe for Lean Software and System Engineering. http://www.scaledagileframework.com/ (2016)
8. blogs.versionone.com: Normalization of Story Points—Part 5 of 5 The Agile ManagementBlog. https://blogs.versionone.com/agile-management/2013/12/02/scalable-agile-estimation-and-normalization-of-story-points-calibrated-normalization-method-for-top-down-estimation-part-5-of-5/ (2013)
9. Cohn, M.: Agile Estimating And Planning. Robert C. Martin Series. Prentice Hall Professional Technical Reference, Upper Saddle River (2006)
10. Derhamy, H., Eliasson, J., Delsing, J., Priller, P.: A survey of commercial frameworks for the Internet of Things. In: 2015 IEEE 20th Conference on Emerging Technologies Factory Automation (ETFA), pp. 1–8 (2015)
11. Edwin, N.M.: Software frameworks, architectural and design patterns. J. Softw. Eng. Appl. **7** (8), 670–678. http://www.scirp.org/journal/PaperDownload.aspx?DOI=10.4236/jsea.2014.78061 (2014)
12. Ivy, A.: The agile dependency manager. http://ant.apache.org/ivy/ (2015)
13. Joshi, R., Bose, A., Scott, B.: Software Platforms for Unmanned Systems, Softwareplatform characteristics and developer advantages (2015)
14. Kaner, C., Bond, W.P.: Software Engineering Metrics: What Do They Measure and How Do We Know? (2004)
15. Küller, P., Dorsch, N., Korsakas, A.: Energy co-operatives business models: Intermediateresult from eight case studies in southern Germany. In: 2015 5th International Youth Conference on Energy (IYCE), pp. 1–8 (2015)
16. Nguyen-Cong, D., Tran-Cao, D.: A review of effort estimation studies in agile,iterative and incremental software development. In: 2013 IEEE RIVF International Conference on Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), pp. 27–30 (2013)
17. Popli, R., Chauhan, N.: A sprint-point based estimation technique in Scrum. In: 2013 International Conference on Information Systems and Computer Networks (ISCON), pp. 98–103 (2013)
18. PricewaterhouseCoopers: Software Revenue Recognition A User-Friendly Guide for Navigating through the Many Complexities. www.pwcrevrec.com (2009)
19. Ramasubbu, N., Balan, R.: Overcoming the challenges in cost estimation for distributed software projects. In: 2012 34th International Conference on Software Engineering (ICSE), pp. 91–101 (2012)

20. Scrum.org: Normalization of story points between Scrum Teams. https://www.scrum.org/Forums/aft/564 (2014)
21. Turek, M., Werewka, J.: Multi-Project Scrum Methodology for Projects Using Software Product Lines. Springer, Berlin (2016)
22. Viswanath, U., Betz, S.: Value Measurement on Performance in Global Software Development: An Industry Study on How R & D can Influence "Value for Money". In: 2015 IEEE 10th International Conference on Global Software Engineering Workshops (ICGSEW), pp. 12–17 (2015)

Environmental Management Systems According to ISO 14001

Jolanta Pakulska and Małgorzata Rutkowska-Podołowska

Abstract The aim of the paper is to analyze the ISO 14001. Environmental management in enterprises has become a requirement of our time. It is hard to imagine the functioning of a large company that does not include the management of environmental aspects. Support for environmental management standard is ISO 14001. It determines how the company should manage the environmental aspects. In the paper the creation of standards from the ISO 14000 series has been presented. Further discussed ISO 14001 which deals with environmental management system. In the next part procedures for the implementation of environmental management according to ISO 14001 have been reviewed. At the conclusion pointed to a growing number of companies that can boast a certified ISO 14001. The study used the method of conceptual and analytical method which used analysis of source materials and the available literature considerations, as well as the available statistical data.

Keywords Environment · ISO · Environmental management system

1 Introduction

ISO (International Organization for Standardization) is an international organization bringing together national standardization organizations. It is a non-governmental organization, although some of its member organizations work in government structures. ISO standard establishes practically all fields, their compliance is

J. Pakulska (✉)
Faculty of History and Social Sciences, Institute of Sociology,
Cardinal Stefan Wyszynski University in Warsaw, Warsaw, Poland
e-mail: j.pakulska@uksw.edu.pl

M. Rutkowska-Podołowska
Wydział Informatyki i Zarządzania, Katedra Infrastruktury Zarządzania,
Politechnika Wrocławska, Wrocław, Poland
e-mail: malgorzata.rutkowska-podolowska@pwr.edu.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*,
Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_11

voluntary. "The ISO 14000 family of standards provides practical tools for companies and organizations of all kinds looking to manage their environmental responsibilities. ISO 14001:2015 and its supporting standards such as ISO 14006:2011 focus on environmental systems to achieve this. The other standards in the family focus on specific approaches such as audits, communications, labeling and life cycle analysis, as well as environmental challenges such as climate change" [1].

Universally valid definition of the environmental management system is based on ISO 14001 "part of the overall management system that includes organizational structure, planning, responsibilities, practices, procedures, processes, and activities for the development, implementation, and improvement, reviewing and maintaining the environmental policy" [2].

2 ISO 1400 Series

The beginning of the ISO 14000 series dates back to 1991 when a group of 100 experts from 20 countries and 11 international organizations began working on the standard which was designed to support the organization's concept of sustainable development which was established at the Earth Summit in Rio de Janeiro. Technical Committee "Environmental Management" was founded in 1993 and the first series of draft standards were presented in 1994. The result is a whole series of standards 14000 (see Table 1). The standard ISO 14001 "Environmental management systems—Specification with guidance for use" applies from September 1996 (Later was modified) [3].

The use of standards 14000 series allows the improvement of environmental management which is especially important in a situation of tightening standards and rising costs for environmental projects undertaken by the company. The aim of the standards of this series is to limit the impact of the organization on the surrounding natural environment. They set the standards for the implementation of environmental management. Included in their organizational structure, action plans, responsibilities, procedures and resources necessary to prepare, use, and review and maintenance of the environmental policy of the company [4].

Table 1 The ISO 14000 series

Specification	Standard
Environmental management system	ISO 14001, ISO 14004
Environmental audit	ISO 14010–14012, ISO 14015
Eco-labelling	ISO 14021, ISO 1024, ISO 14025
Evaluation of environmental performance	ISO 14031, ISO 14032
Life Cycle Assessment	ISO 14040–14043, ISO 14048, ISO 14049
Vocabulary and definitions	ISO 14050
Environmental aspects in product standards	ISO 14061

Source [9], p. 92

3 ISO 14001

Guidelines for environmental management are contained in ISO 14001 which is the basis for issuing the certificate. This standard is derived from ISO 9000 on quality management system in the enterprise. The purpose of this standard is to achieve continuous improvement in environmental quality, as well as improving the relationship between the effects of human activity and the environment. Maintaining a balance in the environment requires uniform management of access to environmental resources; eliminate the negative effects of economic activity and rational use of natural resources. The ISO 14001 standard includes requirements which must be met is the basis for issuing a certificate of conformity for the environmental management system [5]. Environmental management is the management of such activities (processes) organization that adversely affect the environment to minimize the impact. It concerns the planning, organizing, motivating and control activities, resulting in decreases negative impact on the environment organization. The ISO 14001 standard defines the environmental management system as "part of the organization's management system, used to develop and implement its environmental policy and manage its environmental aspects" [6].

Environmental management according to ISO requires follow the cycle of Deming. "W. Edwards Deming in the 1950's proposed that business processes should be analyzed and measured to identify sources of variations that cause products to deviate from customer requirements. He recommended that business processes be placed in a continuous feedback loop so that managers can identify and change the parts of the process that need improvements. As a teacher, Deming created a (rather oversimplified) diagram to illustrate this continuous process, commonly known as the PDCA cycle for Plan, Do, Check, Act:

- PLAN: Design or revise business process components to improve result
- DO: Implement the plan and measure its performance
- CHECK: Assess the measurements and report the results to decision makers
- ACT: Decide on changes needed to improve the process.

Deming's focus was on industrial production processes, and the level of improvements he sought were on the level of production. In the modern post-industrial company, these kinds of improvements are still needed but the real performance drivers often occur on the level of business strategy. Strategic deployment is another process, but it has relatively longer-term variations because large companies cannot change as rapidly as small business units. Still, strategic initiatives can and should be placed in a feedback loop, complete with measurements and planning linked in a PDCA cycle" [7] (Fig. 1).

The standard it can be used by any organization or on its specific organizational units or types of activities. The key is continuous improvement. The basic elements of an environmental management system specified in ISO 14001 are:

- Environmental Policy;
- Planning environmental management system;

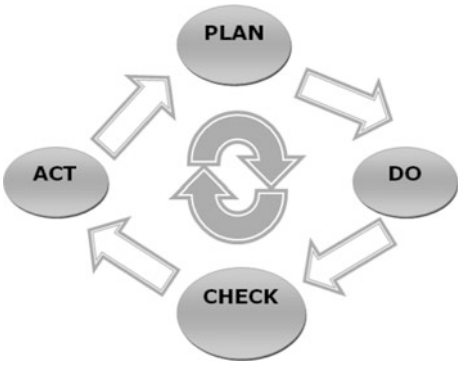

Fig. 1 The Deming cycle [11]

- Implementation and operation of the environmental management system;
- Checking and corrective and preventive action;
- A review of the environmental management system.

The standard defines the environmental policy of the organization as a declaration on its intentions and principles relating to the overall environmental performance, providing a framework for action and setting environmental objectives and targets of the organization. This policy is the starting point for planning the environmental management system. They should find there is a commitment to:

- "continuous improvement which is the continuous improvement in environmental efforts and reducing the negative environmental impact;
- pollution prevention or take action leading to the avoidance, reduction and monitoring of emissions, with the greatest emphasis on the prevention of pollution at source;
- Comply with legal requirements and other regulations relating to the organization" [4].

Environmental policy is the basis for decision-making related to the impact of the organization on the environment, but also indicates the actions and communicates to stakeholders involvement in the activities of pro-ecological entity. This policy should be periodically reviewed and amended regularly [8].

According to the standard continuous improvement is the process of improving the management system, which reduces the impact on the environment. This does not mean that if the level of impact on the environment is in line with current standards organization does not have to be perfect.

Planning the environmental management system, the company should take into account the elements required by the standard, or environmental aspects (e.g. waste or sewage), legal requirements and objectives, tasks and environmental management program. As a result, the company focuses on the key areas of its impact on the environment. Planning is a continuous process [9]. According to the standard environmental aspect is a "member of the organization, the products or services that can interact with the environment" [8]. According to the standard environmental objective is a "specific requirement for environmental performance, applicable to

the organization or parts thereof, arising from the environmental objectives, applicable to the organization or parts thereof, arising from the environmental objectives which should be defined and realized in order to achieve these objectives" [8]. While the environmental task is "specific requirement for environmental performance, applicable to the organization or parts thereof, arising from the environmental objectives which should be defined and realized in order to achieve these objectives" [8]. After defining aspects and impacts, requirements (legal and other), the formulation of environmental policy and setting goals and objectives developed a way to implement them in the practice of the organization.

During the planning process:

- "Are reviewed preliminary, during which analyzed the actual situation of a company in the field of environmental management. Initial inspection is voluntary and is not required by the standard;
- on this basis are identified and evaluated environmental aspects of a company, i.e. the type and extent of the impact of business on the environment;
- in addition to this, a check that legal and other requirements are relevant and whether the company observes them;
- The results of these steps are the basis for defining the environmental policy of the company and determine the specific objectives and environmental program" [2].

After the first plan system follows its implementation. The standard indicates that at this stage you should identify the resources, roles, responsibilities, powers, method of communication (both internal and external). Stages of implementation of the system [10]:

- definition of the responsibilities and powers;
- identification of competencies, training needs and the scope of training;
- develop a communication system;
- developing a system of documentation and supervision of documentation;
- implementation of the operational control through the use of written procedures and instructions for activities that significantly affect or may affect the environment, and to ensure that these activities will be supervised;
- Plan emergency response.
- Implementing the system:
- "they identified are specific to the business management processes, main and supporting and then described and shaped with regard to environmental issues;
- this is the basis for determining the resources, tasks, responsibilities and powers in order to further develop the environmental awareness of employees by means of training and designate instruments of internal and external information and communication environment;
- All activities and determine environmental management are documented and the documents are the book environment. Dossiers shall be prepared on the basis of the standard;
- In addition, you need to identify (potential) emergency situations and to undertake emergency measures to protect against threats" [2].

According to ISO 14001 it is necessary to establish and maintain procedures for regular monitoring activities which have (or could have) a negative impact on the environment. The aim is to achieve the objectives of environmental organizations. Monitoring and measurement is an essential element for the effectiveness of the environmental management system. Detected problems should be corrected as soon as possible. This stage is the implementation of the basic assumptions of the environmental management system, namely continuous improvement [4].

At the stage of verification is done:

- "all the above are checked by internal audits and measured with regard to the designated environmental policy, objectives and programs of environmental and legal requirements;
- The results are communicated regularly through the so-called top management. Maintenance management (management reviews)" [2].

Overview of the implemented system is to assess the compatibility of environmental policy, as well as the comparison of goals and objectives and their practical implementation. It is the basis for making changes in the system and adapting to changing circumstances [2].

The environmental management system is certified independent unit. The certificate is valid for 3 years and is issued on the basis of the controls carried out in the organization. Certification is supplemented by an annual audit control.

ISO 14001 is one of the most popular standards for environmental management which is reflected in more than 325 thousand certified organizations from over 170 countries (as of end 2014). This number is constantly growing (see Fig. 2). More than half of them organization in the East Asia and Pacific; Europe is the second (almost 40 %). The two continents have registered a total of almost 90 % of all certified organizations. Among the European countries, most organizations registered in Italy (over 27 thousand), United Kingdom (almost 17 thousand) and Spain (almost 14 thousand). Poland, with the number over 2.2 thousand, deals 13th place. In 2014, the largest organization in the world had a certificate in China. and the

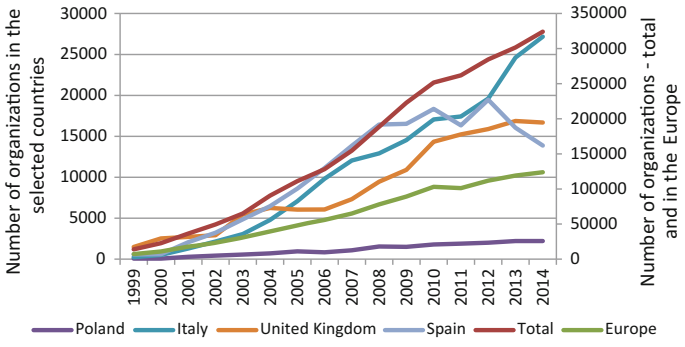

Fig. 2 Number of the certificated organisations (based on the [4])

Table 2 Countries with the largest number of certificates ISO 14001

Year/ country	1999	2000	2001	2002	2003	2004
China	222	510	1085	2803	5064	8862
Italy	243	521	1295	2153	3066	4785
Japan	3015	5556	8123	10,620	13,416	19,584
United Kingdom	1492	2534	2722	2917	5460	6253
Spain	573	600	2064	3228	4860	6473
Romania	1	5	15	45	96	361
France	462	710	1092	1467	2344	2955
Germany	962	1260	3380	3700	4144	4320
USA	636	1042	1645	2620	3553	4759
Czech Republic	60	116	174	318	519	1288

Year/ country	2005	2006	2007	2008	2009
China	12,683	18,842	30,489	39,195	55,316
Italy	7080	9825	12,057	12,922	14,542
Japan	23,466	22,593	27,955	35,573	39,556
United Kingdom	6055	6070	7323	9455	10,912
Spain	8620	11,125	13,852	16,443	16,527
Romania	752	1454	2269	3884	6863
France	3289	3047	3476	3482	4678
Germany	4440	5415	4877	5709	5865
USA	2064	5585	5462	4974	5225
Czech Republic	2122	2211	2731	3318	4684

Year/ country	2010	2011	2012	2013	2014
China	69,784	81,993	91,573	104,735	117,758
Italy	17,064	17,418	19,615	24,618	27,178
Japan	34,852	30,397	27,774	23,723	23,753
United Kingdom	14,346	15,231	15,883	16,879	16,685
Spain	18,347	16,341	19,470	16,051	13,869
Romania	7418	7394	8524	8744	9302
France	5251	7771	7094	7940	8306
Germany	6001	6254	7015	7983	7708
USA	4407	4957	5699	6071	6586
Czech Republic	6629	4451	4215	4792	5831

Based on the Ref. [4]

number of these organizations over the years is steadily growing (see Table 2). Among the European countries already on the second position in Italy. European countries predominate among the top 10 countries since there are up to 7.

Taking into account the sectors (see Fig. 2) in which organizations operate certified ISO 14001 prevails construction (in 2014 almost 44 thousand. Organization, with more than 254 thousand. Registered, which is more than 17 %). Over the years, it has significantly increased the number of organizations in this sector that can boast of having the certificate (in 1998. Less than 300 organizations), systematically increasing the share of this sector (from just over 4 % in 1998).

The situation is different in other leading sectors. Of course, it increased the number of organizations in each sector (which is associated with the popularization

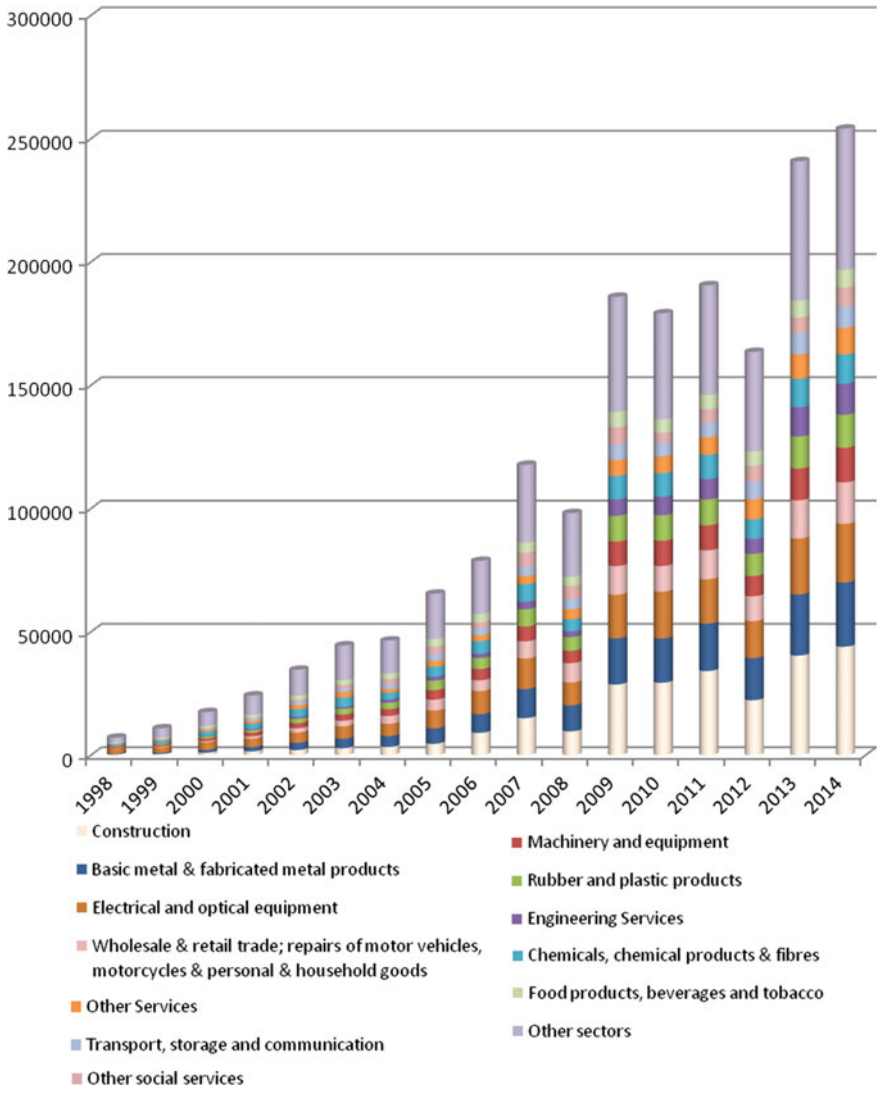

Fig. 3 Number of the certificated organisations sorted by sectors (based on the [4])

of environmental management according to ISO 14001), but the share of sectors in the majority of cases decreased or increased only slightly. The exception is here basic metal and fabricated metal products, where the share increased from just over 4 % to over 10 % (the number of organizations with less than 300 to more than 26 thousand). Significant increase was also recorded in the wholesale and retail trade; repairs of motor vehicles, motorcycles and personal and household foods (the share has increased from less than 2 %, almost 130 organizations, to nearly 6.5 %, more than 15.5 thousand organization). Participation by organizations increased slightly only in Rubber and plastic products (193 organizations, nearly 3 % to almost 13 thousand organizations and 5.4 %), Engineering Services (106 organizations 1.5 %, to almost 12 thousand. The organization and almost 5 %). Other Services (over 200 organizations, nearly 3 % to almost 10 thousand organization and just over 4 %) and other social services (almost 30 organizations, less than 1 % to more than six thousand. organization and over 2.5 %). The share of other sectors over the years has fallen (Fig. 3).

References

1. http://www.iso.org/iso/home/standards/management-standards/iso14000.htm
2. Brauweiler, J.: Rodzaje i zastosowania koncepcji zarządzania środowiskiem, w: Zintegrowane zarządzanie środowiskiem. Systemowe zależności między polityką, prawem, zarządzaniem i techniką, Kryński, A., Kramer, M., Caekelbergh, A.F. (eds.) Oficyna a Wolters Kluwer bussines, Warszawa (2013)
3. Matuszak-Flejszman, A.: Jak skutecznie wdrożyć system zarządzania środowiskowego według normy ISO 14001, 62, Wydawnictwo PZiTS, Poznań (2001)
4. Fura, B.: System zarządzania środowiskowego ISO 14001 a efektywność przedsiębiorstw. Zagadnienia teoretyczne i praktyczne, 62, Wydawnictwo Uniwersytetu Rzeszowskiego, Rzeszów (2011)
5. Koneczna, R.: CSR i jego narzędzia jako element strategii firm—praktyczny przewodnik, Wydawnictwo IGSMiE PAN, Kraków (2014)
6. Systemy zarządzania środowiskowego. Wymagania i wytyczne stosowania, PKN, Warszawa (2005)
7. Arveson, P.: The Deming Cycle. https://balancedscorecard.org/Resources/Articles-White-Papers/The-Deming-Cycle
8. Nierzwicki, W.: Zarządzanie środowiskowe, PWE, Warszawa (2006)
9. Matuszak-Flejszman, A.: Determinanty doskonalenia systemu zarządzania środowiskowego zgodnego z wymaganiami normy ISO 14001, Wydawnictwo UE w Poznaniu, Poznań (2010)
10. Poskrobko, B.: (eds.) Zarządzanie środowiskiem, PWE, Warszawa (2007)
11. http://www.praktykomania.ue.wroc.pl/?page_id=473
12. http://www.iso.org/iso/home/standards/certification/iso-survey.htm?certificate=ISO%209001&countrycode=AF
13. Matuszak-Flejszman, A.: Jak skutecznie wdrożyć system zarządzania środowiskowego według normy ISO 14001, 62, Wydawnictwo PZiTS, Poznań (2001)

Pre-testing of Polish Translation of System Usability Scale (SUS)

Anna Borkowska and Katarzyna Jach

Abstract This study demonstrates the Polish translation of System Usability Scale (SUS). The translation was made according to the procedure of psychologic test adaptation and was made in several steps: (1) forward-translation, (2) formulating the list of problems, (3) comparison of all translations versions and translations synthesis, (4) procedure of collating (5) results comparison of both Polish and English versions of SUS questionnaire, (6) checking the validity of result in comparison to other five usability metrics. It was validated that features with The reliability of Polish SUS version was confirmed by high Cronbach's alpha value (0.805), high correlations between adequate items of Polish and English SUS version as well as between total Polish and English SUS score. The validity of Polish SUS version measured by correlations with other usability metrics, including Computer System Usability Questionnaire (CSUQ) and Net Promoter Score (NPS) is sufficient. The Polish version of System Usability Scale is a valuable tool for quick usability assessment.

Keywords Validity · Reliability · Computer system usability questionnaire · Net promoter score

1 Introduction

The System Usability Scale (SUS) has been introduced as "quick and dirty" usability metric by Brooke [1] and it has been developed as popular usability evaluation measure since 30 years. However, in the meantime the metric changed into validated and reliable tool. Tullis and Stetson [2] found that among five different survey tools for measurement web sites usability, System Usability Scale

A. Borkowska · K. Jach (✉)
Wrocław University of Technology, Wrocław, Poland
e-mail: katarzyna.jach@pwr.edu.pl

A. Borkowska
e-mail: anna.borkowska@pwr.edu.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_12

gives the most reliable results, even if the sample size is small (8–12 users). On the basis of almost 2400 SUS usability tests Bangor et al. [3] found the SUS as highly reliable metric. Other research found the SUS reliability of 0.85, with possible underestimating due to low number of reported users participating [4]. Besides the reliability, Bangor et al. [3] shown the high usefulness of SUS survey for different types of software and hardware interfaces, including graphical user interfaces, web sites, internet-based applications, cell phone equipment, customer premise equipment (phones, modems, etc.), interactive voice response phone systems.

Although System Usability Scale is very popular, which is proved by thousands of researches (i.e. over 2300 SUS usability tests collected by Bangor et al. [3]), there is no validated translations into other languages. Brooke reported some adaptations on European languages were made, but they have not survived. Additionally, some not validated translations are in use in Spanish and French [5]. At least two German versions of SUS already exist [6, 7]. Recently, the attempt of validation of European Portuguese version of SUS was made by Martins et al. [8]. The validity of Persian language version of System Usability Scale was established by Dianat et al. [9], which shows SUS is adaptable in different culture areas.

The main goal of this pilot study was to prepare Polish translation of System Usability Scale for further reliability and validation research.

2 Methods

2.1 The Tool

The original SUS questionnaire comprised 10 statements scored on a 5-point Likert's scale. The positive and negative statements are alternated in order, so during filling the questionnaire the concentration of participant is needed. Because of the opposite purport/meaning of the following statements, different method of result counting are used for different questionnaire items [5]. Final score for the SUS can range from 0 to 100, where higher scores indicate better usability [1]. In comparison to the original SUS questionnaire, one modification was made in the English version of survey. The word "cumbersome" used in 8th item was replace with "awkward". The same modification was made by Bangor et al. [3] and independently by Finstad [10], who investigated SUS among non-native English speakers and found the word "awkward" as more understandable and clear. Lewis and Sauro [11] confirmed this replacement as well. Another debatable point of SUS questionnaire is using the word "system" or "product" or direct reference to the actual assessed product name [5]. According to Lewis and Sauro [11] recommendation all above terms are appropriate but the consistency in using the same terms across all the survey is important. Therefore, for English version of SUS the original term "system" was applied. Another rationale behind this decision was the lexical one, because in Polish language a word "product" is associated more with a physical item than with the software or virtual application.

2.2 Experiment Design and Procedure

The translation was made according to the procedure of psychologic test adaptation [12] and was made in several steps: (1) forward-translation, (2) formulating the list of problems, (3) comparison of all translation versions and translations synthesis, (4) procedure of collating, (5) comparison of results both Polish and English versions of SUS questionnaire, (6) checking the validity of result in comparison to other five usability metrics. Forward-translation made by six different translators, among them two usability experts. All the translators were Polish native speakers, which is compatible with WHO recommendations [13]. The list of problems found was made on the basis of all the translations. Then, six versions of translation were compared with a special attention being paid for the list of problems. According to WHO guidelines, the approach to translating was chosen, emphasizing conceptual rather than literal translations [13].

As an example of procedure translation comparing, 8th item is shown in the Table 2. It was one of the most difficult which resulted in huge discrepancies among different translations. The first problem was choosing between past tense (as original) and present one, which is more natural for Polish language. Because of frequency of use in the common language as well as consistency of the questionnaire, it was decided to use present tense in all the items. The other problem was translation of word "awkward". Experts used the words (English synonyms in brackets): "dziwny" (strange), "dziwaczny" (weird), "skomplikowany" (complicated), "niewygodny" (inconvenient), "niezręczny" (gauchy). In similar cases, like translation of words "inconsistency" (6th item), beside the different versions of expert translation, Polish and English dictionary of synonyms was used. The word "niewygodny" was chosen because it suits the best to the meaning intention of "awkward" (Table 1).

Table 1 Set of version of the translation exemplary SUS item

Original item	Expert translations	The final version
I found the system very awkward to use	Ten produkt wydał mi się dziwny/dziwaczny w użyciu	Odbieram ten system jako bardzo niewygodny w użyciu
	Uważam, że ten produkt był skomplikowany w użyciu	
	Używanie produktu wydawało mi się bardzo niezręczne	
	Odbieram system jako niewygodny, trudny w użyciu	
	Uważam, że produkt był bardzo dziwny w użyciu	
	Stwierdzam, że system jest bardzo niewygodny w użyciu	

Table 2 Items content of Polish and English version of SUS

Item	Item content—Polish version	Item content—English version
1	Myślę, że często używałbym/używałabym tego systemu	I think that I would like to use this system frequently
2	Odbieram ten system jako niepotrzebnie skomplikowany	I found the system unnecessarily complex
3	Myślę, że system jest łatwy w użyciu	I thought the system was easy to use
4	Myślę, że potrzebowałbym/potrzebowałabym wsparcia asystenta, aby używać tego systemu	I think that I would need the support of a technical person to be able to use this system
5	Uważam, że różne funkcje tego system są dobrze zintegrowane	I found the various functions in the system were well integrated
6	Myślę, że w tym systemie jest za dużo niespójności	I thought there was too much inconsistency in this system
7	Wydaje mi się, że większość ludzi nauczyłaby się bardzo szybko używać tego systemu	I imagine that most people would learn to use this system very quickly
8	Odbieram ten system jako bardzo niewygodny w użyciu	I found the system very awkward to use
9	Czułem/czułam się bardzo pewnie używając tego systemu	I felt very confident using the system
10	Musiałem/musiałam nauczyć się wielu rzeczy, zanim zacząłem/zaczęłam właściwie posługiwać się tym systemem	I needed to learn a lot of things before I could get going with this system

Finally, the proposal of synthesis version was chosen for further research. The goal of this step was to identify and resolve the inadequate expressions/concepts of the translation. The gender applicability, which is a characteristic feature of Polish language was considered. The final version of Polish System Usability Scale is shown in the Table 2.

In order to check reliability to compare English and Polish SUS scores, some other usability metrics were added. First of all, the adjective scale proposed by Bangor, Kortum and Miller was applied as 11th item in SUS questionnaire. The respondents are asked for the evaluation of the user-friendliness of the assessed product on the seven point scale from "worst imaginable" to "best imaginable" with "OK" as the middle value. The authors stated high correlation ($r = 0.822$; $\alpha < 0.01$) between the SUS score and identical adjective rating scale and gave the interpretation the adjective rating scale in comparison to System Usability Scale score [14]. Due to lack of Polish version, this item and the adjective rating scale was translated into Polish according to the same procedure as basic SUS questionnaire. Next two metrics added to basic SUS were based on the assumption that besides the SUS total score, the questionnaire lets to find how usable and learnable is the assessed

system, which was confirmed by factor analysis made by Lewis and Sauro [11]. To compare this two additional independent dimensions of System Usability Scale, two questions about usability and learnability of the assessed system with answers on 7 point adjective rating scale were added.

As the final metric used for comparison with SUS score the Net Promoter Score was applied. The Net Promoter Score was proposed as a single question for measuring the customer satisfaction by Reichheld [15]. The metric is created by subtraction the percentage of detractors from the percentage of promoters. Users are asked "How likely is it that you would recommend us to a friend or colleague?" and then provide a rating from 0 ("Not at all likely") to 10 ("Extremely likely"). Unlike the original, ten position scale range from 1 to 10 was used in order to be consistent with other questions used in the survey and to eliminate the middle answer. Using the 10 position scale, as promoters are classified the respondents with 9 or 10 answer and as detractors the subjects with answer from 1 to 6 [16]. An interpretation the Net Promoter Score in comparison to SUS score was made by Sauro [17]. Another advantage of the NPS is widely used and well established Polish version of the key question (i.e. [18–20]). On the other side, it is worth to notice Net Promoter Score is a satisfaction metric, so it covers just one from the three usability dimensions according to ISO 9241 standard.

All the above mentioned metrics are based on single test items. Therefore, the set of questionnaires used in the research was supplemented with Computer System Usability Questionnaire (CSUQ). Similarly to System Usability Scale, CSUQ has been developed since over 30 years. It consists of 19 items with 7 point Likert's scale [21]. Psychometric evaluation of CSUQ was checked by the author (Lewis [22] and earlier) and the questionnaire is widely used in different usability tests (i.e. [23, 24]). The overall CSUQ result is counted as the average of all items. Due to the scale used, lower result means better usability assessment.

The translated questionnaire was tested on the target population, including potential participants of SUS surveys. The validity of the translation was checked by bilingual responses procedure. The full set of questionnaires used during the test comprised three surveys: English version of System Usability Scale with some additional questions, English version of Computer System Usability Questionnaire (CSUQ) and Polish version of System Usability Scale. The half of the subjects start the testing procedure with fulfilling Polish translation SUS questionnaire, while another one start with English original. In order to minimizing the memory effect, both SUS versions were separated with Computer System Usability Questionnaire (CSUQ). Due to reverse scale used in SUS and CSUQ questionnaires the participants were asked for paying special attention on scales in order to avoid mistakes. At the very beginning of the research all the investigated subjects were ask for the assessment of online version or mobile application of Facebook, which was the system used by all the respondents. The real goal of the experiment was revealed to respondents just after filling the full set of questionnaires. The full investigation was conducted on the sample of 37 subjects, mostly students.

3 Results and Discussion

3.1 Reliability of Polish Version of System Usability Scale

The measure of internal consistency of the scale (reliability)—Cronbach's alpha reaches the value above 0.8 (0.805). This result considered the reliability of SUS and additionally shown all the items are significant for the final SUS score (see Table 2).

In order to compare results for Polish and English version of SUS, correlations between each pair of specific items for both languages were checked. The results are shown in Table 3. Besides the item no. 10 (at the limit of significance), all the Spearrman's rho values are relatively high and statistically significant. The lowest results can be observed for 4th and 10th item, which together cover learnability dimension according to Lewis and Sauro [17]. The high correlation between Polish and English SUS score confirm good quality of translation.

3.2 Comparison of Polish Version of System Usability Scale with Other Usability Metrics

The basic statistics of usability metrics used in the experiment are shown in the Table 4. Besides quite good consistency between Polish and English metrics, a visible relation was observed between mean values of SUS score, overall CSUQ result and one-item questions about user-friendliness and learnability. This conclusion can be supported with correlations results shown in the Tables 5 and 6.

All the correlations between Polish and English usability metrics are significant and relatively high, which confirm the reliability of Polish translation. Although the correlation between Polish and English answers on recommendation chance

Table 3 Reliability statistics for all items of Polish version of SUS

Item	Scale mean if item deleted	Scale variance if item deleted	Item total correlations	Cronbach's alpha if item deleted
1	28.8108	21.102	0.546	0.782
2	29.3243	19.503	0.561	0.779
3	28.9459	20.441	0.628	0.772
4	28.4595	23.589	0.411	0.799
5	29.4595	20.311	0.604	0.774
6	29.5405	18.755	0.574	0.778
7	29.1351	21.787	0.394	0.798
8	28.7297	21.369	0.685	0.774
9	29.3243	21.947	0.299	0.812
10	28.9459	22.608	0.289	0.809

Table 4 Correlations (Spearrman's rho) between items of Polish and English version of SUS

Item	1	2	3	4	5	6	7	8	9	10	SUS score
rho	0.874	0.453	0.826	0.358	0.612	0.426	0.734	0.586	0.514	0.307	0.803
Sig. (2-tailed)	0.000	0.000	0.005	0.030	0.000	0.009	0.000	0.000	0.001	0.064	0.000

Table 5 Descriptive statistics of usability metrics

Metrics	Min.	Max.	Mean	MSE*	Median	SD**
SUS score (Polish)	50.0	100.0	80.7432	2.07739	80.0	12.63627
SUS score (English)	40.0	100.0	79.3919	2.41679	82.5	14.70078
Overall CSUQ (English only)	1.11	4.53	2.3855	0.13625	2.3684	0.82876
User friendliness (Polish)	4.0	7.0	5.324	0.1166	5.0	0.7092
User friendliness (English)	4.0	7.0	5.297	0.1494	5.0	0.9087
Learnability (Polish)	4.0	7.0	5.162	0.1371	5.0	0.8338
Learnability (English)	3.0	7.0	5.000	0.1644	5.0	1.0000
Usability (Polish)	2.0	7.0	5.243	0.1914	5.0	1.1644
Usability (English)	2.0	7.0	5.297	0.1893	5.0	1.1514
Recommendation chance (Polish)	1.0	10.0	7.757	0.3496	8.0	2.1266
Recommendation chance (English)	1.0	10.0	7.784	0.3469	8.0	2.1100

*MSE—Mean Standard Error
**SD—Standard Deviation

Table 6 Correlations (Spearrman's rho) between Polish and English version of usability metrics

	User friendliness	Learnability	Usability	Recommendation chance	SUS score
rho	0.595	0.567	0.877	0.948	0.803

**All correlations significant at the 0.01 level (2-tailed)

question is extremely high (0.948; $\alpha < 0.001$), the NPS result counted as the subtraction of percentage of promoters and percentage of distractors is inconsistent (Polish NPS = 18.9, English NPS = 29.8) and does not match to the comparison made by Sauro [17], because with the mean SUS score around 80 the expected NPS value is over 70. The probable reason of this discrepancy is a significant number of neutrals. Therefore additional analysis was made with using data comparing mean value and 95 % confidence intervals for SUS scores for NPS detractors and

Table 7 Correlations (Spearrman's rho) between total scores of Polish and English version of SUS and other usability metrics

Metrics		SUS score (Polish)	SUS score (English)
Overall CSUQ*** (English only)	rho	−0.816**	−0.710**
User friendliness (Polish)	rho	0.596**	0.446**
User friendliness (English)	rho	0.417*	0.442**
Learnability (Polish)	rho	0.571**	0.442**
Learnability (English)	rho	0.535**	0.534**
Usability (Polish)	rho	0.646**	0.580**
Usability (English)	rho	0.590**	0.524**
Recommendation chance (Polish)	rho	0.542**	0.531**
Recommendation chance (English)	rho	0.508**	0.496**

*Correlation significant at the 0.05 level (2-tailed)
**Correlation significant at the 0.01 level (2-tailed)
***Results shown for the reverse scale

promoters for each single subject [17]. For Polish version this comparison does not match for 3 cases only (8 %), and for English one for 5 cases (13.5 %). As it was noticed above, the reason of this discrepancies can be that NPS takes into account satisfaction only, which is not full area of usability according to ISO 9241 standard.

Although all the correlations shown in the Table 6 are significant and relatively high, it is noticeable, that higher correlations can be observed among usability metrics in Polish version. Probably the reason is that for the respondents being Polish natives it was easier to express preferences in Polish. The high correlation between the SUS score and the identical adjective scale (user-friendliness) found by Bangor et al. [14] was not stated, even if values are still significantly high, especially for the relatively small sample (N = 37) (Table 7).

4 Summary

4.1 Limitations of the Research

The research was done among a young participants (the average age 20.7), mostly students. Bangor et al. [3] shown a small but significant negative correlation between age and SUS scores. Therefore, next studies should be provided on the sample of larger age range. Another limitation is using Polish version of System Usability Scale for the assessment of just single system. As it was mentioned above, the respondents assessed Facebook system. It was chosen because of its popularity. On the other side, this feature could influence some aspects of usability evaluation. Especially the recommendation chance declaration (Net Promoter Score) can be disturbed by this. Further investigations should be conducted on the larger and not

so homogenous sample, with using at least two systems on different levels of user-friendliness.

The back-translation procedure was not chosen as according to some research psychometric properties of a rapid translation (three initial translations, one synthesis, and two pretests) and a comprehensive adaptation (back-translations, focus groups, development of equidistant response options, item difficulty and quality ratings, and multiple pretests) gives very similar results [25]. Nevertheless, back-translation is widely used procedure and could be considered for further stages of validation the Polish version of SUS.

The learning effect for Polish and English version of SUS which was provided on the same group was not present because during the testing procedure it was a gap between these two versions. Nevertheless, in order to full eliminate this effect, next investigation can be provided on two independent language groups (Polish or English version only) or, alternatively, with longer time interval between both questionnaires, i.e. one week.

4.2 Final Conclusions

The presented pilot study confirms usefulness of Polish translation of System Usability Scale. The reliability measured by Cronbach's alpha reaches the value above 0.8 (0.805), which is enough. High and statistically significant correlations between single items in Polish and English SUS version as well as between total Polish and English SUS score (rho = 0.803; $\alpha < 0.001$) confirm the reliability of Polish SUS version. The validity of Polish SUS version measured by correlations with other usability metrics is sufficient, especially for Computer System Usability Questionnaire (rho = −0.816; $\alpha < 0.01$; CSUQ operates with a reverse scale), which is a reliable and validated metric [22]. These conclusions give the basis for further investigations of Polish version of SUS. As a next move, a research focuses on clarity and simplicity of SUS language is planned. Due to the lowest correlations results stated for item no. 10, special attention will be paid on this item. Further investigations will be provided on the larger and more age differentiated sample and various systems.

The described method of the assessment the translation process can be applied for other types of texts, e.g. medical data, legal or business documents. Due to elimination of the reverse translation it seems to be easier to use.

References

1. Brooke, J.: SUS: a "quick and dirty" usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) Usability Evaluation in Industry. Taylor and Francis, London (1996)
2. Tullis, T.S., Stetson, J.N.: A Comparison of Questionnaires for Assessing Website Usability. In: Usability Professionals Association (UPA) 2004 Conference, Minneapolis, USA.

Retrieved 02 Mar 2016 from https://www.researchgate.net/publication/228609327_A_Comparison_of_Questionnaires_for_Assessing_Website_Usability (2004)
3. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the System Usability Scale. Int. J. Human-Comput. Interact. **24**(6), 574–594 (2008). doi:10.1080/10447310802205776
4. Kirakowski, J.: The use of questionnaire methods for usability assessment. Unpublished manuscript. Retrieved from http://sumi.ucc.ie/sumipapp.html (1994). Accessed 13 May 2016
5. Brooke, J.: SUS: a retrospective. J. Usability Stud. **8**(2), 29–40 (2013)
6. Lohmann, K., Schäffer, J.: System Usability Scale (SUS)—An Improved German Translation of the Questionnaire. Published 18 Sept 2013, retrieved 08 Mar 2016. https://minds.coremedia.com/2013/09/18/sus-scale-an-improved-german-translation-questionnaire/ (2013)
7. Reinhardt, W.: Crowdsourcing the translation of SUS. Published 16 Jan 2012, retrieved 8 Apr 2016. http://isitjustme.de/2012/01/crowdsourcing-the-translation-of-sus/ (2012)
8. Martins, A.I., Rosa, A.F., Queirósa, A., Silvab, A., Rocha, N.P.: European Portuguese validation of the System Usability Scale (SUS). Procedia Comput. Sci. **67**(1), 293–300 (2015). doi:10.1016/j.procs.2015.09.273
9. Dianat, I., Ghanbari, Z., AshgariJafarabadi, M.: Psychometric properties of the persian language version of the system usability scale. Health. Promot. Perspect. **4**(1), 82–89 (2014). doi:10.5681/hpp.2014.011
10. Finstad, K.: The system usability scale and non-native english speakers. J. Usability Stud. **1** (4), 185–188 (2006)
11. Lewis, J.R., Sauro, J.: The factor structure of the system usability scale. In: Kurosu, M. (ed.) Proceedings of the 1st International Conference on Human Centered Design: Held as Part of HCI International 2009 (HCD 09), pp. 94–103. Springer, Berlin, Heidelberg (2009). doi: http://dx.doi.org/10.1007/978-3-642-02806-9_12
12. Hornowska, E., Paluchowski, W.J.: Kulturowa adaptacja testów psychologicznych. In: Brzeziński, J. (ed.) Metodologia badań psychologicznych—wybór tekstów, pp. 149–190. Wydawnictwo Naukowe PWN, Warszawa (2004)
13. World Health Organization: Process of translation and adaptation of instruments. http://www.who.int/substance_abuse/research_tools/translation/en/ (2016). Retrieved 23 Feb 2016
14. Bangor, A., Kortum, P.T., Miller, J.T.: Determining what individual SUS scores mean: adding an adjective rating scale. J. Usability Stud. **4**(3), 114–123 (2009)
15. Reichheld, F.F.: The one number you need to grow. Harvard Bus. Rev. **81**(12) 46–54 (2003)
16. Schneider, D., Berent, M., Thomas, R., Krosnick, J.: Measuring Customer Satisfaction and Loyalty: Improving the 'Net-Promoter' Score. http://van-haafen.nl/images/documents/pdf/Measuring%20customer%20satisfaction%20and%20loyalty.pdf (2008). Published June 2008, retrieved 10 Mar 2016
17. Sauro, J.: Does better usability increase customer loyalty? The Net Promoter Score and the System Usability Scale (SUS). http://www.measuringu.com/usability-loyalty.php (2009). Published 07 Jan 2009, retrieved 06 Mar 2016
18. Drapińska, A.: Pomiar lojalności klientów—wybrane wskaźniki. Polityki Europejskie, Finanse i Marketing, **58**(9), 125–136 (2013)
19. Kozielski, R.: Wskaźniki marketingowe, Wolters Kluwer Polska (2008)
20. Siejak, M.: Net Promoter Score—jedyny wskaźnik kondycji marki, o który powinieneś się troszczyć. www.marketingowiec.pl (2012). Published 03 July 2012, retrieved 03 May 2016
21. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int. J. Human-Comput. Interact. **7**, 57–78 (1995)
22. Lewis, J.R.: Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int. J. Human-Comput. Interact. **14**(3/4), 463–488 (2002)
23. Frias-Martinez, E., Chen, S.Y., Liu, X.: Evaluation of a personalized digital library based on cognitive styles: adaptivity vs. adaptability. Int. J. Inf. Manage. **29**, 48–56 (2009)

24. Kartakis, S., Stephanidis, C.: A design-and-play approach to accessible user interface development in Ambient Intelligence environments. Comput. Ind. **61**, 318–328 (2010)
25. Perneger, T.V., Leplège, A., Etter, J.-F.: Cross-cultural adaptation of a psychometric instrument: two methods compared. J. Clin. Epidemiol. **52**(11), 1037–1046 (1999). doi:10.1016/S0895-4356(99)00088-8

Part III
Knowledge Discovery and Data Mining

Weighted Fuzzy Genetic Programming Algorithm for Structure and Parameters Selection of Fuzzy Systems for Nonlinear Modelling

Krystian Łapa and Krzysztof Cpałka

Abstract In this paper a weighted fuzzy genetic programming algorithm for selection of structure and parameters of fuzzy systems for nonlinear modelling is proposed. This method is based on fuzzy genetic programming and innovations in this method concern, among the others, using weights of fuzzy aggregation operators, using weights of fuzzy rules, using fitness function criteria designed for fuzzy genetic programming and using dynamic links between fuzzy rules and fuzzy rules base. The proposed method was tested with use of typical nonlinear modelling problems.

Keywords Genetic programming · Weights · Fuzzy system · Nonlinear modelling · Dynamic systems

1 Introduction

The aim of nonlinear modelling is to obtain model with behaviour as close as possible to the testing object. Nonlinear modelling can concern many area of interest, such as physics, engineering, biology, chemistry, etc. and it is an important topic in the literature [16]. One of the mostly used systems for nonlinear modelling are fuzzy systems [20]. These systems can achieve high accuracy and interpretable knowledge in a form of fuzzy rules [9]. Most papers in the literature concerns selecting parameters of fuzzy system with specified structure of the system. For example, genetic algorithms [14], population-based algorithms [5], differential evolution [8], etc. are used to achieve that. From the other hand, an interesting group of methods for solving nonlinear modelling are genetic programming methods [21].

K. Łapa (✉) · K. Cpałka
Institute of Computational Intelligence, Częstochowa University of Technology, Częstochowa, Poland
e-mail: krystian.lapa@iisi.pcz.pl

K. Cpałka
e-mail: krzysztof.cpalka@iisi.pcz.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_13

These methods allow obtaining system structures in a form of computer programs (trees). These systems can also achieve high accuracy. However, disadvantage of typical genetic programming methods is lack in interpretability.

1.1 Fuzzy Systems

Fuzzy systems [20] are based on fuzzy logic and fuzzy rules. Each fuzzy rule consists of fuzzy sets which can represent linguistic values '*low*', '*medium*', '*high*', etc. Fuzzy rules take interpretable form {IF … THEN …}. The interpretability can result not only from the low number of fuzzy rules and fuzzy sets, but also from the semantic of appropriate selected parameters of fuzzy sets [9]. The semantic clarifies understanding of how the systems (models [9], classifiers [12], control systems [13]) work. It is worth to mention that in the nonlinear modelling problems the interpretability is an important issue [9]. It emerges from possibilities of understanding how the current object works and it allows us to model the typical states of the object.

1.2 Genetic Programming

Genetic programming (GP) is, in a general, a computational intelligence method for designing systems (structures) in a form of computer programs for solving optimization problems [21]. These systems can be used as controllers, models of objects, classifiers, etc. The main difference between genetic programming and other computational intelligence methods is a possibility of creation tree structures with use of mathematical operators. GP and other evolutionary algorithms (like evolutionary strategies [20], evolutionary programming [17], genetic algorithm [7], etc.) rely on a population of solutions. These methods are based on a natural evolution (using mechanisms like natural selection, inheritance, survival, etc.) which gives them an advantage over other methods used for optimization problems like analytic methods, gradient methods and random methods (see e.g. [20]).

1.3 Genetic Programming Trees

Typical genetic programming tree contains nodes and leafs (both of them are noted as tree elements in further part of this paper). Each node contains mathematical operator which decides how node works and usually has two child tree elements (see Fig. 1). Each leaf contains a real number value or connection to real number value from the system input (see Fig. 2). The whole tree structure is represented by the root (main node) and child tree elements (see Fig. 3). The output of the system

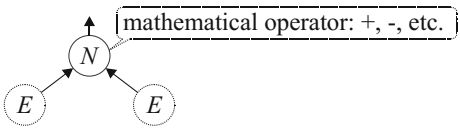

Fig. 1 GP node N and two connected tree elements E

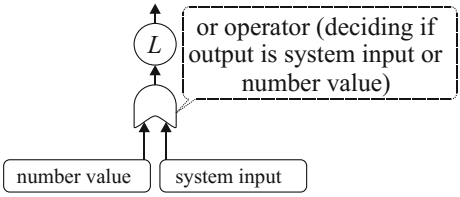

Fig. 2 GP leaf L connected with number value and system input

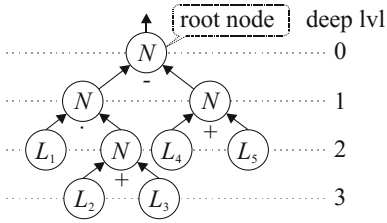

Fig. 3 An example of GP tree with randomly assigned structure and mathematical operators and output calculated as: $(L_1 \cdot (L_2 + L_3)) - (L_4 + L_5)$

(tree) is calculated recursively starting from the root node with use of defined mathematical operators. The mathematical operators are divided into: single argument operators (e.g. "cos(·)"), two argument operators (e.g. "+") and multi arguments operators (e.g. "avg(·)").

1.4 Fuzzy Genetic Programming

In the fuzzy genetic programming a mathematical operators are replaced by fuzzy operators (such as AND, OR, etc.) and leaves can be connected not to the system inputs but to the input of fuzzy sets. In [4] standard fuzzy operator AND was used, in [7] additional fuzzy operators: OR, NOT, '*greater*', '*lesser*' and '*near*' were used. In [14] operators AND, parent operators OR and fuzzy set operator NOT were used. In paper [18] a selected group of operators was used (with multiple versions of AND and OR operators).

1.5 Paper Aim

The aim of this paper is, among the others, to present impact of used weights in proposed weighted fuzzy genetic programing algorithm and to provide accurate and

interpretable fuzzy rules for considered simulation problems. The proposed method is based on fuzzy genetic programming. This method can be distinguished by: (a) use of flexible triangular norms with weights of arguments as fuzzy operators AND and OR, (b) use of weights of fuzzy rules, (c) use of operator NOT for fuzzy sets, (d) use of new encoding of the system, (e) use of fitness function with complexity of the system and new criteria of correct notations of fuzzy rules and (f) use of population algorithm adapted to the proposed system to select GP tree structure and parameters. The full description of the method and learning algorithm is presented in Sect. 2.

1.6 Paper Structure

The structure of the proposed paper consists of: Sect. 2 with description of the proposed method, Sect. 3 with presentation of simulation results and Sect. 4 with conclusions.

2 Proposed Method Description

2.1 Description of Fuzzy System

In this paper a Mamdani type fuzzy system [20] was used, where fuzzy rules can be defined as:

$$R_k : \begin{pmatrix} \text{IF}\left(\bar{x}_1 \text{ IS}[\text{NOT}]A_{1,k}\right) \text{ AND/OR} \ldots \text{AND/OR}\left(\bar{x}_n \text{ IS}[\text{NOT}]A_{1,k}\right) \\ \text{THEN}\left(y_1 \text{ IS } B_{1,k}\right), \ldots, \left(y_m \text{ IS } B_{m,k}\right) \end{pmatrix}, \quad (1)$$

where n is the number of inputs, m is the number of outputs, $\bar{\mathbf{x}} = [\bar{x}_1, \ldots, \bar{x}_n] \in \mathbf{X}$, $\mathbf{y} = [y_1, \ldots, y_m] \in \mathbf{Y}$, $A_{1,k}, \ldots, A_{n,k}$ are input fuzzy sets and $B_{1,k}, \ldots, B_{m,k}$ are output fuzzy sets. In the proposed method fuzzy rules (1) are represented by GP trees (see Fig. 4). Using genetic programming trees creates the need of using fuzzy sets' base, which allows fuzzy sets to connect to the leaves of trees. Proposed fuzzy sets' base $\mathbf{C}$ (stored for each input or output) is defined as:

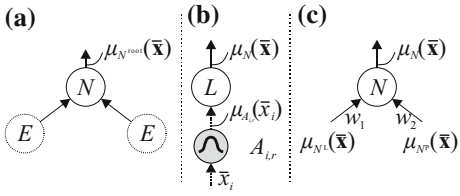

Fig. 4 Proposed structure of: **a** tree, **b** leaf, **c** node

Table 1 Parameters of proposed tree elements ('-' stands for parameters not used in the current type of tree element)

Tree element	Set of parameters							
E (element)	l	o	i	r	w_1	w_2	N^{L}	N^{P}
L (leaf)	$l = 1$	o	i	r	–	–	–	–
N (node)	$l = 0$	o	i	–	w_1	w_2	N^{L}	N^{P}

Table 2 Output of proposed tree elements (leaves and nodes)

Output value $(\mu_N(\bar{\mathbf{x}}))$	o	l	Operator name
$\mu_{A_{i,r}}(\bar{x}_i)$	0	1	IS
$1 - \mu_{A_{i,r}}(\bar{x}_i)$	1	1	IS NOT
$T^*\{\mu_{N^{\mathrm{L}}}(\bar{\mathbf{x}}), \mu_{N^{\mathrm{P}}}(\bar{\mathbf{x}}); w_1, w_2\}$	0	0	AND
$S^*\{\mu_{N^{\mathrm{L}}}(\bar{\mathbf{x}}), \mu_{N^{\mathrm{P}}}(\bar{\mathbf{x}}); w_1, w_2\}$	1	0	OR

$$\mathbf{C} = \left\{A_{1,1}, \ldots, A_{1,R}, \ldots, A_{n,1}, \ldots, A_{n,R}, B_{1,1}, \ldots, B_{1,R}, \ldots, B_{m,1}, \ldots, B_{m,R}\right\}. \quad (2)$$

Each input fuzzy set $A_{i,r}$ from fuzzy set base (2) is represented by the membership function $\mu_{A_{i,r}}(\bar{x})$, while each output fuzzy set $B_{j,r}$ is represented by the membership function $\mu_{B_{j,r}}(\bar{y})$. Each element of the tree (see Table 1) is described by a set of parameters: $l, o, i, r, w_1, w_2, N^{\mathrm{L}}$ and N^{P}. The parameter l decides how the element is treated (as node or leaf—see Table 1). The parameter o indicates operator of a given element: $o = 0$ for $l = 1$ stands for '*IS*', $o = 1$ for $l = 1$ stands for '*IS NOT*', $o = 0$ for $l = 0$ stands for '*AND*' and $o = 1$ for $l = 0$ stands for '*OR*' (see Table 2). The parameter i (for leaf) stands for input index of associated fuzzy set. The parameter r (for leaf) stands for index of associated fuzzy set $A_{i,r}$ (see Fig. 4b). The parameter w_1 stands for weight of left child node, w_2 stands for weight of right child node (see Fig. 4c), N^{L} stands for left child node and N^{P} stands for right child node. Taking into consideration mentioned parameters the output of any element of the tree can be calculated according to Table 2 (in these calculations a $T^*(\cdot)$ triangular t-norm with weights of arguments and $S^*(\cdot)$ triangular t-conorm with weights of arguments [20] were used).

The activation (firing) level of each fuzzy rule based on the structure presented in Fig. 4 is calculated as follows:

$$\tau_k(\bar{\mathbf{x}}) = \mu_{N_k^{\mathrm{root}}}(\bar{\mathbf{x}}), \quad (3)$$

where N_k^{root} stands for root of the tree of kth fuzzy rule ($k = 1, \ldots, K$, K stands for the number of fuzzy rules). Crisp output values of the system for each output j can be calculated (for example) with center of area method [20]:

$$\bar{y}_j(\bar{\mathbf{x}}) = \frac{\sum_{r=1}^{R} y_{j,r}^B \cdot \underset{k=1}{\overset{K}{S}} \left\{ T^* \left\{ \tau_k(\bar{\mathbf{x}}), \mu_{k,j}\left(y_{j,r}^B\right) \right\}; w_k^{\text{rule}} \right\}}{\sum_{r=1}^{R} \underset{k=1}{\overset{K}{S}} \left\{ T^* \left\{ \tau_k(\bar{\mathbf{x}}), \mu_{k,j}\left(y_{j,r}^B\right) \right\}; w_k^{\text{rule}} \right\}}, \tag{4}$$

where $y_{j,r}^B$ are centres of output fuzzy sets $B_{j,r}$, w^{rule} stands for fuzzy rule weight and $\mu_{k,j}\left(y_{j,r}^B\right)$ stands for membership function value of output fuzzy set $B_{j,k}$ calculated for discretization point $y_{j,r}^B$. This value can be calculated (using proposed encoding of the system) as:

$$\mu_{k,j}\left(y_{j,r}^B\right) = \mu_{B_{j,n_{j,k}^B}}\left(y_{j,r}^B\right), \tag{5}$$

where $n_{j,k}^B$ stands for index connecting k-th fuzzy rule with jth output fuzzy set (for example $n_{j=1,k=2}^B = 3$ means that the second fuzzy rule is associated with the third set of the first output $B_{1,3}$).

2.2 Encoding of the System

In the proposed approach encoding of the system (4) is based on the encoding of a tree elements $\mathbf{N}$ (see Table 1 and Fig. 4) as sets of parameters:

$$\mathbf{N} = \left\{ l, o, i, r, w_1, w_2, \mathbf{N}^{\text{L}}, \mathbf{N}^{\text{P}} \right\}, \tag{6}$$

where $\mathbf{N}^{\text{L}}$ and $\mathbf{N}^{\text{P}}$ encodes recursively child tree elements (these values are set to "no value" in case of leaves). The encoding of fuzzy system (4) is defined as:

$$\mathbf{X}_{ch} = \left\{ \mathbf{X}_{ch}^{\text{fsets}}, \mathbf{X}_{ch}^{\text{rules}} \right\}. \tag{7}$$

The part $\mathbf{X}_{ch}^{\text{fsets}}$ encodes parameters of fuzzy sets' base (represented by Gaussian membership functions) (2):

$$\mathbf{X}_{ch}^{\text{fsets}} = \left\{ \begin{array}{c} x_{1,1}^A, \sigma_{1,1}^A, \ldots, x_{1,R}^A, \sigma_{1,R}^A, \ldots, x_{n,1}^A, \sigma_{n,1}^A, \ldots, x_{n,R}^A, \sigma_{n,R}^A, \\ y_{1,1}^B, \sigma_{1,1}^B, \ldots, y_{1,R}^B, \sigma_{1,R}^B, \ldots, y_{m,1}^B, \sigma_{m,1}^B, \ldots, y_{m,R}^B, \sigma_{m,R}^B \end{array} \right\}, \tag{8}$$

thus the part $\mathbf{X}_{ch}^{\text{rules}}$ encodes fuzzy rules:

$$\mathbf{X}_{ch}^{\text{rules}} = \left\{ \mathbf{N}_1^{\text{root}}, n_{1,1}^B, \ldots, n_{m,1}^B, w_1^{\text{rule}}, \mathbf{N}_2^{\text{root}}, n_{1,2}^B, \ldots, n_{m,2}^B, w_2^{\text{rule}}, \ldots, \mathbf{N}_K^{\text{root}}, n_{1,K}^B, \ldots, n_{m,K}^B, w_K^{\text{rule}} \right\}, \tag{9}$$

where $\mathbf{N}_k^{\text{root}}$ is a root of the tree of k-th rule, $n_{j,k}^B$ is an index connecting k-th fuzzy rule with fuzzy set of j-th output, w_k^{rule} is weight of the rule. In the proposed method the part $\mathbf{X}_{ch}^{\text{fsets}}$ encoding parameters of fuzzy sets is processed by a genetic algorithm and the part $\mathbf{X}_{ch}^{\text{rules}}$ encoding fuzzy rules is processed by a genetic programming (for details see Sect. 2.5).

2.3 Initialization of the System

The parameters of fuzzy sets encoded in $\mathbf{X}_{ch}^{\text{fsets}}$ are initialized randomly with adjustments to the considered simulation problems. Next, the number of fuzzy rules $K \in \left[K^{\min}, K^{\max}\right]$ is chosen randomly. After the number of fuzzy rules is chosen, parameters of part $\mathbf{X}_{ch}^{\text{rules}}$ are initialized as follows:

$$\mathbf{X}_{ch}^{\text{rules}} = \left\{ \begin{array}{c} init\left(\mathbf{N}_1^{\text{root}}, 0, 0\right), U_c(1,R), \ldots, U_c(1,R), U_r(0,1), \\ init\left(\mathbf{N}_2^{\text{root}}, 0, 0\right), U_c(1,R), \ldots, U_c(1,R), U_r(0,1), \ldots, \\ init\left(\mathbf{N}_K^{\text{root}}, 0, 0\right), U_c(1,R), \ldots, U_c(1,R), U_r(0,1) \end{array} \right\}, \tag{10}$$

where $U_c(a,b)$ returns random integer value from the set $\{a, \ldots, b\}$, $U_r(a,b)$ returns random number value from the range $[a,b]$, $init(\mathbf{N}, d, e)$ (d stands for "deep lvl" of the tree—see Fig. 3, e stands for type of tree element) is an initialization function for tree elements. The function $init(\cdot)$ for nodes initialization (when $e = 0$ and $d < d^{\max}$, where $d^{\max}$ is maximum "deep lvl" of the tree) takes the following form:

$$init(\mathbf{N}, d, 0) = \left\{ \begin{array}{c} 0, U_c(0,1), U_c(1,n), U_c(1,R), U_r(0,1), U_r(0,1), \\ init\left(\mathbf{N}\{\mathbf{N}^{\text{L}}\}, d+1, U_c(0,1)\right), init\left(\mathbf{N}\{\mathbf{N}^{\text{P}}\}, d+1, U_c(0,1)\right) \end{array} \right\}, \tag{11}$$

where notation $\mathbf{N}\{a\}$ refers to gene a encoded in $\mathbf{N}$. It is worth to mention that this function initializes child elements of the tree recursively with increased value of deep of the tree (d). This process can stop if new element is a leaf or when deep of the tree reaches maximum lvl $d^{\max}$. The function $init(\cdot)$ for leaves initialization (when $e = 1$ or $d = d^{\max}$) takes the following form:

$$init(\mathbf{N}, d, 1) = \{1, U_c(0,1), U_c(1,n), U_c(1,R), U_r(0,1), U_r(0,1), null, null\}, \tag{12}$$

where *null* stands for genes with no assigned values.

2.4 System Evaluation

For evaluation of the system (4) the following fitness function was used:

$$\mathrm{ff}(\mathbf{X}_{ch}) = \underset{f=1}{\overset{F}{T^*}} \left\{ \mathrm{ffc}_f(\mathbf{X}_{ch}), w_f^{\mathrm{ffc}} \right\}, \tag{13}$$

where F stands for the number of fitness function components, component $\mathrm{ffc}_1(\mathbf{X}_{ch}) = \mathrm{ffacc}(\mathbf{X}_{ch})$ specifies the accuracy of the system (4), component $\mathrm{ffc}_2(\mathbf{X}_{ch}) = \mathrm{ffcom}(\mathbf{X}_{ch})$ specifies complexity of the system (4), component $\mathrm{ffc}_3(\mathbf{X}_{ch}) = \mathrm{ffsam}(\mathbf{X}_{ch})$ stands for a penalty for using the same fuzzy set multiple times by fuzzy rules (which is non-desired), component $\mathrm{ffc}_4(\mathbf{X}_{ch}) = \mathrm{ffmul}(\mathbf{X}_{ch})$ stands for a penalty for using the same input multiple times by single fuzzy rule (with is non-desired), w_f^{facc} $(f = 1, \ldots, F)$ are weights of components, $T^*\{\cdot\}$ is a n-argument extension of algebraic triangular norm with weights of arguments. The components aliases (ffcom, ffacc, ffsam, ffmul) were used to increase readability of the paper and presentation of the results.

The component $\mathrm{ffc}_1(\mathbf{X}_{ch})$ of function (13) is defined as:

$$\mathrm{ffc}_1(\mathbf{X}) = \frac{1}{m} \sum_{j=1}^{m} \frac{\frac{1}{Z} \sum_{z=1}^{Z} \left| d_{z,j} - \bar{y}_{z,j} \right|}{\max\limits_{z=1,\ldots,Z} \{d_{z,j}\} - \min\limits_{z=1,\ldots,Z} \{d_{z,j}\}}, \tag{14}$$

where Z is the number of rows of a learning sequence, $d_{z,j}$ is the desired output value of output j for input vector z $(z = 1, \ldots, Z)$, $\bar{y}_{z,j}$ is the real output value j calculated for the input vector $\bar{\mathbf{x}}_z$. Equation (14) takes into account the normalization of errors at different outputs of the system (4), which allows us to use function (14) in triangular norm used in function (13) (Table 3).

The component $\mathrm{ffc}_2(\mathbf{X}_{ch})$ of function (13) is defined as:

$$\mathrm{ffc}_2(\mathbf{X}_{ch}) = \frac{1}{K} \sum_{k=1}^{K} \frac{\mathrm{cm}\left(\mathbf{N}_k^{\mathrm{root}}\right)}{2^{lvl^{\max}} - 1}, \tag{15}$$

where $\mathbf{N}_k^{\mathrm{root}}$ is by default a part of structure $\mathbf{X}_{ch}^{\mathrm{rules}}$ of $\mathbf{X}_{ch}$ (this notation will be used in further part of this paper), denominator stands for maximum number of tree elements (Mersenne's number [15]), numerator stands for actual number of tree elements calculated according to the Table 4.

Table 3 Output of the $\mathrm{cm}(\cdot)$ function

$\mathrm{cm}(\mathbf{N})$	$\mathbf{N}\{l\}$
$1 + \mathrm{cm}\left(\mathbf{N}\{\mathbf{N}^{\mathrm{L}}\}\right) + \mathrm{cm}\left(\mathbf{N}\{\mathbf{N}^{\mathrm{P}}\}\right)$	0
1	1

Table 4 Output of the $\mathrm{sa}(\cdot)$ function

$\mathrm{sa}(\mathbf{N}, i, r)$	$\mathbf{N}\{l\}$	$\mathbf{N}\{i\}$	$\mathbf{N}\{r\}$
1	0	i	r
0	0	not i or not r	
$\mathrm{sa}(\mathbf{N}\{\mathbf{N}^{\mathrm{L}}\}, i, r) + \mathrm{sa}(\mathbf{N}\{\mathbf{N}^{\mathrm{P}}\}, i, r)$	1	–	–

The component $\mathrm{ffc}_3(\mathbf{X}_{ch})$ of function (13) is defined as:

$$\mathrm{ffc}_3(\mathbf{X}_{ch}) = \frac{\left(\begin{array}{l}\sum_{i=1}^{n}\sum_{r=1}^{R}\max\left(0, \sum_{k=1}^{K}\mathrm{sa}\left(\mathbf{N}_k^{\mathrm{root}}, i, r\right) - 1\right) + \\ + \sum_{j=1}^{m}\sum_{r=1}^{R}\max\left(0, \sum_{k=1}^{K}\mathrm{sb}\left(n_{j,k}^{B}, j, r\right) - 1\right)\end{array}\right)}{K \cdot \left(2^{d^{\max}-1} + m\right)} \tag{16}$$

where denominator stands for maximum number of leaves and output fuzzy sets for all K fuzzy rules, numerator stands for penalty for using specified fuzzy set more than 1 time by any fuzzy rule, function $\mathrm{sa}(\mathbf{N_k}, i, r)$ stands for number of used input fuzzy set $A_{i,r}$ by kth rule, function $\mathrm{sb}(n^B, j, r)$ stands for number of used output fuzzy set $B_{j,r}$ by k-th rule. The output of function $\mathrm{sa}(\mathbf{N}, i, r)$ from Eq. (16) is calculated according to Table 4 and the function $\mathrm{sb}(n^B, j, r)$ output is calculated as follows:

$$\mathrm{sb}\left(n^B, j, r\right) = \begin{cases} 1 & \text{for} \quad n^B = r \\ 0 & \text{for} \quad n^B \neq r \end{cases}. \tag{17}$$

The component $\mathrm{ffc}_4(\mathbf{X}_{ch})$ of function (13) was defined with assumption that one fuzzy rule cannot use multiple fuzzy sets which are connected to the same input:

$$\mathrm{ffc}_4(\mathbf{X}_{ch}) = \frac{1}{n \cdot K}\left(\sum_{i=1}^{n}\max\left(0, \sum_{k=1}^{K}\mathrm{mul}\left(\mathbf{N}_k^{\mathrm{root}}, i\right) - 1\right)\right), \tag{18}$$

where function $\mathrm{mul}(\mathbf{N}, i)$ stands for penalty for multiple use of fuzzy sets connected to i-th input calculated according to Table 5. The penalty resulting from using OR operator in minimization of fitness function (13) is smaller than penalty for using AND operator. Using the OR operator (see $\mathbf{N}\{o\} = 1$ in Table 5) for the same inputs is acceptable (opposed to AND operator), but it complicates readability of fuzzy rules.

Table 5 Output of the $\mathrm{mul}(\cdot)$ function

$\mathrm{mul}(\mathbf{N}, i)$	$\mathbf{N}\{l\}$	$\mathbf{N}\{i\}$	$\mathbf{N}\{o\}$
$\mathrm{mul}(\mathbf{N}\{\mathbf{N}^{\mathrm{L}}\}, i) + \mathrm{mul}(\mathbf{N}\{\mathbf{N}^{\mathrm{P}}\}, i)$	0	–	0
$\frac{1}{2}\left(\mathrm{mul}(\mathbf{N}\{\mathbf{N}^{\mathrm{L}}\}, i) + \mathrm{mul}(\mathbf{N}\{\mathbf{N}^{\mathrm{P}}\}, i)\right)$	0	–	1
0	1	not i	–
1	1	i	–

The aim of fitness function is to minimize values of all fitness function components which allow us to obtain accurate system (ffc_1) with simple structure (ffc_2) and consistent interpretable fuzzy rules (ffc_3 and ffc_4).

2.5 Description of Learning Algorithm

The learning algorithm purpose is to select parameters of the fuzzy sets stored in base (2) and to select the structure of the fuzzy rules. The taken into consideration algorithm designed for proposed system structure and encoding works according to the following steps:

- Step 1. In this step N^{pop} individuals of the population $\mathbf{P}$ are initialized according to description from Sect. 2.3.
- Step 2. This step involves evaluation of the individuals of the population $\mathbf{P}$ by fitness function (13).
- Step 3. In this step N^{pop} of child individuals are generated and stored in the temporary population $\mathbf{P}$. Genes $\mathbf{X}_{ch}^{\text{fsets}}$ of these individuals are initialized with use of genetic algorithm crossover operator. The individuals for crossover are selected by roulette wheel method from the population $\mathbf{P}$. The genes $\mathbf{X}_{ch}^{\text{rules}}$ of these individuals are initialized by choosing randomly genes $n^{B_{j,k}}$, weights w_k^{rule} and root nodes from preselected parents.
- Step 4. This step purpose is to mutate individuals from the population $\mathbf{P}$ (each individual is mutated with probability $p_{m1} \in (0,1)$). Genes $\mathbf{X}_{ch}^{\text{fsets}}$ are mutated (with probability $p_{m2} \in (0,1)$) with use of standard genetic mutation operator. Genes $\mathbf{X}_{ch}^{\text{rules}}$ are mutated (with probability $p_{m3} \in (0,1)$). This mutation is based on random changes of parameters $\mathbf{N}\{i\}$, $\mathbf{N}\{r\}$ and $n_{j,r}^{B}$. Independent mutation probabilities $p_{m1} \neq p_{m2} \neq p_{m3}$ (where $p_{m1} \gg p_{m2} > p_{m3}$) balance the mutation in the following way: (a) mutation should be processed on the greater part of the population $\mathbf{P}'$ (p_{m1}) which provides a proper diversity of the population, (b) from the other hand, genes mutation probability (p_{m2}) cannot be high due to degeneration of the population, (c) changes in connection between leafs and nodes (p_{m3}) should be rarely performed, because too intense changes in relationships between the fuzzy rules and fuzzy sets could hinder the convergence of the algorithm.
- Step 5. Next, the individuals from the population $\mathbf{P}'$ are pruned. This process is based on replacing randomly selected node of each genetic programming tree (with probability $p_x \in (0,1)$) by randomly generated leaf ($\text{init}(\mathbf{N}, 0, 1)$).
- Step 6. In this step extension of genetic programming trees from population $\mathbf{P}'$ is performed. This process is based on replacing randomly selected leaf of each genetic programming tree (with probability $p_l \in (0,1)$) by randomly generated node ($\text{init}(\mathbf{N}, lvl, 0)$). The lvl stands for actual height of leaf, which prevents excessive growth of the tree.

- Step 7. In this step for each individual from the population $\mathbf{P}'$ a new fuzzy rule is added (with probability p_d and only when $K < K^{\max}$) or existing randomly chosen fuzzy rule is removed (with probability p_u and when $K > K^{\min}$).
- Step 8. After modification of individuals from the population $\mathbf{P}'$ (Steps 3–7) each individual is evaluated by fitness function (13).
- Step 9. Next, the individuals from populations $\mathbf{P}$ and $\mathbf{P}'$ are merged and only N^{pop} best individuals are chosen to replace the population $\mathbf{P}$.
- Step 10. In the last step of the algorithm the purpose is to check if stop condition is met (for example if the number of executed iterations of algorithm reaches specified value). If this condition is met, the algorithm stops. Otherwise, algorithm goes back to the Step 3.

2.6 Fuzzy Rules Notation

As it was mentioned earlier, in the proposed system (4) a varied fuzzy operators were used to aggregate antecedences of fuzzy rules and to process the fuzzy sets. Due to this and using genetic programming tree structure, the notation of fuzzy rules is defined as:

$$R_k : \left(\text{IF} \overbrace{\text{zp}\left(\mathbf{X}_{ch}^{\text{rules}}\left\{\mathbf{N}_k^{\text{root}}\right\}\right)}^{\text{definitionbyfunction}} |\mathbf{X}_{ch}^{\text{rules}}\left\{w_k^{\text{rule}}\right\} \text{THEN} \begin{pmatrix} y_1 \text{IS} B_{1,\mathbf{X}_{ch}^{\text{rules}}\left\{n_{1,k}^B\right\}}, \ldots, \\ y_m \text{IS} B_{m,\mathbf{X}_{ch}^{\text{rules}}\left\{n_{m,k}^B\right\}} \end{pmatrix} \right), \quad (19)$$

where function $\text{zp}(\mathbf{N})$ defines antecedences of fuzzy rules according to the Table 6. It is worth to mention that values of weights w_1 and w_2 from Eq. (6) and w^{rule} from Eq. (9) can be replaced by their linguistic equivalents: n (not important) for values lower than 0.25, i (important) for values from the range $[0.25, 0.75]$ and v (very important) for values higher than 0.75. The fuzzy rule notation may be written as the following example:

$$R_1 : \left(\text{IF} \left(\begin{pmatrix} x_4 \text{ISNOT} A_{4,5} | v \\ \text{AND} \\ x_6 \text{IS} A_{6,2} | n \end{pmatrix} | n \text{AND} \begin{pmatrix} x_1 \text{IS} A_{1,4} | i \\ \text{OR} \\ x_2 \text{IS} A_{2,2} | n \end{pmatrix} | i \right) | i \text{THEN} \left(y_1 \text{IS} B_{1,4}\right). \right), \quad (20)$$

or in a shorter form as:

$$R_1 : \text{IF} \left(\begin{pmatrix} (x_4 \text{ISNOT} A_{4,5} | v \text{AND} x_6 \text{IS} A_{6,2} | n) | n \\ \text{AND}(x_1 \text{IS} A_{1,4} | i \text{OR} x_2 \text{IS} A_{2,2} | n) | i \end{pmatrix} | i \right) \text{THEN} \left(y_1 \text{IS} B_{1,4}\right). \quad (21)$$

Table 6 Notation output of the $\mathrm{zp}(\cdot)$ function

$\mathrm{zp}(\mathbf{N}, i)$	$\mathbf{N}\{l\}$	$\mathbf{N}\{o\}$
$(\mathrm{zp}(\mathbf{N}\{\mathbf{N}^{\mathrm{L}}\})\|\mathbf{N}\{w_1\}\mathrm{ANDzp}(\mathbf{N}\{\mathbf{N}^{\mathrm{P}}\})\|\mathbf{N}\{w_2\})$	0	0
$(\mathrm{zp}(\mathbf{N}\{\mathbf{N}^{\mathrm{L}}\})\|\mathbf{N}\{w_1\}\mathrm{ORzp}(\mathbf{N}\{\mathbf{N}^{\mathrm{P}}\})\|\mathbf{N}\{w_2\})$	0	1
$x_{\mathbf{N}\{i\}}\mathrm{ISA}_{\mathbf{N}\{i\},\mathbf{N}\{r\}}$	1	0
$x_{\mathbf{N}\{i\}}\mathrm{ISNOTA}_{\mathbf{N}\{i\},\mathbf{N}\{r\}}$	1	1

Table 7 Considered simulation problems

Problem	Label	Inputs	Outputs	Rows
Airfoil self-noise	ASN	5	1	1503
Box and Jenkins gas furnace	BJG	6	1	290
chemical plant problem	CPP	3	1	70
concrete slump test	CST	7	3	103
servo data set	SDS	4	1	167

3 Simulation Results

Simulation was performed using the following benchmarks (for details see Table 7): airfoil self-noise problem [3] (ASN), Box Jenkins gas furnace problem [2] (BJG), chemical plant problem [22] (CPP), concrete slump test [23] (CST), servo data set [19] (SDS). The simulations were executed for four cases:

- case 1—case without using weights (in this case all weights values w_1, w_2 and w^{rule} were set to 1).
- case 2—case with using rule weights (in this case weights values w_1, w_2 were set to 1).
- case 3—case with using fuzzy operators weights (in this case weights values w^{rule} were set to 1).
- case 4—case with using rule weights and fuzzy operators weights.

This way of testing allowed precise determination of the impact of using weights in the system (4) on the results.

3.1 Simulation Parameters

Values of parameters of the algorithm were experimentally selected as follows: number of fuzzy sets in fuzzy sets' base (2) for each input and output $R = 5$, minimum number of fuzzy rules $K^{\min} = 3$, maximum number of fuzzy rules $K^{\max} = 5$, maximum height of the tree $lvl^{\max} = 5$, weights of fitness function

components (13) $w_{\text{ffacc}} = 1.0$, $w_{\text{ffcom}} = 0.5$, $w_{\text{ffsam}} = 0.2$, $w_{\text{ffmul}} = 0.1$, number of individuals in population $N^{\text{pop}} = 100$, number of algorithm iterations $Nstep = 1000$, individual mutation probability $p_{m1} = 0.7$, genes mutation probability $p_{m2} = 0.2$, rules mutation probability $p_{m3} = 0.1$, pruning of tree probability $p_x = 0.3$, extending of tree probability $p_l = 0.2$, adding new fuzzy rule probability and removing fuzzy rule probability to $p_u = 0.3$. For each benchmark and case, simulations were repeat 100 times and results were averaged.

3.2 Obtained Results

Obtained results for all simulation problems are presented in Table 9. The normalized and averaged results for all simulation problems are presented in Fig. 5 and in Table 8. The example of obtained fuzzy rules and fuzzy sets are presented in Fig. 6 and in Table 10.

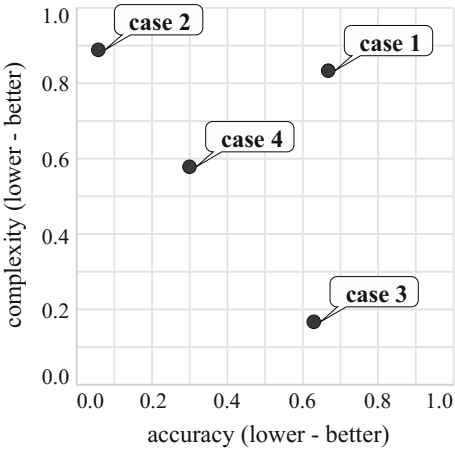

Fig. 5 Obtained accuracy and complexity normalized and averaged for all considered simulation problems

Table 8 Obtained accuracy and complexity, normalized and averaged for all considered simulation problems (see also Fig. 5)

Rule weights	Norm weights	ffacc	ffcom
No	No	0.6675	0.8334
Yes	No	0.0574	0.8888
No	Yes	0.6292	0.1671
Yes	Yes	0.2994	0.5785

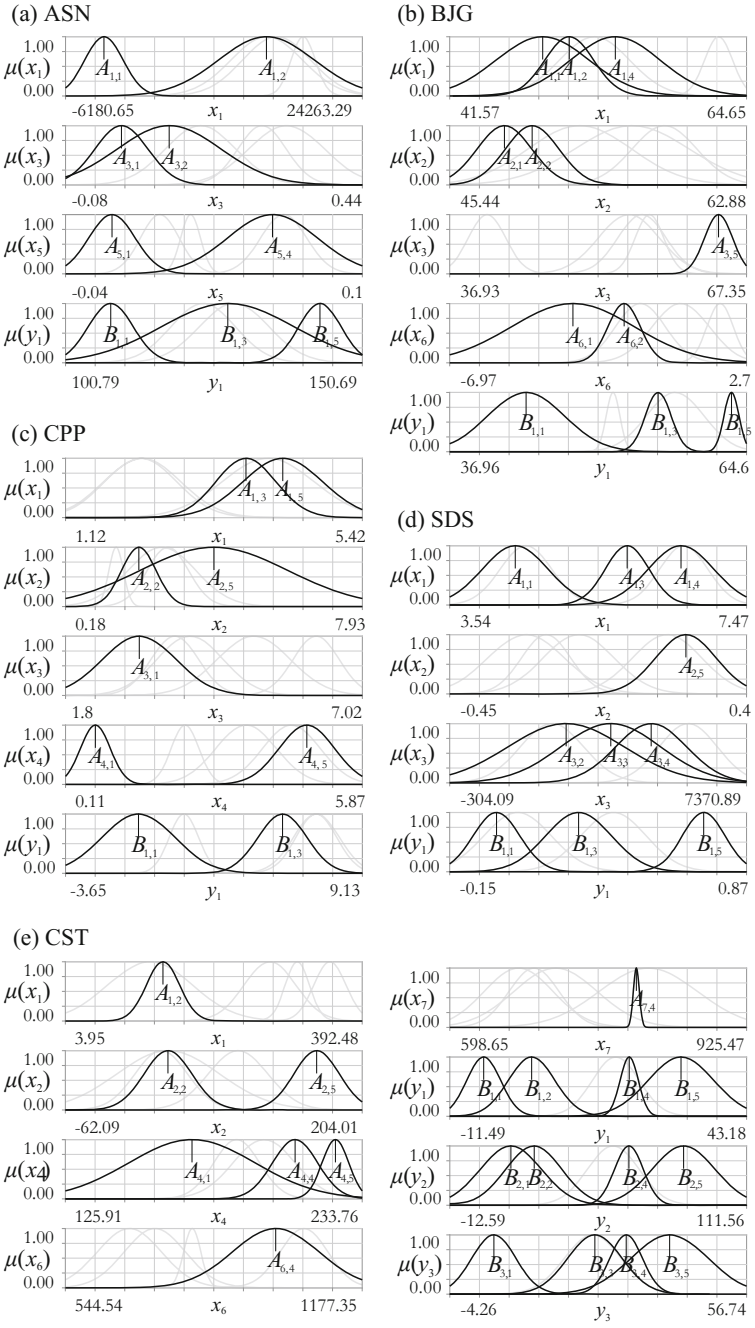

Fig. 6 Fuzzy sets obtained for simulation problems corresponding to fuzzy rules presented in Table 10. Grey fuzzy sets stands for fuzzy sets from fuzzy set base not used by any fuzzy rules

Table 9 Obtained simulation results in comparison with the best results (aimed on accuracy) obtained by other authors [1, 6, 10, 11]

Label	Rule weights	Norm weights	avg.ff	avg.K	avg. RMSE	RMSE	Best RMSE (other authors)
ASN	No	No	0.1601	3.7778	4.4576	4.1256	From 2.4280 to 5.270
	Yes	No	0.1585	4.0000	4.3578	3.9965	
	No	Yes	0.1529	3.5556	4.3882	3.8559	
	Yes	Yes	0.1566	3.6000	4.3459	4.1694	
BJG	No	No	0.0831	4.4500	0.4315	0.3720	From 0.2190 to 0.4490
	Yes	No	0.0854	4.7500	0.4123	0.3633	
	No	Yes	0.0746	3.7000	0.5124	0.4470	
	Yes	Yes	0.0778	4.0588	0.4894	0.4036	
CPP	No	No	0.0676	4.6000	0.0087	0.0065	From 0.0042 to 0.0092
	Yes	No	0.0719	4.5882	0.0082	0.0069	
	No	Yes	0.0631	4.0500	0.0085	0.0071	
	Yes	Yes	0.0770	4.7000	0.0081	0.0065	
CST	No	No	0.1193	3.1500	13.9072	12.8881	From 11.9410 to 15.3440
	Yes	No	0.1193	3.1579	13.7611	12.1798	
	No	Yes	0.1186	3.0588	13.7612	12.6257	
	Yes	Yes	0.1159	3.0588	13.7616	13.3177	
SDS	No	No	0.1016	3.8000	0.4188	0.2976	From 0.1177 to 0.7480
	Yes	No	0.0961	3.5500	0.4007	0.3280	
	No	Yes	0.1098	3.4167	0.5248	0.3858	
	No	No	0.1601	3.7778	4.4576	4.1256	

3.3 Simulation Conclusions

The simulation conclusions are following: (a) using rule weights allowed us to obtain better accuracy and similar complexity in a comparison to case without using weights (see Tables 8, 9 and case 2 on Fig. 5), (b) using fuzzy operators weights allowed us to obtain lower complexity and similar accuracy in a comparison to case without using weights (see Tables 8, 9 and case 3 on Fig. 5), (c) using fuzzy operators weights and fuzzy rules weights allowed us to obtain both the lower complexity and better accuracy in a comparison to case without using weights (see Table 8, 9 and case 4 in Fig. 5), (d) obtained results do not differ from the results of other authors (see Table 9). It is worth to mention that other authors' results are concentrated mostly on accuracy or on using more complex systems (see e.g. [1, 10]), (e) proposed approach is characterized by clear and interpretable fuzzy rules (see Fig. 6 and Table 10).

Table 10 Obtained examples of fuzzy rules for all simulation problems

Label	Fuzzy rules notation	RMSE
ASN	$\begin{cases} R_1 : \text{IF}\left(\begin{matrix} x_1 \text{IS} A_{1,1} \mid v \\ \text{AND}\, x_3 \text{ IS} A_{3,1} \mid n \end{matrix}\right) \mid i \text{ THEN } \left(y_1 \text{IS } B_{1,5}\right) \\ R_2 : \text{IF}\left(x_1 \text{ IS } A_{1,1} \mid v \text{ AND } x_5 \text{ IS } A_{5,1} \mid n\right) \mid v \text{ THEN } \left(y_1 \text{ IS } B_{1,3}\right) \\ R_3 : \text{IF}\left(\left(\begin{matrix} x_5 \text{ IS } A_{5,4} \mid v \\ \text{AND } x_3 \text{ IS NOT } A_{3,2} \mid n \end{matrix}\right) \mid v \text{ OR } x_1 \text{ IS } A_{1,2} \mid n\right) \mid v \text{ THEN } \left(y_1 \text{ IS } B_{1,1}\right) \end{cases}$	4.2719
BJG	$\begin{cases} R_1 : \text{IF}\left(x_6 \text{ IS } A_{6,1} \mid v \text{ AND} \left(\begin{matrix} x_3 \text{ IS } A_{3,5} \mid i \\ \text{OR } x_1 \text{ IS } A_{1,2} \mid n \end{matrix}\right) \mid v\right) \mid i \text{ THEN } \left(y_1 \text{ IS } B_{1,5}\right) \\ R_2 : \text{IF}\left(x_1 \text{ IS } A_{1,1} \mid v \text{ OR } x_2 \text{ IS } A_{2,1} \mid i\right) \mid v \text{ THEN } \left(y_1 \text{ IS } B_{1,1}\right) \\ R_3 : \text{IF}\left(x_1 \text{ IS } A_{1,4} \mid v \text{ OR} x_6 \text{ IS } A_{6,2} \mid i\right) \mid v \text{ THEN } \left(y_1 \text{ IS } B_{1,3}\right) \\ R_4 : \text{IF}\left(x_1 \text{ IS } A_{1,1} \mid v \text{ OR } x_2 \text{ IS } A_{2,2} \mid i\right) \mid v \text{ THEN} \left(y_1 \text{ IS } B_{1,1}\right) \end{cases}$	0.4725
CPP	$\begin{cases} R_1 : \text{IF}\left(x_3 \text{ IS } A_{3,4} \mid i \text{ AND } x_3 \text{ IS } A_{3,3} \mid i\right) \mid v \text{ THEN} \left(y_1 \text{ IS } B_{1,5}\right) \\ R_2 : \text{IF}\left(\left(x_3 \text{ IS } A_{3,2} \mid v \text{ OR } x_1 \text{IS } A_{1,4} \mid i\right) \mid v \text{ OR } x_1 \text{ IS } A_{1,3} \mid i\right) \mid i \text{ THEN} \left(y_1 \text{ IS } B_{1,1}\right) \\ R_3 : \text{IF}\left(\left(x_3 \text{ IS } A_{3,3} \mid i \text{OR } x_3 \text{ IS } A_{3,4} \mid i\right) \mid i \text{OR } x_2 \text{ IS } A_{2,5} \mid n\right) \mid i \text{ THEN} \left(y_1 \text{ IS } B_{1,3}\right) \\ R_4 : \text{IF}\left(x_1 \text{ IS } A_{1,1} \mid i \text{ AND } x_3 \text{ IS } A_{3,3} \mid i\right) \mid v \text{ THEN} \left(y_1 \text{ IS } B_{1,5}\right) \end{cases}$	0.0082
CST	$\begin{cases} R_1 : \text{IF}\left(x_7 \text{ IS } A_{7,4} \mid i \text{ OR } x_2 \text{ IS } A_{2,5} \mid v\right) \mid v \text{ THEN} \left(y_1 \text{ IS } B_{1,1}, y_2 \text{IS } B_{2,2}, y_3 \text{IS} B_{3,4}\right) \\ R_2 : \text{IF}\left(x_1 \text{IS} A_{1,2} \mid i \text{AND} x_4 \text{IS} A_{4,5} \mid v\right) \mid i \text{THEN} \left(y_1 \text{IS} B_{1,4}, y_2 \text{IS} B_{2,4}, y_3 \text{ IS } B_{3,1}\right) \\ R_3 : \text{IF}\left(x_6 \text{ IS } A_{6,4} \mid i \text{ OR } x_4 \text{ IS } A_{4,4} \mid n\right) \mid i \text{ THEN} \left(y_1 \text{ IS } B_{1,2}, y_2 \text{ IS } B_{2,1}, \ y_3 \text{ IS } B_{3,3}\right) \\ R_4 : \text{IF}\left(x_4 \text{ IS NOT } A_{4,1} \mid i \text{ AND } x_2 \text{ IS } A_{2,2} \mid i\right) \mid v \text{ THEN} \left(\begin{matrix} y_1 \text{ IS } B_{1,5}, \ y_2 \text{ IS} \\ B_{2,5}, \ y_3 \text{ IS } B_{3,5} \end{matrix}\right) \end{cases}$	13.4402
SDS	$\begin{cases} R_1 : \text{IF}\left(x_3 \text{ IS NOT } A_{3,1} \mid v \text{ AND } x_1 \text{ IS } A_{1,5} \mid n\right) \mid i \text{ THEN} \left(y_1 \text{ IS } B_{1,1}\right) \\ R_2 : \text{IF} \left(\begin{matrix} x_3 \text{ IS NOT } A_{3,1} \mid v \\ \text{AND} \left(\begin{matrix} x_1 \text{ IS } A_{1,3} \mid v \\ \text{OR} \left(\begin{matrix} x_4 \text{ IS } A_{4,5} \mid i \\ \text{AND } x_2 \text{ IS } A_{2,2} \mid n \end{matrix}\right) \mid v \end{matrix}\right) \mid n \end{matrix}\right) \mid i \text{ THEN} \left(y_1 \text{ IS } B_{1,1}\right) \\ R_3 : \text{IF}\left(x_2 \text{ IS NOT } A_{2,5} \mid i \text{ OR } x_4 \text{IS NOT } A_{4,1} \mid n\right) \mid n \text{ THEN} \left(y_1 \text{ IS } B_{1,3}\right) \end{cases}$	0.4251

The corresponding fuzzy sets are shown in Fig. 6

4 Conclusions

In this paper a weighted fuzzy genetic programming algorithm for selection of the structure and the parameters of the fuzzy systems for nonlinear modelling is presented. In presented approach fuzzy rules take the form of binary trees where nodes of these trees decide on aggregation operators (AND/OR) and the leaves of these trees are connected to the input fuzzy sets. The proposed method allows us to obtain accurate fuzzy systems with clear and interpretable fuzzy rules. The obtained accuracy is similar to the accuracy obtained by other authors, achieved using systems which usually do not take into account interpretability. The use of the system weights shown possibilities in increasing the system accuracy (with use of rule weights), decrease the system complexity (with use of fuzzy operators'

weights) or improve both accuracy and complexity (with use of both weights). The proposed approach was tested on typical nonlinear modelling benchmarks and it can be said that obtained results are satisfying.

Acknowledgment The project was financed by the National Science Centre (Poland) on the basis of the decision number DEC-2012/05/B/ST7/02138.

References

1. Bosnic, Z., Kononenko, I.: Correction of regression predictions using the secondary learning on the sensitivity analysis outputs. Comput. Inform. **20**, 1–17 (2001)
2. Box, G., Jenkins, G.: Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco (1970)
3. Brooks, T.F., Pope, D.S., Marcolini, A.M.: Airfoil self-noise and prediction. Technical report, NASA RP-1218 (1989)
4. Carmona, C.J., Ruiz-Rodado, V., del Jesus, M.J., Weber, A., Grootveld, M., González, P., Elizondo, D.: A fuzzy genetic programming-based algorithm for subgroup discovery and the application to one problem of pathogenesis of acute sore throat conditions in humans. Inf. Sci. **298**, 180–197 (2015)
5. Cheng, S., Shi, Y., Qin, Q., Zhang, Q., Bai, R.: Population diversity maintenance in brain storm optimization algorithm. J. Artif. Intell. Soft Comput. Res. **4**(2), 83–97 (2014)
6. Cpałka, K., Łapa, K., Przybył, A., Zalasiński, M.: A new method for designing neuro-fuzzy systems for nonlinear modelling with interpretability aspects. Neurocomputing **135**, 203–217 (2014)
7. Edmonds, A.N., Kershaw, P.S.: Genetic programming of Fuzzy logic production rules with application to financial trading. In: Proceedings of the IEEE World Conference on Computational Intelligence, Orlando, Florida (1994)
8. Gabryel, M., Woźniak, M., Damaševičius, R.: An application of differential evolution to positioning queueing systems. Lect. Notes Comput. Sci. **9120**, 379–390 (2015)
9. Gacto, M.J., Alcalá, R., Herrera, F.: Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures. Inf. Sci. **181**(20), 4340–4360 (2011)
10. Huang, G.-B., Zhu, Q.-Y., Siew, C.-K.: Extreme learning machine: theory and applications. Neurocomputing **70**, 489–501 (2006)
11. Łapa, K.: Algorithms for extracting interpretable expert knowledge in nonlinear modeling issues. Ph.D. thesis (in polish), Czestochowa University of Technology (2015)
12. Łapa, K., Cpałka, K., Galushkin, A.I.: A new interpretability criteria for neuro-fuzzy systems for nonlinear classification. Artif. Intell. Soft Comput. Lect. Notes Comput. Sci. **9119**, 448–468 (2015)
13. Łapa, K., Cpałka, K.: On the application of a hybrid genetic-firework algorithm for controllers structure and parameters selection. Adv. Intell. Syst. Comput. **429**, 111–123 (2015)
14. Mendes, R.R.F., Voznika, F.B., Freitas, A.A., Nievola, J.C.: Discovering fuzzy classification rules with genetic programming and co-evolution. In: De Raedt, L., Siebes, A. (eds.) PKDD 2001, LNAI 2168, pp. 314–325 (2001)
15. Robinson, M.R.: Mersenne and Fermat numbers. Proc. Am. Math. Soc. **5**, 842–846 (1954)
16. Motulsky, H.J., Christopoulos, A.: Fitting models to biological data using linear and nonlinear regression. A practical guide to curve fitting. GraphPad Software Inc., San Diego, CA (2003)
17. Nallasamy, K., Ratnavelu, K.: Optimal control for stochastic linear quadratic singular Takagi-Sugeno fuzzy delay system using genetic programming. Appl. Soft Comput. **12**, 2085–2090 (2012)

18. Preen, R.J., Bull, L.: Fuzzy dynamical genetic programming in XCSF. In: GECCO'11, July 12–16, 2011, pp. 167–168
19. Quinlan, J.R.: Learning with continuous classes. In: Adams, A., Sterling, L. (eds.) Proceedings 5th Australian Joint Conference on AI, World Scientific, Singapore (1992)
20. Rutkowski, L.: Computational Intelligence. Springer (2008)
21. Stanimirovic, Z., Maric, M., Bozovic, S., Stanojevic, P.: An efficient evolutionary algorithm for locating long-term care facilities. Inf. Technol. Control **41**(1), 77–89 (2012)
22. Sugeno, M., Yasukawa, T.: A fuzzy-logic based approach to qualitative modelling. IEEE Trans. Fuzzy Syst. **1**, 7–31 (1993)
23. Yeh, I.C.: Modeling slump flow of concrete using second–order regressions and artificial neural networks. Cement Concr. Compos. **29**(6), 474–480 (2007)

Queuing Model of Wireless Access System

Sławomir Hanczewski, Maciej Stasiak and Piotr Zwierzykowski

Abstract The paper presents a new multi-dimensional Erlang's Ideal Grading (EIG) model with queues and priority that can service a number of call classes with differentiated access to resources. The model was used to determine delays and packet loss probabilities in the wireless access system. The analytical results obtained in the study were then compared with the results of a simulation, which confirmed the essential and required accuracy of the proposed model. The model developed in the study can be used to analyse, design and optimize present-day wireless access system.

Keywords EIG · Queuing systems · Priority

1 Introduction

For many years, wireless technologies have been one of the most dynamically developing branches of the IT [1] and wireless access systems are an important part of the present-day operator networks [2]. The scope, quality and responsibilities supporting the cooperation between network operators and users are described in SLA (Service Level Agreement) [3]. These agreements also include technical parameters that define the quality of service for network services (QoS—Quality of Service). A good example of these parameters can be the maximum delay or the acceptable level of packet loss. To guarantee appropriate values for the QoS parameters, it is necessary to take them into consideration at the network designing

S. Hanczewski (✉) · M. Stasiak · P. Zwierzykowski
Faculty of Electronics and Telecommunications, Poznan University of Technology, Poznań, Poland
e-mail: slawomir.hanczewski@put.poznan.pl

M. Stasiak
e-mail: maciej.stasiak@put.poznan.pl

P. Zwierzykowski
e-mail: piotr.zwierzykowski@put.poznan.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_14

stage and to monitor their values when the operation of the network is in progress. The present paper proposes a simplified analytical model of the wireless access part of a service provider's network, further in the text referred to as the wireless access system.

In the study, a model of Erlang's Ideal Grading (EIG) is used to analyse the wireless access system [4–6]. Ideal grading with one call stream was proposed by Erlang [7–9]. In the works [10, 11], a model of EIG with one call stream and infinite queues in each of the load groups was proposed. Gradings with different structures, with finite and infinite queues and different types of call streams, are addressed in [12–15], among others. [16] proposes a model of EIG without queues in which the system can service a number of call classes. In [4], a model of EIG in which call classes can have differentiated, non-integer availability to the resources of the group, was presented. In works [5, 17–19], the model is used to analyse UMTS system, mobile overflow system and VoD system. In [20], an approximated model of EIG with a number of call stream and finite queues for each load group was introduced. Systems with priorities were discussed in many previous analyzes, e.g., in [21], scheme for singe-service systems with losses was proposed, which in [22] was extended to multi-service systems with losses.

This paper proposes a queueing model of the EIG that can service a number of call classes with differentiated access to resources. Additionally, it is assumed in the model that one class has priority.

The paper is structured as follows. Following the Introduction, Sect. 2 describes the multi-dimensional EIG model with losses and with queues and Sect. 3 proposes the new model of EIG with queues and priority. In Sect. 4, the analysed wireless access model is parametrized in line with the EIG model with queues and priority proposed earlier. In Sect. 5, the results of the analytical calculations are compared with the results of a digital simulation for selected structures of the wireless access system. Section 6 sums up the paper.

2 Model of Erlang's Ideal Grading with Queues

The basis for the analysis of the wireless access system is the Erlang's Ideal Grading model that is offered a traffic stream composed of a number of call classes [23]. The group has the capacity of V allocation units (AU), i.e., units that have been adopted in the system capacity discretization process [24]. Traffic of class i is offered to gi load groups. A call of class i appearing at the input of a single load group has access to di AU's from the number of all V AU's. The di parameter is called the availability of class i. The number of load groups gi for calls of class i in EIG is equal to:

$$g_i = \binom{V}{d_i}. \tag{1}$$

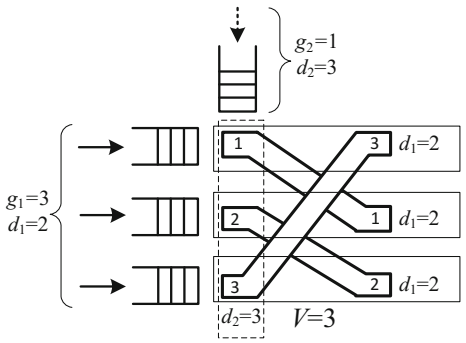

Fig. 1 Erlang's Ideal Grading with queues

Figure 1 shows a simple model of EIG with the capacity $V = 3$ AU's. The group services two classes of calls with the availability $d_1 = 2$, $d_2 = 3$. Hence, the number of incoming groups for relevant call classes is equal to $g_1 = 3$ and $g_2 = 1$.

EIG has properties that enable the combinatorial determination of the conditional blocking probability $\beta_i(n)$ for calls of class i in EIG servicing n AU's. The $\beta_i(n)$ parameter is equal to the probability of the appearance of a new call of class i in these load groups in which all AU's necessary for setting up class i connection are busy:

$$\beta_i(n) = \begin{cases} \binom{n}{d_i} \Big/ \binom{V}{d_i} & \text{for } d_i \le n \le V \\ 0 & \text{for } 0 \le n \le d_i \end{cases}. \tag{2}$$

A complementary event to the blocking phenomenon is the event of service acceptance for a call of class i in the occupancy state n. The probability of such an event is defined as the conditional transition probability $\sigma_i(n)$, and is equal to:

$$\sigma_i(n) = 1 - \beta_i(n). \tag{3}$$

Observe that the conditional transition probability $\sigma_i(n)$ and the conditional blocking probability $\beta_i(n)$ for calls of class i depend on the total number n of serviced calls of all classes and do not depend on the distribution of these calls between individual classes.

The occupancy distribution in a model of EIG with losses which services M call streams can be expressed by the formula [23]:

$$\begin{cases} n[P_n]_V = [P_{n-1}]_V \sum_{i=1}^{M} A_i \sigma_i(n-1) & \text{for } 0 \le n \le V \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

where:

- $[P_n]_V$—probability of n AU's being busy in the system,
- A_i—offered traffic of class i:

$$A_i = \frac{\lambda_i}{\mu_i}, \tag{5}$$

- λ_i—intensity of Poisson call stream of class i,
- μ_i—intensity of exponential holding time of class i.

Blocking probability for class i call in such a group can be calculated as follows:

$$B_n = \sum_{n=0}^{V} \beta_i(n)[P_n]_V. \tag{6}$$

Let us consider a model of EIG with a number of call streams and finite queues for each load group [20]. A class i call which cannot be serviced due to the occupancy of all AU's available in a given load group is directed to class $\underline{i}$ queue, respectively (Fig. 1). We assume that after the termination of servicing one call of class i, a new call of class i is selected from one of class i queues of an arbitrary load group which has access to released resources. This assumption leads to the conclusion that the service process taking place in EIG with delay is independent of the queue service discipline for particular classes. The occupancy distribution in such a system can be approximated by the following formulae [20]:

$$n\left[P_{n,z}\right]_{V,U} = \left[P_{n-1,z}\right]_{V,U} \sum_{i=1}^{M} A_i \sigma_i(n-1) \quad \text{for } 0 \le n \le V, z = 0, \tag{7}$$

$$n\left[P_{n,z}\right]_{V,U} = \left[P_{n,z-1}\right]_{V,U} \sum_{i=1}^{M} A_i \beta_i(n) + \left[P_{n-1,z}\right]_{V,U} \sum_{i=1}^{M} A_i \sigma_i(n-1) \quad \text{for } 0 \le n \le V,\ 1 \le z \le U \tag{8}$$

along with the normalisation condition:

$$\sum_{n=0}^{V} \sum_{z=0}^{U} \left[P_{n,z}\right]_{V,U} = 1. \tag{9}$$

where $\left[P_{n,z}\right]_{V,U}$ is the probability that the system services n calls and z calls are waiting in all queues. Parameter U defines the capacity of all EIG queues. The blocking probability for a call of class i is determined by the following equation:

$$B_n = \sum_{n=d_i}^{V+U} \beta_i(n) \left[P_{n,z}\right]_{V,U}. \tag{10}$$

The average length of the queue $q_i(n, z)$ for calls of class i in the microstate (n, z) can be expressed approximately by the following reasoning [20]: a transition from any randomly chosen macrostate (n, k) to (n, k + 1) causes an increase in length of a queue of class i by one, with the probability $\pi_i(n, k)$ that is equal to the participation of a stream of class i in the total stream that increases the total queue size by one waiting call:

$$\pi_i(n,k) = \frac{A_i \beta_i(n)}{\sum_{j=1}^{M} A_j \beta_j(n)}. \tag{11}$$

Therefore, the total average length of queues of class i in the macrostate (n, z) can be determined by the following formula:

$$q_i(n,z) = \sum_{k=0}^{z-1} 1 \cdot \pi_i(n,k) = z \frac{A_i \beta_i(n)}{\sum_{j=1}^{M} A_j \beta_j(n)}. \tag{12}$$

Now, we are in a position to determine the total average length of queues of class i in the system:

$$Q_i = \sum_{n=d_i}^{V} \sum_{z=0}^{U} q_i(n,z) \left[P_{n,z}\right]_{V,U}. \tag{13}$$

Since there are g_i queues of class i in the system (the number of queues is equal to the number of load groups of a given call class), then the average length of one queue q_i of class i and the average waiting time T_i in one queue of class i (determined on the basis of Little's formula [25]) are respectively as follows:

$$q_i = \frac{Q_i}{g_i}, \quad T_i = \frac{Q_i}{\lambda_i g_i}. \tag{14}$$

3 Model of EIG with Queues and Priority

Let us assume that one of the classes offered to the system has priority and calls of this class are not placed in a queue. Because the system works like a queuing system without interrupts, therefore, for class z calls with priority, short buffers are also provided with the capacity of one or two packets, which can be omitted in the modeling process. Each packet with a lower priority, in the absence of free resources, will be replaced by higher priority packet immediately after the

termination of the processing. Assume also that EIG services M classes of calls and that the calls of the first class have a higher priority than the others. The service of a call of the lower priority class does not have any influence on the servicing process of the first (higher priority) class.

First consider an EIG group servicing only one call of calls with priority. The blocking probability in such a system can be written as follows:

$$B_1 = f(A_1, d_1, V). \tag{15}$$

where function $f(\cdot)$ can be obtained on the basis of the EIG model with losses—Eqs. (4)–(6) for $M = 1$.

After determining the blocking probability, it is possible to determine the total traffic carried, i.e., the carried traffic of the first class calls in the considered system:

$$Y_1 = A_1[1 - B_1]. \tag{16}$$

Let us now construct a fictitious EIG model with queues for the other classes of calls. In the model, we assume that the EIG capacity will be reduced by the average number of AU's occupied by the first class calls (with priority), equal to the carried traffic of the first class. Similarly, the availability for each class of calls will be reduced in proportion to capacity changes, i.e.:

$$V^* = V - Y_1, \quad d_i^* = d_i\left(1 - \frac{Y_1}{V}\right), \quad \text{for } 2 \le i \le M. \tag{17}$$

All important characteristics of the system for call classes without priority can be obtained on the basis of the following formula:

$$(\mathbf{B}, \mathbf{q}) = f(\mathbf{A}, \mathbf{d}^*, V), \tag{18}$$

where function $f(\cdot)$ is determined based on the EIG model with queues—Eqs. (7)–(14).

4 Model of Access to Resources

The access to the resources of a system with the capacity of V AU's is available for G users (users correspond to load groups in EIG), whereas user j has access to a link with capacities V_j AU's, where $1 \le j \le G$. Each link can be offered a mixture of M packet streams. If a packet of class i cannot be serviced due to the occupancy of the access link or the occupancy of resources, then it is redirected to a queue of class i of a given user. Calls of the first class have a higher priority than other classes. This means that after the termination of servicing a given call in a given load group, a packet of the first class will be accepted for service, if only a queue of

this class in the given load group is not empty. The availability of the system for the first class $d_{1,j}$ for packets via the access link j is directly proportional to the traffic participation of the first class in the j link in the total traffic of class 1 offered to the access system:

$$d_{1,j} = \frac{V_j A_{1,j}}{\sum_{k-1}^{G} A_{1,k}}. \tag{19}$$

It was assumed that the availability in the system for calls of the first class (with priorities) will be equal to the average availability of particular access links:

$$d_1 = \frac{1}{G} \sum_{j=1}^{G} d_{1,j}. \tag{20}$$

The availability of the system for a class i call without priority $(2 \leq i \leq M)$ is a measure of access to resources and can be estimated on the basis of the reasoning proposed in [17]. In the access link j, the resources are shared between $M - 1$ packet streams. It can be assumed that the number of occupied resources in the link by particular classes will be directly proportional to the traffic participation of these classes in the j link in the total traffic offered to the access system.

Therefore, availability $d_{i,j}$ for class i packets by access link j can be expressed as follows:

$$d_{i,j} = \frac{V_j A_i}{\sum_{k=1}^{M} A_k}. \tag{21}$$

Taking into account (17) formula (22) can be rewritten as:

$$d_{i,j}^{*} = \left(1 - \frac{Y_1}{V}\right) d_{i,j}. \tag{22}$$

The total average availability for packets of class i is equal to:

$$d_i^{*} = \frac{1}{G} \sum_{j=1}^{G} d_{i,j}, \tag{23}$$

where G_i is the number of access links that service packets of class i $(1 \leq G_i \leq G)$.

5 Numerical Evaluation

In order to verify the proper functioning of the analytical model of a wireless access system proposed in the paper, the results obtained on its basis were compared with the results of simulation experiments. To do so, a wireless access system simulator

was developed. The operator network node capacity was assumed to be $C = 300$ Mbps. 10 clients ($G = 10$) were connected with the wireless node through wireless access links.

The clients were offered $M = 3$ classes: the first class (with priority) without queue and the second and the third classes with queues. The simulator's input parameters were as follows:

- λ_i—intensity of packets of class i,
- μ_i—intensity of service stream of class i,
- l_i—average length of packets of class i.

It was assumed that subscribers generated data with the speed of 50 Mbps.

In order to determine the bit length of each packet belonging to a stream of class i, a pseudo-random number following an exponential distribution with parameter μ_i was generated, corresponding to packet duration τ_i. Next, the packet bit length was calculated:

$$l_{\tau_i} = \tau_i c_i, \tag{24}$$

where c_i is the average bitrate of the packet stream:

$$c_i = \lambda_i l_i. \tag{25}$$

The access system capacity was expressed by the number of resource allocation units according to the following formula [24]:

$$V = \frac{C}{c_{\mathrm{AU}}}, \tag{26}$$

where c_{AU} is the system resource allocation unit. For the purpose of the experiment, it was assumed that an allocation unit was equal to $c_{\mathrm{AU}} = 10$ kbps. The value was determined on the basis of the maximum packet possible to be sent via the Ethernet (1526 bytes). The adopted total buffer capacity for all services was 1200 AU's.

The experiment also assumed that the traffic offered in each access network fulfilled the following condition: $A_1 : A_2 : A_3 = 1 : 5 : 8$.

The obtained results are presented in graphs in the function of traffic offered per one AU of the system:

$$a = \frac{\sum_{i=1}^{M} A_i}{V}. \tag{27}$$

The results obtained on the basis of the analytical model are shown by a solid line, whereas those obtained in the simulation are indicated by appropriate symbols.

The study made it possible to evaluate the average value of packet delays and the levels of packet losses for individual traffic classes. Figure 2 shows the average waiting time for packets in one queue, whereas Fig. 3 presents the values of packet

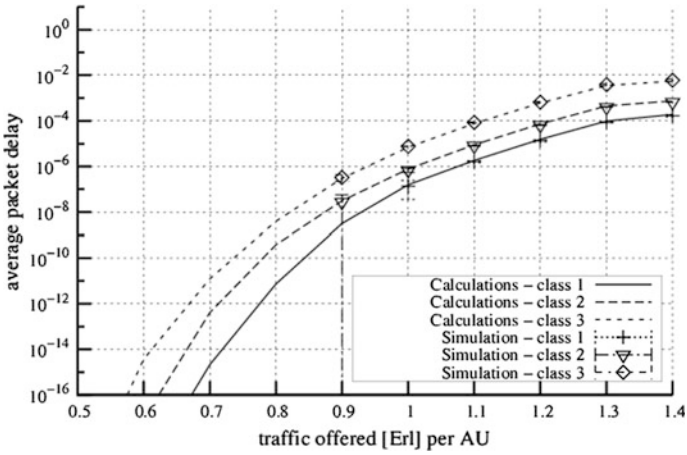

Fig. 2 Average packet delay of the second and the third classes

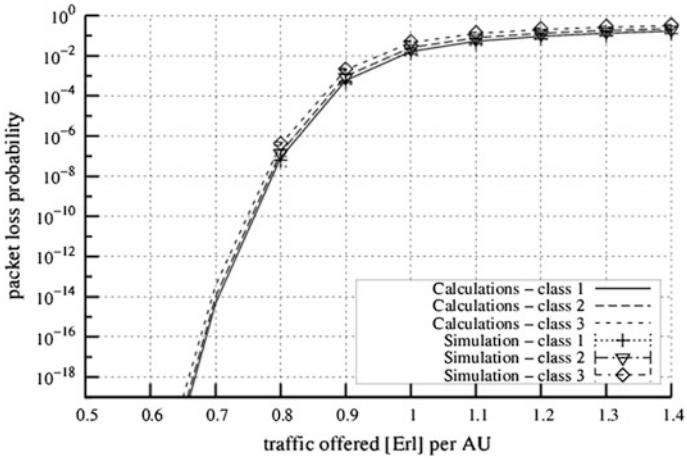

Fig. 3 Packet loss probability of individual traffic classes

loss probability. The study was conducted with reference to changing traffic intensities.

The values of loss probabilities presented in Fig. 3 also show the dependence between the loss probabilities obtained for the priority and non-priority classes. The first class has priority, which causes a lower value of loss probability.

To determine one point in the graph, 5 series of simulation runs, each involving 10,000,000 packets, were performed. This made it possible for confidence intervals to be determined at the level of 95 %. These intervals are small enough that, in the graphs, they do not exceed the values of the symbol that defines the result of a simulation.

6 Conclusion

This paper proposes a new analytical model of Erlang's Ideal Grading with queues for traffic classes that are differentiated by their availability to group resources. In the model it was assumed that one realtime class has priority. The proposed model is used to analyse a wireless access system in which QoS mechanisms are used. The study confirms that satisfactory accuracy can be obtained by the application of the analytical model presented in the paper. This means that the developed model can be used in evaluate QoS parameters in designing present-day wireless access systems.

The queueing model proposed in the paper will be further addressed in research projects in the future and extended to include selected packet scheduling mechanisms [26] and buffer management mechanisms [27, 28].

References

1. Yin, Z., Yu, F., BU, S., Han, Z.: Joint cloud and wireless networks operations in mobile cloud computing environments with telecom operator cloud. IEEE Trans. Wireless Commun. **99**, 1–1 (2015)
2. Martignon, F., Paris, S., Filippini, I., Chen, L., Capone, A.: Efficient and truthful bandwidth allocation in wireless mesh community networks. IEEE/ACM Trans. Netw. **23**(1), 161–174 (2015)
3. Wikisource: Telecommunications act of 1996. http://en.wikisource.org (2009) (Online)
4. Stasiak, M., Hanczewski, S.: Approximation for multi-service systems with reservation by systems with limited-availability. In: Computer Performance Engineering, LNCS, vol. 5261, pp. 257–267. Springer (2008)
5. Stasiak, M., Głąbowski, M., Hanczewski, S.: The application of the Erlang's Ideal Grading for modelling of UMTS cells. In: 8th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP), July 2012, pp. 1–6
6. Hanczewski, S., Stasiak, M., Zwierzykowski, P.: Modelling of the access part of a multi-service mobile network with service priorities. EURASIP J. Wireless Commun. Netw. **2015**(1), 1–14 (2015)
7. Brockmeyer, E.: A survey of A.K. Erlang's mathematical works. Danish Acad. Tech. Sci. **2**, 101–126 (1948)
8. Lotze, A.: History and development of grading theory. In: Proceedings of 5th International Teletraffic Congress, pp. 148–161 (1967)
9. Brockmeyer, E., Halstrom, H., Jensen, A.: The life and works of A.K. Erlang. Acta Polytechnika Scandinavia **6**, 287 (1960)
10. Thierer, M.: Delay System with Limited Accessibility. In Prebook of the 5th International Teletraffic Congress, 1967, pp. 203–213
11. Thierer, M.: Delay system with limited availability and constant holding time. In: Prebook of the 6th International Teletraffic Congress, Munich, pp. 322/1–322/6 (1970)
12. Gambe, E.: A study on the efficiency of graded multiple delay systems through artificial traffic trials. In: Proceedings of 3rd International Teletraffic Congress, Paris, doc. 16 (1961)
13. Kühn, P. (1970) Combined delay and loss systems with several input queues, full and limited accessibility. In: Proceedings of 6th International Teletraffic Congress, Munich, pp. 323/1–7 (1970)

14. Kühn, P.: Waiting time distributions in multi-queue delay systems with gradings. In: Proceedings of 7th International Teletraffic Congress, Stockholm, pp. 53–61 (1973)
15. Kampe, G., Kühn, P.: Graded systems with infinite or finite source traffic and exponential or constant holding time. In: Proceedings of 8th International Teletraffic Congress, Melbourne, pp. 251/1–10 (1976)
16. Stasiak, M.: An approximate model of a switching network carrying mixture of different multichannel traffic streams. IEEE Trans. Commun. **41**(6), 836–840 (1993)
17. Hanczewski, S., Stasiak, M.: Performance modelling of Video-on-Demand systems. In: 17th Asia-Pacific Conference on Communications (APCC), pp. 784–788 (2011)
18. Głąbowski, M., Hanczewski, S., Stasiak, M.: Erlang's Ideal Grading in Diffserv modelling. In: AFRICON, 2011, pp. 1–6 (2011)
19. Głąbowski, M., Hanczewski, S., Stasiak, M.: Modelling of Cellular Networks with Traffic Overflow. Mathematical Problems in Engineering, Article ID 286490 (2015)
20. Hanczewski, S., Stasiak, M., Weissenberg, J., Zwierzykowski, P.: Queuing model of the access system in the packet network. In: Computer Networks, ser. Communications in Computer and Information Science. Springer, vol. accepted for publication (2016)
21. Katzschner, L., Scheller, R.: Probability of loss of data traffics with different bit rates hunting one common PCM-channel, pp. 525/1–8 (1976)
22. Stasiak, M., Wiewióra, J., Zwierzykowski, P., Parniewicz, D.: An approximate model of the WCDMA interface servicing a mixture of multi-rate traffic streams with priorities. In: Thomas, N., Juiz, C. (eds.) 5th European Performance Engineering Workshop, vol. 5261, pp. 168–180 (2008)
23. Głąbowski, M., Hanczewski, S., Stasiak, M., Weissenberg, J.: Modeling Erlang's Ideal Grading with Multirate BPP traffic. Mathematical Problems in Engineering, Article ID 456910 (2012)
24. Roberts, J. (ed.): Performance evaluation and design of multiservice networks. Final Report COST 224. Commission of the European Communities, Brussels (1992)
25. Little, J.: A proof for the queueing formula: L = w. Oper. Res. **9**(3), 383–387 (1961)
26. Chydziński, A., Adamczyk, B.: Analysis of a scheduler for virtualization of links with performance isolation. Appl. Math. Inf. Sci. **8**(6), 2653–2666 (2014)
27. Chydziński, A.: Delay analysis for an AQM queue with dropping function. In: Proceedings of the 11th WSEAS International Conference on Data Networks, Communications and Computers. World Scientific and Engineering Academy and Society, September 2012, pp. 53–57
28. Domański, A., Domańska, J., Czachórski, T.: Comparison of aqm control systems with the use of fluid flow approximation. In: Computer Networks, ser. Communications in Computer and Information Science, vol. 291, pp. 82–90. Springer (2012)

Clustering Algorithms and Data Streams for Supervised Control of Data and Prevention of Threats in Mobile Application Systems

Aneta Majchrzycka and Aneta Poniszewska-Marańda

Abstract The paper aims to present the possibilities of application of various data mining techniques in order to increase the security level of mobile application systems. The scope of work incorporates the usage of clustering algorithms—particularly Density-Based Spatial Clustering of Applications with Noise (*DBScan*)—as well as other mechanisms connected with data streams. The proposed solution is based on the process of monitoring the incoming server requests obtained from mobile devices which use server application to connect to the data.

Keywords Mobile systems · Security of mobile applications · Data exploration · Data mining · Clustering · Data streams

1 Introduction

With the increased popularity of mobile devices, the number of applications and entire systems designed specifically for this kind of appliances has grown rapidly over the past decade. Right now almost any—big or small—business enterprise finds the need of having a mobile application or at least a mobile web site indispensable as the lack of such solution quickly becomes a substantial drawback in the eyes of their customers.

After the initial phase of going into raptures over the numerous possibilities that the mobile applications offer, it has become clear that one of the weakest links regarding their usage is security. In case of many systems, using the mobile application is bound with the necessity to share private confidential data like login credentials, credit card numbers, personal information and others. This data is then

A. Majchrzycka (✉) · A. Poniszewska-Marańda
Institute of Information Technology, Lodz University of Technology, Łódź, Poland
e-mail: aneta.majchrzycka@p.lodz.pl

A. Poniszewska-Marańda
e-mail: aneta.poniszewska-maranda@p.lodz.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_15

sent over the Internet to the remote server where the request for the application resources is processed, verified and responded to.

This process gives the opportunity for the potential attackers to manipulate the data or the entire process itself and gain unauthorized access by impersonating valid mobile requests through stealing the users' identity. This issue should not be neglected when implementing the data transfer mechanisms. Nonetheless, it could also be supervised in order to detect and avoid potential threats.

In order to achieve this goal an idea of intermediate verification mechanism for analyzing and verifying the requests coming from mobile devices arouse. As the amount of data received by the server-side of mobile system will always be significant it was established that a suitable approach for such analysis would be data mining techniques. Initial assumption concerned the usage of the DBScan algorithm to detect intruding requests based mainly on the time and location attributes. In the course of work it turned out that the quantity of the requests to be analyzed in a real-time environment will cause the verification mechanism to be ineffective if all the requests were to be analyzed. Thus, the application of a common technique used for data streams—a sliding window was suggested in order to improve the performance of proposed solution [1–3].

The proposed solution is based on the process of monitoring the incoming server requests obtained from mobile devices which use server application to connect to the data. The analysis of these requests is performed based on series of selected attributes. In the framework of research the RequestAnalyzer application was created, which enables supervision over the request traffic on the server.

The paper is structured as follows: Sect. 2 presents the outline of data mining techniques used for request analysis process. Section 3 gives the short description of architecture of mobile systems. Chapter 4 deals with the details on implementation of chosen techniques in terms of mobile traffic analysis based on created *RequestAnalyzer* application while Sect. 5 describes the results of clustering analysis for security of mobile applications.

2 Clustering and Data Streams

One of the frequently used techniques for analyzing the data is clustering and associated outlier detection. Clustering of the data consists in dividing them into groups called clusters where all the elements in a single cluster are somehow more similar to one other than to other objects which are outside of this particular cluster [4–6].

The measure of similarity between the objects and the principle of assigning them to the clusters is different depending on the characteristics of data and the attributes which may be derived from them.

There exist various algorithms to achieve data clusterization and they are based on different properties of the data space. One of the most popular approaches is density-based clusterization and the example of an algorithm which realizes this task is the *Density-Based Spatial Clustering of Application with Noise* (*DBScan*) algorithm [4, 7].

The idea behind the *DBScan* algorithm is to create a data space where similar objects will be situated close to one another as neighbours. Then this data space should be divided into the densely-populated areas and these areas marked as separate clusters. The area can be considered a cluster if it satisfies a certain condition for the minimum number of objects which constitute a cluster, popularly referred to as *minPtr* parameter. All objects which could not have been assigned to any of the clusters can be considered as noise or in other words an anomaly from the normal behaviour of other objects in the data set. This is useful in detecting exceptions or potential risks within the analyzed scope of data [5, 7].

Data streams in data mining is a concept connected with rapidly growing sets of data due to the constant and fast flow of new information. Data streams are frequently used in cases where sensor systems are used. Such systems are fed with readouts from multiple sources within very short periods of time. All the readouts provide vital information although it is not possible to maintain satisfying performance quality while analyzing such vast amount of data. This is where data streaming techniques come in handy.

One of the approaches in dealing with large datasets is to limit the amount of analyzed data. Instead of the entire data set we choose a representative sample and base the analysis on a reasonably sized set [8]. This assumption is correct for the sets where the differentiation between consecutive data records is not significant. It means that a random sample selected on the basis of a given attribute can be considered as representative. The attribute used for the selection of representative data set may be any feature which is adequate taking into consideration the data domain in question. In many cases such characteristic attribute is the time when the record was included in the data set, other example may be the need to maintain a constant amount of samples for measurements. The concept of limiting the analyzed data set and varying it in time according to the newly obtained data is referred to as a concept of a sliding window. During every iteration when we are analyzing the data we use a different data set as a basis for analysis. This data set is modified with each iteration according to a pre-defined condition. This changing data set can be viewed as a sliding window as it slides along the data, sampling it and giving only the indispensable number of records submitted for further analysis [2, 3].

A sliding window concept can be useful in analyzing the request traffic in mobile application systems which fulfils the condition of rapidly growing data set with slight variance after imposing appropriate constraints.

3 Mobile Systems Architecture

Mobile application systems may vary depending on the nature of mobile application itself [9]. Nonetheless, rarely it happens that a mobile application acts without the necessity to connect to the external system. The most typical architecture of mobile application system beside mobile application itself includes:

- web service deployed on the application server,
- database where the system information is stored,
- other components of business system which may be complementary web or desktop application.

The mobile application connects to the server by means of a web service designed for this purpose. The web service provides methods which provide access to the application resources by ensuring the connection with application database, thus the mobile users can request all resources they require if they have appropriate permissions to obtain them. The most often the mobile devices use the same database as the rest of business system (Fig. 1).

The transfer of data in this case is frequent and very often it includes sending *sensitive data* information. The term *sensitive data* refers to all confidential pieces of information which should not be distributed or captured by any unauthorized third-party users or organizations. They may include: passwords, logins, access identifiers, credit card numbers and other application- or domain-specific data.

The transfer of such data cannot be completely avoided, however it should be limited to only indispensable operations and in case where it has to occur the data should be properly secured by the encryption algorithms. If possible the connection with the web service should be also secured by means of HTTPS and SSL protocols for connection.

The application of the aforementioned data mining techniques in case of helping to improve the overall security of the mobile systems will concern the analysis of the requests which are sent from the mobile devices to the application server.

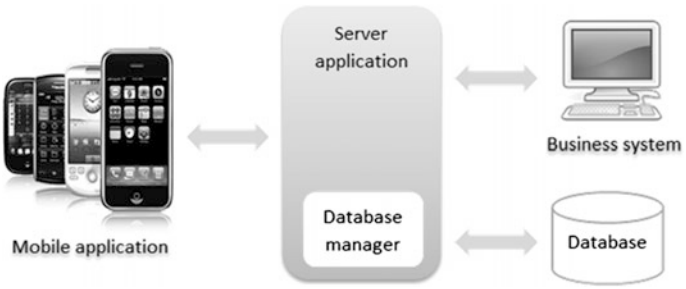

Fig. 1 Exemplary architecture of mobile application system

4 Clustering Techniques for Security of Mobile Applications

Maintaining security of an IT system of any kind is one of the crucial tasks of the system administrators. The following section describes the ways to incorporate data mining techniques in order to facilitate the process of maintaining system safety. It also presents the initial version of created *RequestAnalyzer* application, which provides an interface for described solution.

4.1 Initial Assumptions

The HTTP requests coming to the server can be easily captured, logged and stored as a source of data. The initial phase of the analysis process is to extract the necessary information from the data according to the defined list of attributes. For the purposes of data exploration of this kind it was established that the attributes of HTTP request which have to be taken into consideration are: time of the request, IP address from which the request was issued, application identifier, user identifier, device identifier, geolocation coordinates of incoming request, name of web service method called.

As far the geolocation coordinates are concerned they may be obtained two folds. They can be either extracted based on the IP address or they can be sent as the parameters in the request itself. The first case is more inaccurate as the IP addresses may simply indicate the internet provider location instead of the real user ones. The second case would require an assumption that all the requests come with geolocation parameters. Such an assumption was made as for the system to work effectively it would be the most beneficial. The same argument is valid in case of the application, user, and device identifier parameters. The mobile requests should be complemented with such data in order to be able to analyze them.

The usability of clustering technique was tested by detecting two separate threats within the request data set: *location-based intrusion detection* and *time-based intrusion detection*.

4.2 Location-Based Intrusion Detection

The attributes chosen for clustering in location-based intrusion detection are: time of the request, IP address together with extracted geolocation coordinates, application identifier, user identifier, device identifier.

The similarity function was based on the calculation of distance between points using the above mentioned attributes. The algorithm used the *minPtr* and *rad* as the parameters specifying the minimum size of a group to be considered as cluster and

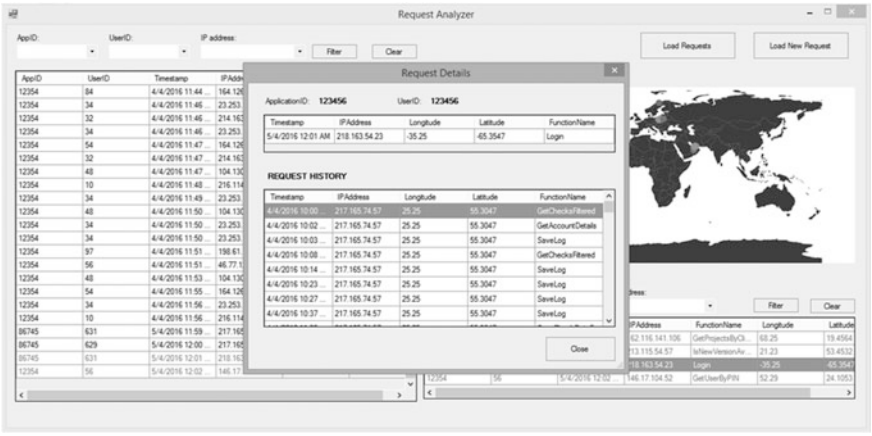

Fig. 2 *RequestAnalyzer* application—location-based intrusion detection

the maximum distance between points to be considered as belonging to the same group respectively. The setting of these parameters depends mainly on the characteristics of application the requests refer to.

The clustering should allow to group all requests indicating the regular usage of the application. Every cluster corresponds to a single user using a single application. The range of the cluster is defined by the radius parameter. Every request which was not assigned to a cluster is treated as noise and it is a potential threat. There is however one exception. The algorithm detects requests coming from new users and singles them out from the rest. If the new user sends more requests to the server then they will finally start to constitute separate cluster and they will be treated as regular ones.

Figure 2 presents a view of created *RequestAnalyzer* applications for location-based intrusion detection. Left-hand panel is an overview of all requests incoming to the server. The colours enable to differentiate suspicious server calls and regular ones.

The right-hand panel of the bottom window provides the information on requests which could not have been assigned to any of the existing clusters. The top window shows the potentially insecure request with respect to the most similar ones.

4.3 Time-Based Intrusion Detection

In case of *time-based intrusion detection* the attributes chosen for clustering are: time of the request, IP address from which the request was issued, application identifier, user identifier, device identifier, name of the web service method called.

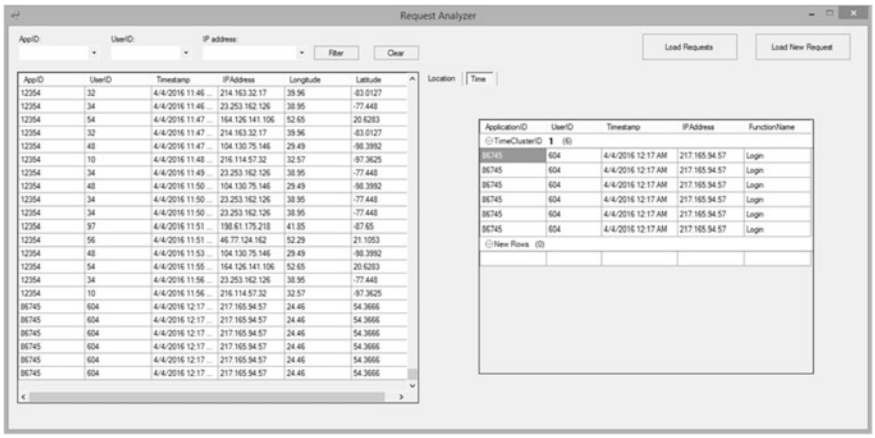

Fig. 3 *RequestAnalyzer* application—time-based intrusion detection

The algorithm uses the same concept of minimum cluster size and maximum radius. For two approaches the two values can be different and adjusted to the application specification.

Figure 3 presents the second view of *RequestAnalyzer* application. Right-hand side differs from the previously described view as the panel provides the information on the densely clustered requests with respect to the time dimension. Thanks to this information it is possible to see the details of the requests which may be a brute force attack to for instance log into the application.

5 Results of Cluster Analysis for Mobile Applications Security

The implementation of the previously described mechanisms allowed to carry out a series of tests in the specially created test environment. The test environment was static which means that the traffic of requests was artificially generated on demand. Also the potential intrusion requests were generated by the alteration of the real requests gathered from multiple mobile application servers.

The test environment included two sets of data:

- constant training set containing 2563 pre-processed requests with selected attributes; all requests in this set were valid—they didn't contain potential threats,
- test set containing 30 pre-processed requests; requests were modified from the original ones in order to be able to generate potential threats.

Based on these two data sets the cluster analysis was performed in following way:

1. training set was loaded and treated as a point of reference for other requests,
2. manual process of sending a new request to *RequestAnalyzer* application was invoked,
3. new request from the test set was loaded into the application and analyzed with reference to the older ones,
4. new request was either:
   - added to the existing cluster,
   - allowed to create a new cluster (the new user situation),
   - regarded as a noise.

Figure 4 depicts the division of the test data space into clusters. The diagram was prepared based on the test data set with generated noise records. Depending on chosen radius of neighbourhood the clusters may vary. The light gray dots represent the records grouped into a cluster together with their close neighbours. The dark gray dots were marked as a noise—the potentially threatening requests.

In the second case the aim of the research was to find the very densely populated regions with respect to the time attributes. Instead of looking for the objects constituting noise, the relevant requests were gathered very close to one another. Such tight groups of objects were potentially harmful requests aiming to break into application using the so called brute force method. Figure 5 presents the distribution of requests within the data space in such a situation. This time dark dots indicate not the noise but the clusters of data with very small radius of neighbourhood. These requests are the records to be reported as threats.

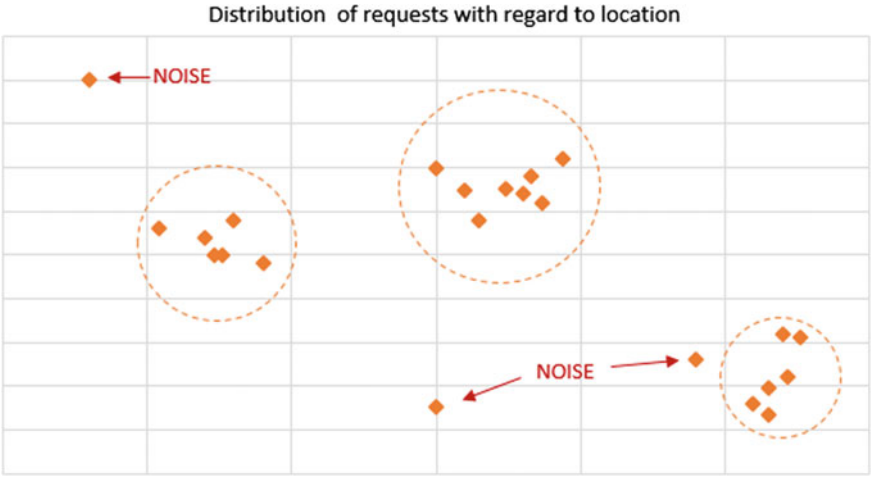

Fig. 4 Distribution of clusters of requests obtained from the test set based on location attributes

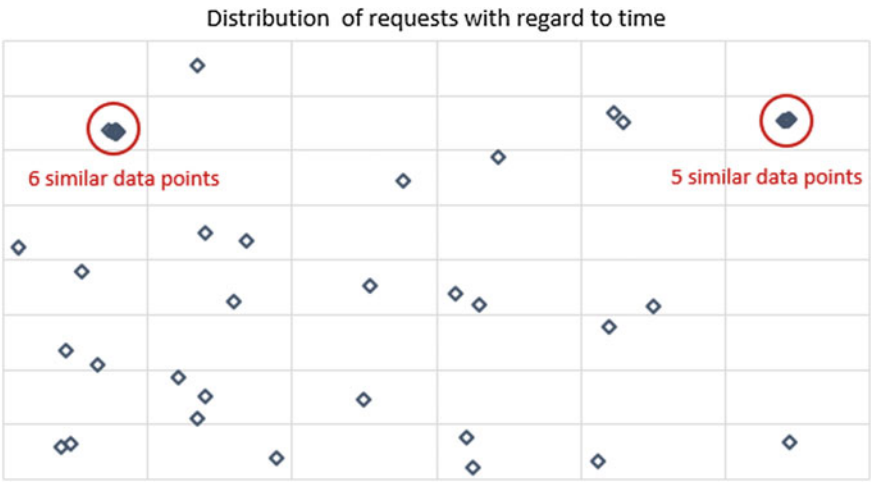

Fig. 5 Distribution of clusters of requests obtained from the test set based on time attributes

5.1 Performance Analysis

The initial tests were carried out for the entire set of test records which in the first phase of the experiment counted 2563 requests with extracted attributes. After establishing empirically the most optimal values for the DBScan algorithm parameters the test set was extended to include 10034 different requests. This has shown that increasing the number of the elements in the analyzed data set significantly increases the time of computation. If it was visible in a static test environment, then it would probably constitute a larger problem in a real-life dynamic situation. Thus, a sliding window technique commonly applied for rapidly growing streams of data was used. Instead of analyzing all the requests every time, the analysis was performed on a representative sample of the records in the data set. The elements from the data set were selected basing on the attributes. The series of tests were carried out to determine which attribute combination provides the most accurate indications. The results of these tests are shown in Table 1.

Table 1 Results of tests carried out for location-based intrusion detection in order to determine the most optimal selection of attributes for the sliding window

Selected attributes collection	Found threats	Actual threats	Found new users	Actual new users	Time of search
No attributes	10	10	2	2	Slow
Time	7	10	6	2	Medium
Application ID	10	10	2	2	Slow
Application ID, User ID	10	10	2	2	Medium
Time, Application ID, User ID	10	10	3	2	Fast

In terms of performance the most efficient method turned out to be the one using the most limited data set—the largest number of attributes. However, its effectiveness to recognize the new users was not optimal. The usage of time attribute caused that one of the requests was identified incorrectly as a new user request, due to the fact that this particular user ID was not used within the selected time period. A small improvement of the algorithm—to extend the time range if no records are found within the specified one could easily solve this problem.

6 Conclusions and Future Development

The test environment created for the purposes of effectiveness verification of implemented data mining techniques proved that they can be useful in monitoring the mobile request traffic on the application server. The prepared test cases enabled to detect all potential attack attempts. Thus, one may conclude that incorporating a similar mechanism in a real-life environment could bring comparable results and significantly contribute to security improvement of mobile application systems.

One of the observed drawbacks of suggested solution is its low tolerance for the unusual behaviours which may appear during the application usage especially in case of mobile devices. The example of such unusual behaviour may be the permanent or temporary relocation, which the system will interpret as a potential threat while in fact it is a common practice. The additional inclusion of time factor contributes to decreasing the quantity of incorrectly raised alerts. However, in order to reduce the amount of unnecessarily detected threat cases a complementary pattern recognition mechanism could be incorporated within the framework of the created solution.

The application of data mining techniques enabled to successfully detect a majority of suspicious requests incoming from hostile mobile devices. The created application can prove to be a valuable tool for server or application administrators in controlling the access to application resources. Nonetheless, a vast potential for improvement and additional functionalities has to be noted.

As far as RequestAnalyzer application is concerned the crucial issue is a complete parameterization of detection process of suspicious requests. This means that the users should be provided with an interface where they could parameterize the variables of the clustering and data streaming algorithms by setting the values in accordance to the application characteristics. They should be able to define for instance the minimum cluster size and adjust it based on the experience gained from previously gathered data. Another point of view is the adaptation of learning mechanism in order to automatically calculate the statistically appropriate parameters.

Moreover, as in the example of new application user, different patterns of using the application could be extracted based on the user's activity. The implementation

of detection mechanism for the known behavioural patterns together with machine learning techniques could vitally improve the process of distinguishing suspicious requests from natural changes in the application usage.

References

1. Tandon, G.: Machine Learning for Host-based Anomaly Detection. Florida Institute of Technology, Florida (2008)
2. Aggarwal, C.C., Yu, P.S.: Privacy-Preserving Data Mining: Models and Algorithms, Advances in Database Systems. Springer, Switzerland (2008)
3. Aggarwal, C.C.: Data Streams: Models and Algorithms, Advances in Database Systems. IBM T.J. Watson Research Centre, Springer, Switzerland (2007)
4. Markov, Z., Larose, D.T.: Data Mining the Web: Uncovering Patterns in Web Content, Structure and Usage. Wiley, New Jersey (2007)
5. Maloof, M.A.: Machine Learning and Data Mining for Computer: Methods and Applications, Advanced Information and Knowledge Processing. Springer, Switzerland (2006)
6. Dua, S., Du. X.: Data Mining and Machine Learning in Cybersecurity. CRC Press, Auerbach Publications, Taylor & Francis Group, LLC, Boca Raton (2011)
7. Witten, I.H., Frank, E., Hall M.A.: Data Mining: Practical Machine Learning Tools and Techniques. 3rd edn. Morgan Kaufamn, Burlington (2011)
8. Babcock, B., Babu, S., Datar, M., Motwani, R., Widom, J.: Models and issues in data stream systems. In: Proceedings of 21st ACM SIGMOD-SIGACT-SIGART, pp. 1–16 (2002)
9. Michalska, A., Poniszewska-Marańda, A.: Security risks and their prevention capabilities in mobile application development. Inf. Syst. Manag. **4**(2), 123–134 (2015) (WULS Press)

Using Fuzzy Knowledge Base to Evaluate a Classical Potential Barrier of Liquid Phase Reactions of *Tert*-Butyloxyl Radicals with Hydrocarbons

Vladimir E. Tumanov, Elena S. Amosova and Andrei I. Prokhorov

Abstract A classical potential barrier of liquid phase reactions of *tert*-Butyloxyl radical with hydrocarbons has been approximated using fuzzy knowledge base built from the experimental sample. The predicted values of the classical potential barrier have been compared with the experimental values on the testing sample. Experimental and calculated values are in good agreement within the error. Weak dependence of activation energy of such reactions on the solvent type is discovered. A feedforward artificial neural network has been developed to approximate the classical potential barrier of studied reactions and the obtained results have been compared with the results obtained from the fuzzy knowledge base. Using fuzzy knowledge base produces more precise prediction of the classical potential barrier of given reactions.

Keywords Knowledge discovery · Fuzzy knowledge base · Feedforward artificial neural network · Radical abstraction reaction

1 Introduction

Chemical society watches closely the development of methods of artificial intelligence and applies them in the research [1–3]. In particular, it studies the use of applied methods of artificial intelligence for production of new knowledge on the kinetics from the electronic collections.

V.E. Tumanov (✉) · E.S. Amosova · A.I. Prokhorov
Department of Computing and Information Resources, Institute of Problems of Chemical Physics RAS, 142432, Chernogolovka, Semenov Ave, 1, Moscow, Russian Federation
e-mail: tve@icp.ac.ru

E.S. Amosova
e-mail: aes@icp.ac.ru

A.I. Prokhorov
e-mail: aipro@icp.ac.ru

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_16

The study [4] is dedicated to the problem of application of fuzzy logic in chemistry. However, using fuzzy logics in the research of radical chemical kinetics in the liquid phase is still an open issue.

Reactivity of organic compounds in radical abstraction reactions both in the gas and the liquid phases has a great significance for research on technological process engineering for various chemical technologies, including creation of new materials. This parameter is calculated from the rate constant of radical abstraction reaction and might be used in the computer simulation of processes involving liquid-phase radical reactions. Typically, modeling of technological chemical processes includes calculation of parameters from 10 to 100 chemical reactions. For the half of studied reactions the rate constants are unknown and it is necessary to apply theoretical and empirical methods to evaluate them. To date, in the liquid-phase kinetics it was determined experimentally no more than 40,000 rate constants of radical reactions [5], which is not sufficient for the computer simulation of radical reactions in the liquid phase processes. Methods of artificial intelligence open up new opportunities for theoretical evaluation of the rate constants of radical abstraction reactions, including difficult to study, in the liquid phase through the calculation of the classical potential barrier of such reactions based on the experimental data. Therefore, using fuzzy logic methods for prediction of the classical potential barrier of liquid phase radical reactions of abstraction is a scientific and technical problem of current purposeful interest.

The purpose of this paper is to use a fuzzy knowledge base built from the experimental sample to evaluate the classical potential barrier of liquid phase reactions of *tert*-Butyloxyl radicals with hydrocarbons.

Authors previously have made an attempt to approximate the classical potential barrier of liquid phase reactions of phenyl radical with hydrocarbons [6].

2 Problem Formulation

In [7], an empirical model of elementary bimolecular radical reactions of abstraction has been proposed, which allows to build non-linear correlations between the classical potential barrier of the radical bimolecular reaction and thermochemical properties of the reactants (nonlinear correlation):

$$br_e = D_{ei}^{1/2} \ln\left(\frac{D_{ei}^{1/2}}{D_{ei}^{1/2} - E_e^{1/2}}\right) + \alpha D_{ef}^{1/2} \ln\left(\frac{D_{ef}^{1/2}}{D_{ef}^{1/2} - (E_e - \Delta H_e)^{1/2}}\right) \quad (1)$$

where $\Delta H_e = D_i - D_f + 0.5\ (hLv_i - hLv_f)$ is the enthalpy of reaction, including the difference between the zero-point energies to rupture or form the bonds, ν_i and ν_f are the frequencies of the stretching vibrations for the bond being broken and the bond being formed, respectively; D_i and D_f are the dissociation energies for the breaking and the forming bond, respectively; $D_{ei} = D_i + 0.5hLv_i$ and $D_{ef} = D_f + 0.5hLv_f$;

μ_i and μ_f are the reduced masses of atoms for the breaking and the forming bond, respectively; the coefficients $b = \pi(2\mu_i)^{1/2}v_i$, $b_f = \pi(2\mu_f)^{1/2}v_f$ and $\alpha = b/b_f$, the distance r_e, which the abstracted atom is displaced in the course of the reaction.

Within the proposed empirical model of radical abstraction reactions the intersection point of two potential curves approximates the value of the classical potential barrier of such reaction. The physical meaning of the parameter br_e is that it is defined by the value of the classical potential barrier of the thermoneutral reaction for the considered reaction ($\Delta H_e = 0$). As shown in [8], this model is a generalization of the well known correlation of Polanyi-Semenov [9].

We will designate br_e empirical index of the reactionary center of the radical reaction.

Experience has shown that for the same type of reaction the empirical index of the reactionary center may be assumed as constant within the statistical calculation error. Exceptions are the water radical reactions where the classical potential barrier depends on pH, and the empirical index of the reactionary center will not be constant for a particular reaction.

Knowing the empirical index of the reactionary center and thermo chemical parameters of the radical reaction of abstraction, it is possible to evaluate the classical potential barrier by solving the Eq. (1). However, having experimental set of the rate constants of reactions of hydrogen atoms and hydrocarbons, we can assume that the classical potential barrier of such reactions nonlinearly depends on the thermochemical characteristics of the reagents and kinetic characteristics of such reaction:

$$E_e = f\left(D_{ei}, D_{ef}, b_{re}, \alpha\right) \tag{2}$$

and set the task of approximation of this dependence by a feed forward artificial neural network, or by a fuzzy knowledge base.

Currently, several types of neural networks and their training methods have been developed; various approaches to design fuzzy knowledge bases on the empirical data to solve difference applied and theoretical problems exist. Research on comparison of proposed algorithms has been made [10].

In this work we focus on a feedforward artificial neural network trained by the backpropagation method and on building a fuzzy knowledge base by experts, and further using Mamdani inference method. The reason for this choice was the simplicity of software implementation.

3 Problem Solution

The experimental sample of the rate constants

$$(CH_3)_3CO^{O} + RH \rightarrow (CH_3)_3COH + R^{O} \tag{3}$$

where RH—is hydrocarbon, R^O—is alkyl radical, $(CH_3)_3CO^O$—*tert*-butyloxyl radical, has been obtained from the intelligent system database of rate constants of the liquid phase radical reactions [5].

The initial sample has been preprocessed. At first, all the rate constants have been brought to the same temperature T = 298 K. The classical potential barrier of radical reaction for this temperature has been calculated

$$E_e = E - 0.5(hLv_i - RT) \tag{4}$$

where v_i is the frequency of the stretching vibrations for the bond being broken, R is the gas constant, h is the Planck constant, L is the Avogadro constant, and T is the reaction temperature (K).

Activation energy of radical reaction has been calculated by the formula

$$E = -RT \ln\left(\frac{k}{nA_0}\right) \tag{5}$$

where A_0 is the pre-exponential factor per equireactive bond in the molecule and equal to 108–109 L mol^{-1} s^{-1}, n is the number of equireactive bonds in the molecule, k is the reaction rate constant, and E is the activation energy from (5).

Then duplicates of the same reaction have been deleted from the sample. Solvents have been divided into four classes: non-polar, complex non-polar, complex polar and polar. The experiment has shown that the constant rate of the same reaction in different solvents depends weakly on the solvent type. To neutralize this dependence in the knowledge base in the sample reactions the polar solvents have been preferred.

As a result, the experimental sample used for the fuzzy knowledge base consists of 192 radical reactions, and the testing sample consisted of 31 radical reactions.

The dissociation energy of the C–H bonds is taken from [11]. For this sample D_f = 449.4 kJ/mol and α = 0.796 are constant. Therefore, the dependence (2) takes the form:

$$E_e = f_e(D_{ei}, b_{re\_ind}) \tag{6}$$

Then the analysis of the values br_e has been performed for various reaction centers and the empirical index of the reactionary center br_{e_ind} has been calculated, as shown in Table 1. We used br_{e_ind} calculated for the certain reaction centers for prediction of the classical potential barrier.

The fuzzy knowledge base has been built by the experts and includes 192 linguistic rules of the form:

$$R_i \textbf{ IF } D_{ei} \text{ about } x_1 \textbf{ AND } br_e \text{ about } x_2 \textbf{ THEN } E_e \text{ about } y$$

Symmetrical triangular function has been chosen as the membership function. To approximate the values of the classical potential barrier the Mamdani's fuzzy

Table 1 Empirical index of the reactionary centers

Organic compound class	Reactionary center	Empirical index of the reactionary center $(kJ/mol)^{0.5}$
Alkanes	$–C^{\cdot}HCH_3$	15.73 ± 0.11
	$–C^{\cdot}(CH_3)_2$	15.44
	$(CH_3)_2–\ldots–C^{\cdot}(CH_3)_2$	16.34 ± 0.59
	cyclo-$[C^{\cdot}H(CH_2)_k]$	15.94 ± 0.39
	cyclo-$[C^{\cdot}(CH_3)(CH_2)_k]$	15.16
	bicycle-$C^{\cdot}$	16.31 ± 0.39
	Decalin	14.93
Alkenes	$=CHC^{\cdot}H_2$	18.52 ± 0.25
	$=CHC^{\cdot}H(CH_2)_kCH_3$	18.71 ± 0.13
	$–C^{\cdot}CH=CH_2$	18.81 ± 0.19
	cyclo-$[CH=CHC^{\cdot}H(CH_2)_k]$	18.61 ± 0.63
	$=CHC^{\cdot}HCH=$	20.22
Arenes	$C_6H_5C^{\cdot}H_2$	17.70
	$C_6H_5C^{\cdot}H–$	17.65 ± 0.03
	$C_6H_5C^{\cdot}–$	19.01
	n–X–$C_6H_4C^{\cdot}H_2$	18.15 ± 0.66
Alcohols	$C^{\cdot}H_2OH$	15.88
	$CH_3C^{\cdot}HOH$	16.03
	$C^{\cdot}H_2(CH_3)_2COH$	15.13
	$(CH_3)_2C^{\cdot}OH$	16.18
	cyclo-$[(CH_2)_kC^{\cdot}(OH)]$	16.23
	$CH_2=CHC^{\cdot}HOH$	17.29
	$CH_3(C_6H_5)C^{\cdot}OH$	16.87
	$C_6H_5C^{\cdot}HOH$	16.09
	4–X–$C_6H_4C^{\cdot}HOH$	22.89
Aldehydes	$–CH_2C^{\cdot}(O)$	15.44
	$C_6H_5C^{\cdot}(O)$	16.33
	n–X–$C_6H_4C^{\cdot}(O)$	15.49 ± 0.15
Ketones	$–C(O)C^{\cdot}H_2$	17.00
	$–C(O)C^{\cdot}HCH_3$	17.29
	cyclo-$[C(O)C^{\cdot}H(CH_2)_k]$	16.44
	$C_6H_5C(O)C^{\cdot}H_2$	16.62
	$C_6H_5C(O)C^{\cdot}HCH_3$	17.16
Ethers	$–OC^{\cdot}H_2$	15.67
	$–OC^{\cdot}HO–$	16.18 ± 0.49
	$>C^{\cdot}(OCH_3)_2$	16.11 ± 0.33
	cyclo-$[OC^{\cdot}H(CH_2)_k]$	15.95 ± 0.17
	cyclo-$[OC^{\cdot}(CH_3)(CH_2)_k]$	16.22 ± 0.29
	cyclo-$[OC(CH_3)_2SC^{\cdot}H(CH_2)_k]$	16.13
	cyclo-$[OC^{\cdot}(CH3)S(CH_2)2]$	15.93 ± 0.19
	$–OC^{\cdot}HC_6H_5$	17.57
Esters	$–C(O)OC^{\cdot}H_2$	17.32
	$C_6H_5C(O)OC^{\cdot}H_2$	15.81

(continued)

Table 1 (continued)

Organic compound class	Reactionary center	Empirical index of the reactionary center $(kJ/mol)^{0.5}$
Asids	*cyclo*-[C˙(C(O)OH)$(CH_2)_k$]	16.48
Nitriles	C˙H_2CN –C˙HCN	18.60 18.81
Nitro compounds	C˙H_2NO_2 –C˙HNO_2	16.48 18.12
Sulfides	–SC˙HS– $(–S)_2$CHC˙< *cyclo*-[SC$(CH_3)_2$SC˙H$(CH_2)_k$] *cyclo*-[SC˙(CH_3)S$(CH_2)_k$]	16.15 15.57 15.37 15.58 ± 0.67
Amines	(CH_3)C˙H_2NH $(CH_3)_2$C˙H_2N $(CH_3CH_2)_2(CH_3$C˙H)N $(CH_2$=CHC$H_2)_2(CH_2$=CHC˙H)N *cyclo*-[NHC˙H$(CH_2)_k$] *cyclo*-[C˙$(NH_2)(CH_2)_k$] $(C_6H_5CH_2)_2(C_6H_5$C˙H)N C_6H_5N(C˙$HCH_3)(CH_2CH_3)$ C_6H_5N(C˙$H_2)(CH_3)$	14.50 16.29 15.05 17.13 15.36 16.08 ± 0.17 15.02 13.91 15.49
Benzene	C_6H_5˙	12.97

inference method has been applied based on the use of the matching degree [4, 5]. The Mean of Maximum (MoM) method has been used for defuzzification.

The obtained theoretical values have been compared with the experimental data for the given class of radical abstraction reactions.

Table 2 shows a comparison of the values of the classical potential barrier of reactions of *tert*-Butyloxyl radicals with hydrocarbons E_e, obtained with the fuzzy knowledge base, and the calculated ones from the experimental values of the activation energy of these reactions E_{exp}.

As can be seen in Table 2, there is a good agreement between the experimental and obtained by fuzzy inference values of the classical potential barrier.

Figure 1 shows correlation between calculated and experimental values in the testing sample. The simple regression coefficient for given linear dependence equals 0.988, Fisher's F-criterion equals 39.2, and coefficient of determination equals 0.975.

It is probable that the difference between the experimental and predicted values is related to the weak dependence of activation energy of radical reactions on the solvent type. To account the solvent effect in the fuzzy knowledge base it is required to fill the missing data depending on the solvent type. Experimental data is not enough to fill the gap. In future research it is necessary to study the dependence of activation energy of liquid phase radical reactions on the solvent.

Table 2 Comparison of the values of classical potential barrier of reactions for *tert*-Butyloxyl radicals with hydrocarbons for the fuzzy knowledge base

Hydrocarbons	E_e	E_{exp}	ΔE
	kJ/mol		
$CH_3(CH_2)_5CH_3$	38.42	35.90	2.52
cyclo-[$(CH_2)_6$]	36.99	36.01	0.98
cyclo-[$CH(CH_3)(CH_2)_5$]	28.57	28.84	−0.27
cyclo-[$CH{=}CH(CH_2)_3$]	25.14	24.80	0.34
cyclo-[$CH{=}CH(CH_2)_6$]	27.02	27.02	0.00
4-OCH_3–$C_6H_4CH_3$	30.78	30.76	0.02
3-Cl–$C_6H_4CH_3$	43.68	43.68	0.00
4-Cl–$C_6H_4CH_3$	34.83	34.09	0.74
4-F–$C_6H_4CH_3$	33.82	32.14	1.68
1,3-$(CH_3)_2$–C_6H_4	33.23	33.38	−0.15
1,3,5-$(CH_3)_3$–C_6H_3	33.33	33.03	0.30
1,4-$(CH_3)_2$–C_6H_4	33.60	33.00	0.60
3–CN–$C_6H_4CH_3$	33.44	34.63	−1.19
3-NO_2–$C_6H_4CH_3$	37.02	37.02	0.00
4-OC_6H_5–$C_6H_4CH_3$	32.95	32.88	0.07
$C_6H_5CH_3$	33.39	33.78	−0.39
1,2-*cyclo*-$(CH_2)_4$–C_6H_4	24.10	25.53	−1.43
$C_6H_5CH_2CH_3$	28.96	29.06	−0.10
$C_6H_5CH(CH_3)_2$	27.82	28.36	−0.54
CH_3OH	38.39	38.86	−0.47
$CH_3C(O)CH_3$	44.60	44.60	0.00
cyclo-[$C(O)(CH_2)_5$]	36.42	37.31	−0.89
CH_3OCH_3	37.66	38.51	−0.85
cyclo-[$O(CH_2)_4$]	31.34	31.61	−0.27
$CHCl_3$	35.38	34.35	1.03
CH_3CH_2Cl	36.58	36.58	0.00
CH_3CH_2Br	41.08	41.08	0.00
$(CH_3CH_2)_3N$	22.09	22.40	−0.31
$CH_3CH_2SCH_2SCH_2CH_3$	31.44	34.42	2.98
cyclo-[$SCH(CH(CH_3)_2)S(CH_2)_2$]	22.87	22.87	0.00
cyclo-[$SCH(CH_3)S(CH_2)_2$]	22.63	22.97	−0.34

In the previous paper [6] we used Mamdani's algorithm with the bell-shaped membership function and defuzzification by "center of gravity" method to approximate the classical potential barrier of radical reactions of abstraction. Such approach requires adjusting parameters of membership function during the training.

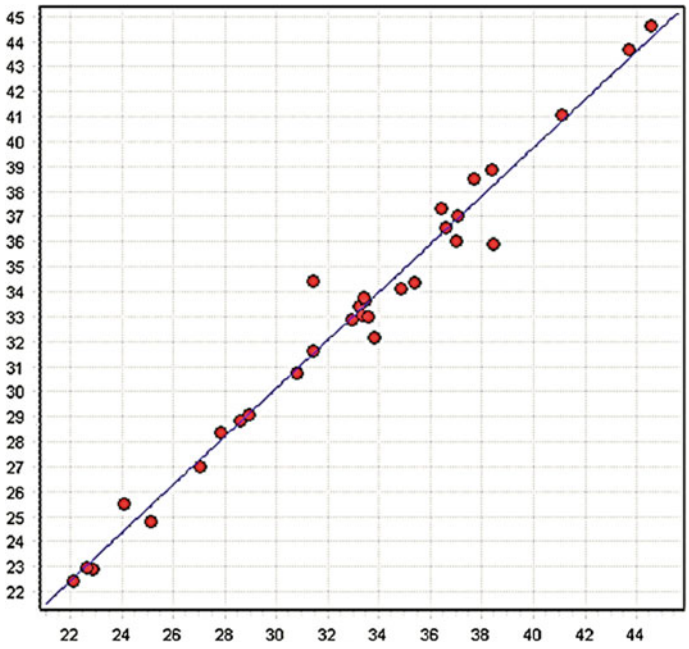

Fig. 1 Correlation between the experimental and calculated values in the testing sample in case of the fuzzy knowledge base

Using average maximum for defuzzification allows avoiding such adjusting in construction of the fuzzy rule base by the experts and, in our opinion; it is more efficient from the point of view of inference.

Function (6) can be approximated by the artificial neural network built according to the method from [12]. The feedforward artificial network with 2 inputs, 2 hidden layers of 5 and 7 neurons and one output has been trained by the backpropagation method, in order to compare the prediction of the classical potential barrier by the fuzzy knowledge base and the artificial neural network. A sigmoid function has been used to activate neurons.

The training results are given in Table 3 and Fig. 2.

Figure 2 (the correlation coefficient equals 0.935, Fisher's F-criterion equals 7.655, and the coefficient of determination equals 0.874) and Table 3 show that the results obtained from the artificial neural network are less accurate. Thus we can make a conclusion that using fuzzy knowledge bases in this case has some advantages: the fuzzy knowledge base can interpret the obtained knowledge from

Table 3 Comparison of the values of classical potential barrier of reactions of *tert*-Butyloxyl radicals with hydrocarbons for the artificial neural network

Hydrocarbons	E_e	E_{exp}	ΔE
	kJ/mol		
$CH_3(CH_2)_5CH_3$	37.26	35.90	1.37
cyclo-[$(CH_2)_6$]	36.74	36.01	0.73
cyclo-[$CH(CH_3)(CH_2)_5$]	27.97	28.84	−0.87
cyclo-[$CH{=}CH(CH_2)_3$]	25.93	24.80	0.13
cyclo-[$CH{=}CH(CH_2)_6$]	25.93	27.02	−1.09
4-OCH_3–$C_6H_4CH_3$	30.60	30.76	−0.16
3-Cl–$C_6H_4CH_3$	35.94	43.68	−7.74
4-Cl–$C_6H_4CH_3$	35.60	34.09	1.51
4-F–$C_6H_4CH_3$	33.85	32.14	1.71
1,3-$(CH_3)_2$–C_6H_4	32.66	33.38	−0.72
1,3,5-$(CH_3)_3$–C_6H_3	32.51	33.03	−0.52
1,4-$(CH_3)_2$–C_6H_4	31.98	33.00	−1.02
3-CN–$C_6H_4CH_3$	33.93	34.63	−0.07
3-NO_2–$C_6H_4CH_3$	35.48	37.02	−1.54
4-OC_6H_5–$C_6H_4CH_3$	36.44	32.88	3.56
$C_6H_5CH_3$	32.99	33.78	−0.79
1,2-*cyclo*-$(CH_2)_4$–C_6H_4	24.99	25.53	−0.54
$C_6H_5CH_2CH_3$	28.75	29.06	−0.31
$C_6H_5CH(CH_3)_2$	30.93	28.36	2.57
CH_3OH	37.30	38.86	−1.56
$CH_3C(O)CH_3$	45.14	44.60	0.54
cyclo-[$C(O)(CH_2)_5$]	33.66	37.31	−3.65
CH_3OCH_3	36.47	38.51	−2.04
cyclo-[$O(CH_2)_4$]	30.28	31.61	−1.33
$CHCl_3$	34.65	34.35	0.03
CH_3CH_2Cl	35.52	36.58	−1.06
CH_3CH_2Br	40.65	41.08	−0.43
$(CH_3CH_2)_3N$	22.92	22.40	0.52
$CH_3CH_2SCH_2SCH_2CH_3$	30.74	34.42	−3.68
cyclo-[$SCH(CH(CH_3)_2)S(CH_2)_2$]	22.10	22.87	−0.77
cyclo-[$SCH(CH_3)S(CH_2)_2$]	21.69	22.97	−1.28

the point of view of chemical kinetics and can predict more precisely the classical potential barrier of the studied radical reactions.

Programs used in this study have been written by authors in Visual Basic 6 for OS Windows. The results were statistically analyzed using the program Statgraf.

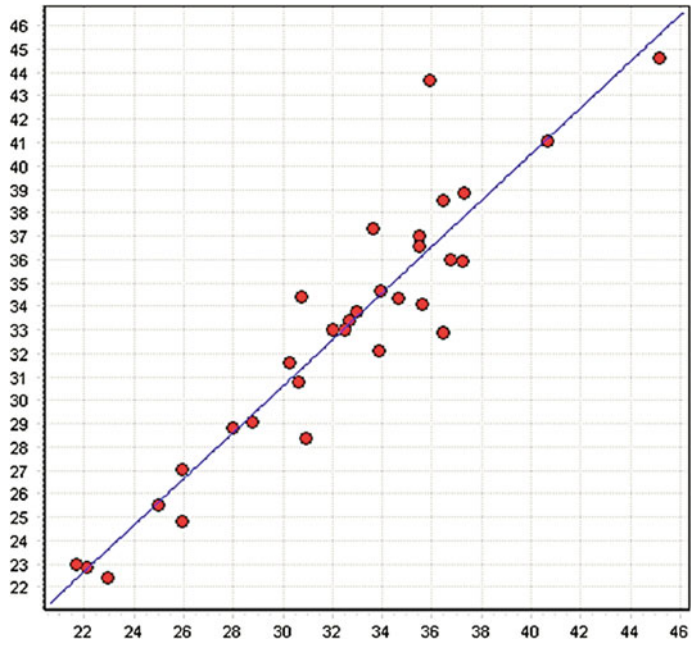

Fig. 2 The correlation between the calculated and experimental values for the test sample in the case of the artificial neural network

4 Conclusion

In this paper the classical potential barrier of liquid phase reactions of *tert*-Butyloxyl radicals with hydrocarbons has been approximated using fuzzy knowledge base built on the experimental sample. The study shows that the calculated value of a classical potential barrier is in a good agreement with the experiment (±4.0 kJ/mol).

Also, the feedforward artificial neural network has been developed to approximate the classical potential barrier of the studied reactions. The obtained results have been compared with the results obtained from the fuzzy knowledge base. Using the fuzzy knowledge base gives a more accurate prediction of the classical potential barrier of the given reaction class according to the made research.

To modify fuzzy knowledge base it is necessary to study additionally the effect of the solvent on the values of activation energy of such reactions.

References

1. Brereton, R.G.: Applied Chemometrics for Scientists. Wiley, West Sussex (2007)
2. Banville D.L.: Chemical Information Mining, Facilitating Literature-Based Discovery. CRC Press, Taylor & Francis Group, Boca Raton (2009)
3. Gasteiger, J., Engel, T.: Chemoinformatics: A Texbook. Wiley-VCH GmbH & Co. KGaA, Weinheim (2003)
4. Rouvray, D.H.: Fuzzy Logic in Chemistry. Academic Press, New York (1997)
5. Tumanov, V., Gaifullin, B.: Subject-oriented science intelligent system on physical chemistry of radical reactions. In: Ding, W., Jiang, H., Ali, M., Li, M. (eds.) Modern Advances in Intelligent Systems and Tools, SCI, vol. 431, pp. 121–126, Springer, Heidelberg (2012)
6. Tumanov, V.E.: Knowledge extraction from the problem-oriented data warehouses on chemical kinetics and thermochemistry. WSEAS Trans. Comput. **15**(9), 93–102 (2016)
7. Denisov, E.T., Tumanov, V.E.: Transition-state model as the result of 2 morse terms crossing applied to atomic-hydrogen reactions. Zurnal fiziceskoj himii **68**(4), 719–725 (1994)
8. Denisov, E.T.: New empirical models of radical abstraction reactions. Russ. Chem. Rev. **66** (10), 953–971 (1997)
9. Semenov, N.N.: Some problems in chemical kinetics and reactivity. Academy of Sciences of the USSR, Moscow (1958)
10. Larose D.T., Data mining methods and models. Wiley, New Jersey (2006)
11. Luo, Y.-R.: Comprehensive Handbook of Chemical Bond Energies. CRC Press, Boca Raton, London, New York (2007)
12. Tumanov, V.E.: Hybrid algorithm of application of artificial neuronets for an evaluation of rate constants of radical bimolecular reactions. In: Balicski, J. (ed) Advances in neural networks, fuzzy systems and artificial intelligence. Recent Advances in Computer Engineering Series, vol. 21, pp. 58–61, WSEAS Press, Gdansk, Poland (2014)

Tree Based Reduction of Concept Lattices Based on Conceptual Indexes

Miroslav Smatana, Peter Butka and Lenka Cöveková

Abstract There are many approaches and tools which deal with conceptual structures in datasets and their main goal is to support user in understanding of data and structure. One of methods is formal concept analysis (FCA) which is suitable for processing and analyzing input data of object-attributes models based on their relationship. One from FCA family is model of generalized one-sided concept lattice (GOSCL). It is suitable to work with different type of attributes. While generating one-sided concept lattices in FCA improved understanding and interpretation of analysis, one of the lasting problem is to provide the users a result of FCA in appropriate form, if there is large number of concept lattices and generated structure is complex. This is one of the main topics in the FCA and solution can be reached with the reduction methods. In this paper we propose some of the reduction techniques and their combinations.

Keywords Formal concept analysis · One-sided concept lattices · Reductions · Conceptual indexes

1 Introduction

The main point of this paper is one of the approaches used for data analysis, which applies to the theory of concept lattices, known as FCA and is suitable for processing input data of object-attributes models. Concept lattices are hierarchically

M. Smatana (✉) · P. Butka · L. Cöveková
Faculty of Electrical Engineering and Informatics, Department of Cybernetics and Artificial Intelligence, Technical University of Košice, Letná 9/B, 042 00 Košice, Slovakia
e-mail: miroslav.smatana@tuke.sk

P. Butka
e-mail: peter.butka@tuke.sk

L. Cöveková
e-mail: lenka.covekova@student.tuke.sk

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_17

organized structure of clusters of objects-concepts based on presence of their shared attributes. The methods from this field have found application in various field such as concept analysis of data, machine learning, data mining, information retrieval, software engineering and knowledge management. The standard approach to FCA is based on a "crips" approach, where object-attribute models are based on binary relations (object should/should not attribute). Given that much data is described other types of attributes have been proposed several approaches for working with multiple contextual inputs such as method of concept scaling and fuzzy approach. In order to provide more practical data mining tools, one-sided fuzzification is used instead of full fuzzy models. It means that only one part of object-attribute model is fuzzificated and the resulting model is usually called one-sided conceptual lattice. The algorithm for the processing one-sided concept lattice—GOSCL is capable to process table with different type of attributes, that is for each attribute can be used various structures of true values. This models simplifies interpretation of the analysis, but problem remains to provide the users a result of FCA in appropriate form, if there is large number of concept lattices and generated structure is complex. It is one of the main topics in FCA and solution can have reached with reduction methods and their combinations. In this paper we describe reduction methods of concept lattices and their combinations.

2 Formal Concept Analysis (FCA)

When people talk about objects and attributes they formulate their knowledge about the outside world. They express various arguments that some objects share the same values for certain attributes (properties). These relationships between objects and attributes are in most cases represented by a table, where the rows are objects and columns represent attributes. One of the methods of analysis of tabular data is a formal concept analysis [1, 2]. The term FCA is often replaced by term "method of concept lattices", where the input for analysis of data are tabular data. FCA is a method of exploratory data analysis. This method gives the user non-trivial information about input data, which can be used directly or may be used in the further processing of data.

It has been applied in different areas as knowledge discovery and data/text mining [3, 4], association rule mining [5, 6], or information retrieval.

FCA provides two primary output:

1. Concept lattice is hierarchically arranged set of clusters—formal concepts that are contained in the input table of data. Formal concepts can be compared with each other. Unless it can be compared, so we can say about them, that one is wider than the other or specific. Important feature of concept lattice is its simple graphical representation. The user can examine the data in graphical form, which incorporates the hierarchical arrangement.
2. Attribute implications describe dependences between attributes of input data.

To understand the FCA need to explain some terms and these terms are explained in [7].

There are several algorithms for building of concept lattices as NextClosure [8], UpperNeighbor [7] and Generalized One-Sided Concept Lattice (GOSCL) which is described in the next chapter.

3 Generalized One-Side Concept Lattice (GOSCL)

For the purpose of this paper we only describe some basic details regarding the theory of GOSCL [9, 10], some other properties on heterogeneous concept lattices are described in [11, 12]. For generalized one-sided formal context can be considered as a 4-tuple $c = (A, B, L, R)$ that meets the following conditions:

1. B is a non-empty set of all attributes and A is non-empty set of all objects.
2. L: $B \rightarrow CL$ is a view of a set of attributes to a class all complete of lattices CL. $L(b)$ is the structure of truth values of attribute b for each attribute b.
3. R is a generalized incidence session, $R(a,b) \in L(b)$ for all $b \in B$ a $a \in A$. $R(a, b)$ represents the degree of $L(b)$, which has object $a \in A$ in attribute b.

A data table for analysis represents the relation R. The ability to create concept lattice from table that contain attributes of different types is one main difference compared to other approaches.

```
ALGORITHM (Generalized One- Sided Concept Lattice)
1: Input (A, B, L, R) - generalized one- sided concept
lattice
2: begin
3:  create lattice  L := Π_{b∈B} L(b)
4:  C := {1_L}, C ⊆ L       (set of all intents)
5:  while (A ≠ ∅ )
6:  {
7:    choose a ∈ A
8:    C* := C
9:    for each c ∈ C*
10:       C := C ∪{c ∧ R (a)}    (meet operation on lattice)
11:   A := A / {a}
12: }
13:   C(A, B, L, R) := ∅
14:   for each c ∈ C
15:     C(A, B, L, R) := C(A, B, L, R) ∪{(c', c)}
16: end
17: Output: C(A, B, L, R)     (set of all concepts)
```

This algorithm produces concepts of generalized one-sided concept lattice. It starts with the one concept (defined as 1_L in the algorithm) containing largest value for every attribute. Operator $\wedge$ is standard lattice operation *meet* on attribute lattices truth value structure defined by L. The notation x' is used if we want to find all objects that have at least x level of attributes (x is subset of attributes values defining the concept). Similarly, we can use same notation for opposite function, e.g., if y is some subset of objects defining the concept, then y' is respective definition in attributes values.

4 Related Work

Formal concept analysis is used in many applications. However, when used for the analysis of medium and large number of data, concept lattices can be large and opaque and they should therefore be reduced. The results are so useful and easier to understand. For that purpose, some methods were introduced. Some of the techniques are threshold-based, where main goal is to find relevant concepts using some ranking method and remove the concepts under the threshold and examples can be found in [13] and [14]. In [15] authors proposed another interesting way of concept lattice reduction based on removing links between concepts in order to get tree-based structure. Other reduction methods are fault tolerance introduced in [16] and concept clustering described in [15]. Reductions based on the usage of singular value decomposition (SVD) can be found in [17] and [18].

4.1 Reduction Methods Based on Conceptual Indexes

In this paper we will focus on reduction based on removing the edges between concepts in concept lattice consider helping conceptual indexes. This method begins with the most specific terms, e.g. at the bottom of lattice and recursively calculate an index for each of the candidates and remove links to all higher-level concepts. For this method of extraction exists algorithm, where we assume that $c = (A,B,L,R)$ is a generalized formal context and (X,Y), (Z,U) are formal concepts of $C(A,B,L,R)$, with $X,Z \subseteq A$ and $Y,U \subseteq U$.

```
ALGORITHM ExtractTree
1: Concepts := Algorithm_GOSCL(c)
2: for each (X, Y) in Concepts
3:  Parents [(X, Y)] := list of parent concepts of (X, Y)
4:  if |Parents [(X, Y)] > 1
5:    max_score = -1
6:    for each (Z, U) in Parents [(X, Y)]
7:      Score [(Z, U)] := the score for concept (Z, U)
8:      if Score [(Z, U)] > max_score
9:        selected_parents [(X, Y)] := (Z, U)
10:         max_score := Score [(Z, U)]
11:     end if
12:     end for
13:   end if
14: end for
15: return selected_parents
```

Choosing a single parent concept at each step leads to some information loss. Our goal is to minimize this loss by selecting parents using the most relevant criteria. There are strategies for selecting parent concepts, including the stability and support indexes from FCA literature [19], as well as the confidence and similarity measures.

To improve understanding of the analysis by FCA there are 2 ways. One is visualization of concept lattices and second which were approximately the reduction of concept lattices on tree using conceptual indexes. These reduction methods based on conceptual indexes we tried to improve the use of combinations of these indexes. For all indexes definitions same assumption on context $c = (A,B,L,R)$ and concepts (X,Y), (Z,U) are applied.

4.2 Parent Selection Based on Stability

The stability index measures the proportion of subset of object of a given concept whose derivation is equal to the intent of this concept [20]. In other words, stability indicates the probability of preserving a concepts intent while removing some objects of its extent. A more stable concept is less dependent on individual members in the extension.

Definition of stability: *The stability index of concept (X,Y) is defined as follows:*

$$stability\,(X,Y) = \frac{|\{M \subseteq X | M' = Y\}|}{2^{|X|}} \tag{1}$$

4.3 Parent Selection Based on Support

Using support as criteria for selecting the parents has a tendency to group concepts that are part of the object.

Definition of support: *Let $Y \subseteq A$. The support count of the attribute set Y in C(A,B,L,R) is:*

$$support\,(Y) = \frac{\mathrm{Y}'}{\mathrm{A}} \tag{2}$$

4.4 Parent Selection Based on Similarity

This criterion is based on the clustering of parent and child concepts that share most of their attributes or objects. Parent and child having a large number of attributes that are grouped based on the principle of similarity clustering and local predictability.

Definition of similarity: *Let (X,Y) is a parent concept of its child concept (Z,U) in C(A,B,L,R). A shared attributes index of their connection edge E (Z,U) → (X,Y) is:*

$$similarity(E) = \frac{|\mathrm{Y} \cap \mathrm{U}|}{|\mathrm{B}|} \tag{3}$$

4.5 Parent Selection Based on Confidence

The value of concept for confidence estimates that a suitable object has an attribute A and attribute C also [21]. Other words, the criterion is trying to measure what is the strong implication of concept of parent and concept of child.

Definition of confidence: *Let (X, Y) is a parent concept of its child concept (Z,U) in C(A,B,L,R). The confidence of an edge E (Z,U) → (X,Y) is then:*

$$confidence(E) = \frac{|\mathrm{Z}|}{|\mathrm{X}|} \tag{4}$$

5 Experiments

In this chapter we present the results of reductions. We experimented with the selected dataset and generated data with specifying sparsity. The selected data consist of 101 objects (each object represents some animal) and 17 attributes (15 of

diamond type, 1 chain, 1 interval-based). Individual concepts are connected by edges. The number of collision represents the number, when the weight of edge is the same—occurs collision. In this case the program randomly selects the edge or by other criteria (combination of conceptual indexes). The number of top collision is the number, when the program puts into variable the highest weight edge of concept and compares it with the previous concept.

5.1 Reduction of the Selected Dataset

Reduction of concept lattice with the selected dataset (which consist of 101 objects, 17 attributes, 558 concepts and 1787 edges between concepts) based on each conceptual index deliver results that can be seen at Table 1.

From these results we can see that the best score has conceptual index of CONFDENCE, because number of collision and top collision are the smallest.

At Tables 2, 3, 4 and 5 can be seen reductions made from combinations of conceptual indexes.

If compare criteria Stability, Support and Similarity with the criterion of Confidence the best result have combination SIMILARITY-CONFIDENCE and if compare criteria with Similarity, Support and Stability the best result have combinations CONFIDENCE-SIMILARITY, CONFIDENCE-SUPPORT, CONFIDENCE-STABILITY.

Table 1 Results reduction of the selected dataset based on each conceptual index

Conceptual index	Number of edges	Number of collision	Number of top collision
Stability	557	1230	1230
Support	557	1230	1230
Similarity	557	583	462
Confidence	557	60	50

Table 2 Comparison conceptual indexes with conceptual index of CONFIDENCE

Conceptual index	Number of edges	Number of collision	Number of top collision
Stability	557	60	50
Support	557	60	50
Similarity	557	23	16

Table 3 Comparison conceptual indexes with conceptual index of SIMILARITY

Conceptual index	Number of edges	Number of collision	Number of top collision
Stability	557	583	462
Support	557	583	462
Confidence	557	22	10

Table 4 Comparison conceptual indexes with conceptual index of SUPPORT

Conceptual index	Number of edges	Number of collision	Number of top collision
Stability	557	1230	1230
Similarity	557	576	465
Confidence	557	64	48

Table 5 Comparison conceptual indexes with conceptual index of STABILITY

Conceptual index	Number of edges	Number of collision	Number of top collision
Support	557	1230	1230
Similarity	557	576	465
Confidence	557	64	48

From these results we can generally conclude that where is index of confidence the results are the best because number of collision and top collision are the lowest. The lowest number of collision and top collision have combination CONFIDENCE-SIMILARITY.

5.2 Reduction for Generated Data with Specifying Sparsity

With this reduction of concept lattices for generated data we added sparsity. Random data includes 10 object and 10 attributes while sparsity will change. In this case, the reduction will not be performed by combining conceptual indexes. Result can be seen at Table 6 (where concept lattice consist of 10 objects, 10 attributes, 189 concepts and 573 edges between concepts), Table 7 (where concept lattice consist of 10 objects, 10 attributes, 97 concepts and 244 edges between concepts), Table 8 (where concept lattice consist of 10 objects, 10 attributes, 40 concepts and 70 edges between concepts).

Of these reductions, where was introduces sparsity have found that the sparsity is larger, the size of concept lattices decreases and there are fewer collisions and top collisions. The different values have only index CONFIDENCE.

Table 6 Random data with specifying sparsity—sparsity 30

Conceptual index	Number of edges	Number of collision	Number of top collision
Stability	188	385	385
Support	188	385	385
Similarity	188	385	385
Confidence	188	343	342

Table 7 Random data with specifying sparsity—sparsity 50

Conceptual index	Number of edges	Number of collision	Number of top collision
Stability	96	148	148
Support	96	148	148
Similarity	96	148	148
Confidence	96	130	124

Table 8 Random data with specifying sparsity—sparsity 70

Conceptual index	Number of edges	Number of collision	Number of top collision
Stability	39	39	39
Support	39	39	39
Similarity	39	39	39
Confidence	39	29	29

6 Conclusion

From these experiments we have found that number of edges in each indexes are not changed. Considering each criterion separately in selected dataset, the best results has conceptual index of CONFIDENCE. The combination indexes CONFIDENCE-SIMILARITY have the lowest number of collisions and top collisions. With increasing sparsity is concept lattices diminish, reducing the number of edges, collisions and top collisions.

Based on results we can conclude that the best effect in the reduction of concept lattice has conceptual index of CONFIDENCE. On each dataset gave the index the best results with the lowest number of collisions and top collisions.

Acknowledgments The work presented in this paper was partially supported by the Slovak Cultural and Educational Grant Agency of Ministry of Education, Science, Research and Sport of the Slovak Republic (KEGA) under grant No. 025TUKE-4/2015 and also by the Slovak Grant Agency of Ministry of Education and Academy of Science of Slovak Republic (VEGA) under grant No. 1/0493/16.

References

1. Wille, R.: Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts. Springer, Netherlands (1982)
2. Birkhoff, G.: Lattice Theory. American Mathematical Soc. (1940)
3. Thomas, J., Cook, K.: Illuminating the path: research and development agenda for visual analytics. In: National Visualization and Analytics Ctr. (2005)
4. Keim, D.A.: Information visualization and visual data mining. IEEE Trans. Vis. Comput. Graph. **8**(1), 1–8 (2002)
5. Hamrouni, T., Yahia, S.B., Slimani, Y.: Avoiding the itemser clousure computation pitfall. In: Belohlavek, R., Snasel, V. (eds.) CLA (2005)

6. Hermann, M., Sertkaya, B.: On the complexity of computing generators of closed sets. In: ICFCA, LNAI 4933, pp. 158–168, Springer, Berlin (2008)
7. Belohlavek, R.: Introduction to Formal Concept Analysis. Palacky University, Department of Computer Science, Olomouc (2008)
8. Borchmann, D.: A Generalized Next-Closure Algorithm–Enumerating Semilattice Elements from a Generating Set. arXiv (2011)
9. Butka, P., Pocs, J.: Generalization of one-sided concept lattices. Comput. Informat. **32**(2), 355–370 (2013)
10. Butka, P., Pocs, J., Pocsova, J.: On equivalence of conceptual scaling and generalized one-sided concept lattices. Inform. Sci. **259**, 57–70 (2014)
11. Pocs, J., Pocsova, J.: Basic theorem as representation of heterogeneous concept lattices. Front. Comput. Sci. **9**(4), 636–642 (2015)
12. Pocs, J., Pocsova, J.: Bipolarized extension of heterogeneous concept lattices. Appl. Math. Sci. **8**(125–128), 6359–6365 (2014)
13. Butka, P., Pocs, J., Pocsová, J.: Reduction of concepts from generalized one-sided concept lattice based on subsets quality measure. Adv. Intell. Syst. Comput. **314**, 101–111 (2015)
14. Antoni, L., Krajci, S., Kridlo, O.: Randomized fuzzy formal contexts and relevance of one-sided concepts. In: LNAI (Subseries of LNCS) 9113, pp. 183–199 (2014)
15. Melo, C., Le-Grand, B., Aufaure, A.: Browsing large concept lattices through tree extraction and reduction methods. Int. J. Intell. Inf. Technol. (IJIIT) **9**(4), 16–34 (2013)
16. Pensa, R., Boulicaut, J.-F.: Towards fault-tolerant formal concept analysis. In: Proceedings of 9th Congress of the Italian Association for Artificial Intelligence, LNAI, pp. 212–223, Springer (2005)
17. Gajdos, P., Moravec, P., Snasel, V.: Concept lattice generation by singular value decomposition. In: Proceedings of CLA (2004)
18. Snasel, V., Polovincak, M., Abdulla, H.: Concept lattice reduction by singular value decomposition. In: Proceedings of the SYRCoDIS, Moscow, Russia (2007)
19. Ganter, B. Stumme, G. Wille, R.: Formal Concept Analysis: Foundations and Applications. Springer (2005)
20. Kuznetsov, S.O.: Stability as an estimate of the degree of substantiation of hypotheses derived on the basis of operational similarity. In: Automatic Documentation and Mathematical Linguistics (1990)
21. Ganter, B. Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer Science & Business Media (2012)

An Experiment with Facebook as an Advertising Channel for Books and Audiobooks

Grzegorz Chodak and Grażyna Suchacka

Abstract The paper addresses the problem of using social media to promote innovative products available in online stores. Motivated by the fast development of the audiobook market, on the one hand, and the efficiency of social media marketing, on the other hand, we conducted an experiment with a marketing campaign of books and audiobooks on the most popular social networking site, Facebook, and discussed it in the paper. The goal of the experiment was exploring possible differences in FB users' reaction to FB advertisements of traditional books and audiobooks. The experiment was implemented by using a real Facebook fanpage of a Polish publishing house having its own online bookstore. Results show some differences in Facebook users' reaction to the individual items' advertisements, however no significant differences in efficiency of the marketing campaign for books and for audiobooks were observed.

Keywords Audiobook · Social media marketing · Online marketing campaign · Facebook · Innovative products · Online bookstore

1 Introduction

The audiobook market is one of the most innovative branches of the book publishing market. Talking books have a long history but a very intensive market growth can be observed after the invention of CD MP3 format and development of technologies for distributing files on the Internet. Furthermore, the audiobook

G. Chodak (✉)
Department of Operational Research, Wroclaw University of Technology,
Wyb. Wyspiańskiego 27, 50-370 Wroclaw, Poland
e-mail: Grzegorz.Chodak@pwr.wroc.pl

G. Suchacka
Institute of Mathematics and Computer Science, Opole University, Ul. Oleska 48,
45-052 Opole, Poland
e-mail: gsuchacka@math.uni.opole.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*,
Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_18

market has a great potential. People spend more and more time in traffic jams and travel a lot. Popularization of sports like running, biking, riding on rollers, and so on, encourages to listening to the music and audiobooks, which enhances the audiobook market development possibilities. At the same time less and less (especially young) people read books—the percentage of people in Poland who did not read any book in 2015 was 63.1 % [1]. A hypothesis that some of these people listen to audiobooks is questionable but possible. On the other hand, it is known and confirmed by many studies that young people spend many hours every day using the Internet and interacting with friends through social networking sites, like Facebook (FB). That is why using Facebook by publishing houses as a medium to reach potential customers may provide them with additional outcomes.

The main purpose of our study was to conduct a comparative analysis of the effectiveness of advertising audiobooks and traditional books with Facebook and checking whether this most popular social medium is more suitable for innovative products, like audiobooks, than traditional products, like books. To achieve this goal we conducted an experiment using a real FB fanpage of a Polish publishing house which sells books and other products through its own online bookstore.

The structure of the article is the following. In Sect. 2 some information about social media marketing with a focus on Facebook marketing tools is presented. In Sect. 3 related work is briefly discussed and in Sect. 4 a history of audiobook-related technology development is outlined. A design of our experiment is talked over in Sect. 5 and experimental results are discussed in Sect. 6. At the end of the article some conclusions are presented.

2 Social Media Marketing

Social media are dedicated websites, such as forums, blogs, wikis, or social networking sites, through which users create online communities to interact, collaborate, and share contents [2, 3]. Examples of extremely popular social media include Facebook, Twitter, Google+, or Wikipedia. These applications are becoming an integral part of life for more and more people, especially the young ones.

Besides enriching individuals' private lives, social media have been successfully used to support business, especially in the area of social media marketing (SMM). Social networking sites are an essential marketing channel for competitive companies. Those managers who were skeptical about it a few years ago, now are becoming more open to this innovative form of marketing activities. Especially Facebook, with more than 1.5 billion active users [4], has proven to be a very effective platform to market products, promote brands and manage relationships with customers. "Facebook for business" offers a number of mechanisms supporting business: driving online sales, increasing local sales, promoting apps, raising brand awareness, generating leads [5].

The starting point for a company wanting to use Facebook in marketing is creating an FB fanpage, which may be then used to publish interesting contents and gain "fans". The basic tool for contacting with FB users is publishing on the fanpage so-called posts, containing text enriched with images, videos, and links to other websites. Posts may be targeted to specific groups of fans. Companies often offer promotions and discount to fans for "liking" their fanpage, post, or taking part in a competition. Other forms of promotion on Facebook include creating events, recommendations, or paid ads [6].

3 Related Work

Previous studies have shown that Facebook users have differentiated attitudes towards Facebook advertising and various factors may positively and negatively affect these attitudes. The finding of [7] have indicated a positive perception of fast food ads on Facebook by young Egyptians. Similarly, in [8] clearly positive attitudes of FB users' towards ads of Slovak companies were detected and the efficiency of social media advertising was shown to be even higher than efficiency of banner and contextual advertising in Slovakia. The study [9] has revealed that FB users do not actively use Facebook to find information on products and FB ads do not influence their purchase decision but the social media allows companies to significantly improve the communication with their customers.

In [10] the cluster analysis was applied to distinguish users' groups with different interest and attitudes towards FB advertising. Users from a negative group tended to block and avoid ads. Users from a neutral group did not blocked ads and liked participating in FB activities whereas users assigned to the third group revealed clearly positive attitudes towards FB ads. Various factors may influence FB advertising perception and usage. The study [11] has shown a significant positive impact of such factors as informativeness, entertainment, credibility, and personalization as well as a negative impact of message irritation.

Some studies examined the effectiveness of FB advertising. In [12] social advertising was found to be effective, mainly due to the ability of social networks to link together similar users. However, this efficiency is lower if ads contain the explicit information showing their promoting character. The results of [13] have shown that FB advertising has a significant impact on brand image and equity and that these two factors contribute to changes in Cypriot consumers' purchase intention.

The effectiveness of ads' content on user engagement in terms of the number of "likes" and comments was analyzed in [14]. Ads with persuasive content, e.g. appealing to emotions, were shown to increase users' engagement. Ads containing only informative content, like product prices, availability and features, were shown to reduce engagement but if they were combined with persuasive content, then users' engagement increased. In [15] online factors influencing users' attitude towards Facebook advertising were examined. Perceived interactivity, privacy, and

advertising avoidance were identified as significant factors whereas credibility was not.

To the best of our knowledge, in the literature there are no papers concerning the comparison of books' and audiobooks' advertisements in the Facebook environment. Therefore, our experiment fills this research gap.

It is obvious that the social media marketing does not work in all areas with the same strength but in the case of the publishing market it can be surely considered an effective way to reach potential customers. In Sects. 5 and 6 an experiment aiming at verifying this hypothesis is discussed.

4 Development of Audiobook-Related Technology

The history of audiobooks (or talking books) started in 1877 with the invention of a phonograph. The first sentence recorded with the phonograph by its inventor Thomas Edison was "Mary had a little lamb" [16].

A successor of the phonograph was a gramophone, invented in 1886 and developed over next few decades [17]. In the 30s, vinyl (originally known as vinylite) was introduced as a record material for radio transcription discs [18].

In 1931 American Foundation for the Blind started "Talking Book Program". Talking books were a chance for blind people to enhance the offer of Braille books. The only problem was the technique of recording books in those days, which enabled recording only short texts. The spoken audio has been available in public libraries and to a lesser extent in music shops since the 30s [19]. The first company which specialized in selling talking books was Caedmon Records, founded in 1952 [20].

In the United Kingdom on 7th November 1935 the Royal National Institute for the Blind (RNIB) delivered first talking books for blinds. Their first recordings were Agatha Christie's "The murder of Roger Ackroyd" and "Typhoon" by Joseph Conrad. The books were recorded on Long Play (LP) records which were 25 min per side. Discs could be played on a gramophone [21].

A magnetic tape was invented for recording sound by Fritz Pfleumer in 1928 in Germany. The Pfleumer's invention used a ferric oxide powder coating on a long strip of paper. This invention was further developed by the electronics company AEG, which manufactured recording machines and BASF, which manufactured tapes [22].

CDs (invented and popularized in the early 80s) firstly enabled saving in an audio format up to 74-min sound without compression [23] but the invention of MP3 standard greatly expanded possibilities of recording books. MP3 standard was a joint achievement of Fraunhofer Institute, University of Hannover, AT&T, Bell Labs, and Thomson-Brandt. It enabled recording 16 h of voice on one CD. This can be considered as the starting point for the growth of the audiobook market. The term "audiobook" was officially announced as the standard in 1994 by the Audio Publishers Association (APA). One may observe the digital renaissance of

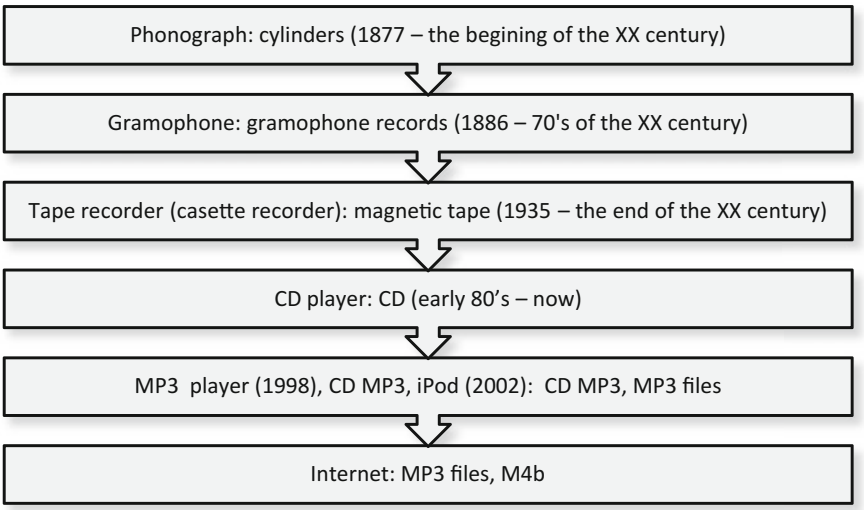

Fig. 1 History of audiobook-related devices and media [16–22, 25]

audiobooks over the next decades, when new portable delivery technologies and distribution channels have radically expanded both its use and its user groups. Unlike physical media, such as cassettes and CD-ROMs, which required transportation through the mail, digital libraries on the Web allow an instant download and access to the enormous number of audiobook titles [24, 25]. The history of devices which enabled listening to audiobooks and media on which audiobooks were stored is shown in Fig. 1.

An audiobook is an interesting hybrid medium, bringing together the form of a novel with new audio technologies, and as a hybrid it falls between the audio recording and print publishing industries [26].

The success of audiobook sales is dependent mainly on lector skills. Mainstream productions of audiobooks usually employ well-known movie stars to read books. In 2014, Bob and Debra Deyan of Deyan Audio opened the Deyan Institute of Vocal Artistry and Technology (DIVA), the world's first school for teaching the art and technology of audiobook production [27].

The future of audiobooks may be the possibility to choose the voice of a professional lector who reads a book without employing him/her. Nowadays there are two options: a book may be read by a human or by an automat (Text to Speech—TTS technology). Actually the quality of the automate reading is far from good actor's reading and especially the interpretation of text is still the future of TTS.

Phillips [26] claims that an audiobook offers a case study in the transition of cultural forms into new media. Facebook is the phenomenon of social media so the analysis of audiobook advertisement efficiency on a Facebook fanpage is an interesting combination of two cultural phenomena.

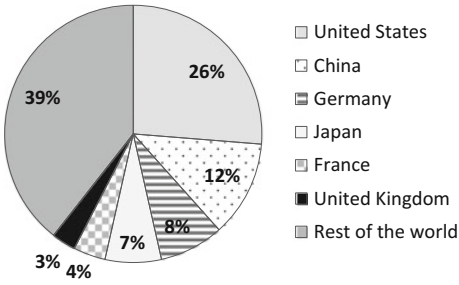

Fig. 2 The largest book publishing markets, by value (2013) [30]

The global audiobook industry is evaluated at 2.8 billion dollars. 43,000 new audiobooks were released in 2015 which is a slight increase from the 36,000 that came out in 2014 and a huge increase from 20,000 that were issued in 2013 [28].

A global book market value in 2013 was 114 billion euro [29]. Comparing values of the audiobook market and the whole book market it can be seen that talking books constitute only about 2 % of the whole book market. The dynamics of the audiobook market is much higher than that of the book market but even with such high dynamics it will take it decades to become a considerate and important piece of the book market. In Fig. 2 six largest book publishing markets are presented. This list is not surprising if we compare it to the list of countries with the biggest GDP.

5 Experimental Design

The purpose of the experiment was to examine possible differences between traditional books and audiobooks in FB users' reaction to social media advertisements. The experiment was realized in collaboration with a Polish publishing house and bookstore which sells books, multimedia, and films online (the name of the company is not given due to a non-disclosure agreement; in the paper we call it simply "the Bookstore").

The Bookstore has its Facebook fanpage, which at the time of conducting the experiment had 2910 fans. The company uses Facebook as one of two main advertising channels. FB posts are used as a mean to inform customers about recommended products and novelties available on the Bookstore site. Facebook is a good communication channel with actual and potential customers as it enables obtaining customers' opinions about products and learning their needs. The fanpage is also used to inform customers and potential customers about special promotions, discount coupons, and prize competitions.

The idea of the experiment was the following. From over three thousand products in the Bookstore database 20 books having their counterparts in audiobooks issued on CD were selected (20 titles corresponds to 40 items). The assumption was that the corresponding two items with the same title should differ

only in the publishing format (a paper book or an audiobook), without any additions. For each selected item a post was published on a Facebook fanpage of the Bookstore. Each post consisted of a little Bookstore logo, the Bookstore name, date of publishing the post, a picture of a book or audiobook cover, a short description and a link to the Bookstore Web page with a full item's description.

The only difference between two posts informing about two corresponding items (a book and an audiobook) for the same title was in a word "book" or "audiobook" in a short description of the item, and sometimes also in proportions and some details of the book cover picture.

The posts were published on Facebook in the following sequence: book, audiobook, book, audiobook… and after five titles (i.e., ten posts) the order was reversed to the sequence: audiobook, book… Such an order change was made after each ten posts. A rationale for such a way of the experiment implementation was to present to Facebook users almost the same products which differ only in the medium (CD or the paper).

The experiment started on 29 March 2016, when all the posts were published (from 3:00 to 6:00 pm) and it was finished on 06 April 2016, when results were read (from 8:00 to 10:00 pm).

Through the experiment we wanted to answer two questions:

1. Are there any differences in FB users' reaction to ads of audiobooks and books?
2. Is the strength of users' reaction proportional to market shares of audiobooks and books?

For each advertised title t (t = 1, 2, …, 20), i.e. for each two FB posts, the following measures were analyzed:

- S^t_{books}, $S^t_{audiobooks}$—the number of FB users who saw the post about the book and the audiobook, correspondingly, read from the FB data "*Views*" for each post;
- L^t_{books}, $L^t_{audiobooks}$—the number of FB users who saw the post about the book and the audiobook, respectively, and then reacted to it by liking, commenting, or sharing it, measured as the sum of FB data "*Likes*", "*Comments*", and "*Shares*" read for each post;
- C^t_{books}, $C^t_{audiobooks}$—the number of FB users who saw the post about the book and the audiobook, respectively, and then clicked one of its elements: the cover picture, the link in the text, the logo, or the author link, measured as the FB data "*Post Clicks*".

Furthermore, aggregated measures for all advertised books and for all advertised audiobooks were determined, e.g. the number of FB users' displays of all the 20 posts about the books was determined as:

$$S_{books} = \sum_{t=1}^{20} S^t_{books}.$$

To examine how the observed users' reactions to books' and audiobooks' ads as regards individual titles are related, we calculate the Pearson correlation: (1) for the number of the resulting likes/comments/shares, $r_{L_bL_a}$, (2) for the number of clicks, $r_{C_bC_a}$, and (3) for the aggregated number of likes/comments/shares and clicks, $r_{(L+C)_b(L+C)_a}$. For example, the Pearson correlation for the users' clicks is determined according to the formula:

$$r_{C_bC_a} = \frac{\sum_{t=1}^{20}\left(C^t_{books} - \overline{C_{books}}\right)\left(C^t_{audiobooks} - \overline{C_{audiobooks}}\right)}{\sqrt{\sum_{t=1}^{20}\left(C^t_{books} - \overline{C_{books}}\right)^2}\sqrt{\sum_{t=1}^{20}\left(C^t_{audiobooks} - \overline{C_{audiobooks}}\right)^2}}, \quad (1)$$

where $\overline{C_{books}} = \frac{1}{20}\sum_{t=1}^{20} L^t_{books}$, and $\overline{C_{audiobooks}} = \frac{1}{20}\sum_{t=1}^{20} L^t_{audiobooks}$. The coefficient is within the range $[-1, 1]$. The higher its absolute value is, the stronger the linear relationship between the variables occurs. The coefficient value equal to 0 means the lack of linear relationship between the variables.

6 Experimental Results

At the beginning, the volume of sales of books and audiobooks through the Bookstore website in the experiment period was roughly analyzed to give an idea of a sales scale. Sales data for the experiment period and for the entire previous year are shown in Table 1.

As it can be seen in Table 1, the sales ratio of audiobooks to books in 2015 was 0.21 and for the experiment period it was almost three times higher (0.62). It can be interpreted as a success of the Facebook marketing campaign, however one should notice that there are many additional factors affecting the volume of sales in the Bookstore, like seasonality, AdWords campaigns, and other external factors (e.g. the death of a well-known book's author in March 2016, which greatly increased sales of his books and audiobooks).

Results of the experiment are presented in Table 2, Figs. 3 and 4. Table 2 shows how many Facebook users saw the individual posts informing about the advertised items and how many viewers reacted to them. It can be seen that most popular products in categories of books and audiobooks are different. The mean number of

Table 1 Data on the sale of books and audiobooks in the Bookstore

Period	The number of books sold	The number of audiobooks sold	Sales ratio of audiobooks to books
2015	2425	520	0.21
29 March 2016–06 April 2016 (the experiment period)	69	43	0.62

Table 2 Results of the Facebook marketing campaign [the number of FB users]

Title no. (t)	S^t_{books}	$S^t_{audiobooks}$	L^t_{books}	$L^t_{audiobooks}$	C^t_{books}	$C^t_{audiobooks}$
1	173	327	0	4	4	7
2	263	228	3	1	4	1
3	285	212	3	0	7	1
4	560	319	8	1	9	1
5	597	461	12	6	12	9
6	319	322	3	0	2	8
7	526	406	15	6	5	12
8	413	353	8	3	4	2
9	408	162	6	0	6	0
10	244	145	3	0	1	0
11	327	155	6	0	8	1
12	219	201	1	2	1	2
13	227	1041	2	18	3	25
14	330	223	5	0	3	1
15	414	198	5	0	5	0
16	296	489	4	5	5	0
17	464	537	7	8	7	14
18	457	227	8	3	0	0
19	361	194	0	1	11	3
20	276	283	2	0	3	10
Sum	7159	6483	101	58	100	97
Mean	358.0	324.2	5.1	2.9	5.0	4.9
Std. dev.	118.7	204.0	3.9	4.3	3.2	6.5

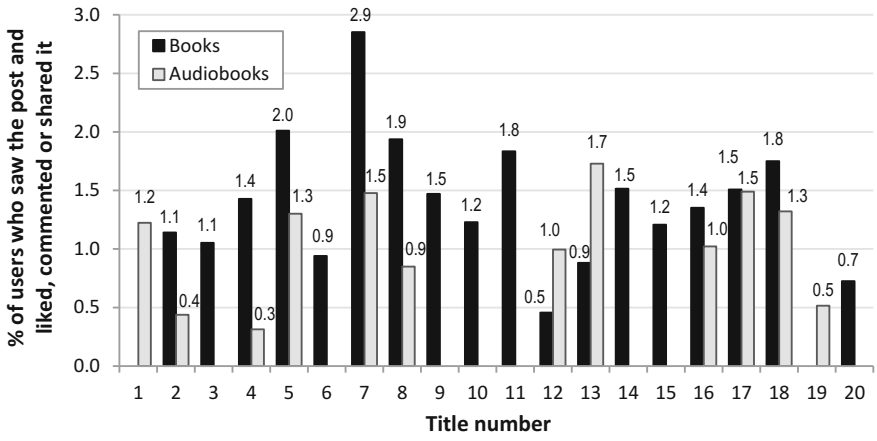

Fig. 3 Percentage of FB users who saw the posts and then liked, commented or shared them

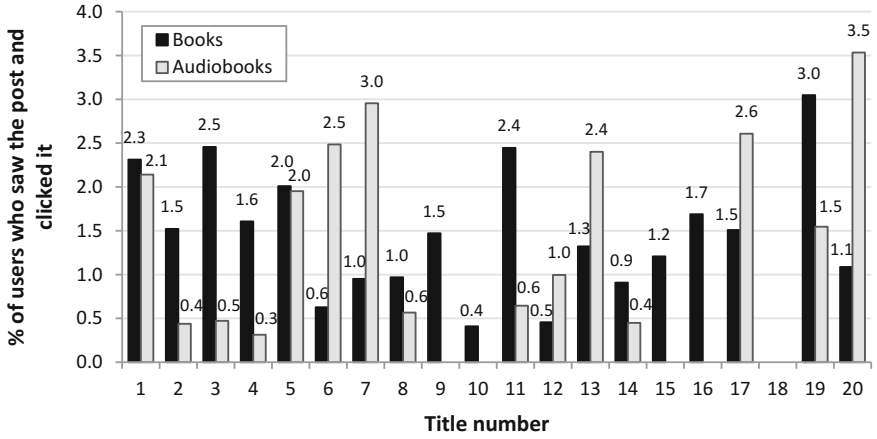

Fig. 4 Percentage of FB users who saw the posts and then clicked them

viewers for books (equal to 358.0) is only slightly higher than that for audiobooks (equal to 324.2)—it may be a little bit surprising if we compare it to the market shares of books and audiobooks. Standard deviation of the mean number of viewers is high both for books (118.7) and audiobooks (204.0).

Figure 3 shows how many Facebook users who saw the posts actively reacted to them by liking, commenting, or sharing them. In general, the reaction of users was limited both to the ads of books (1.4 % of viewers reacted) and audiobooks (0.9 % reactions), however users' reactions differ greatly for individual items in both categories. It is worth mentioning that in the case of books only two titles did not cause any users' activity and in the case of audiobooks it was eight titles.

Figure 4 presents percentages of customers who saw the individual posts and clicked either on the picture or the link referring to the Bookstore website. These measures reflect activity of users who were potentially interested in the advertised products. The average activity expressed with the percentage of user clicks was very low both for books (mean equal to 1.4 %) and audiobooks (mean equal to 1.5 %). The standard deviation was much higher for audiobooks (6.5) than for books (3.2).

Figure 5 presents the aggregated percentage results in terms of the number of users' reactions expressed as liking, commenting, sharing, or clicking the posts. Generally, there were 201 reactions to book ads (i.e. 2.8 % of all book ads' displays) and 155 reactions (2.4 %) to audiobook ads. One can see in Fig. 5 that the most popular products in both categories were different—only the titles numbered 5 and 7 were very popular both for books and audiobooks.

The Pearson correlation was very low both for likes ($r_{L_bL_a} = 0.16$) and clicks ($r_{C_bC_a} = 0.07$), and it was extremely low for the aggregated results ($r_{(L+C)_b(L+C)_a} = 0.01$). This means that users' reactions to ads of books and

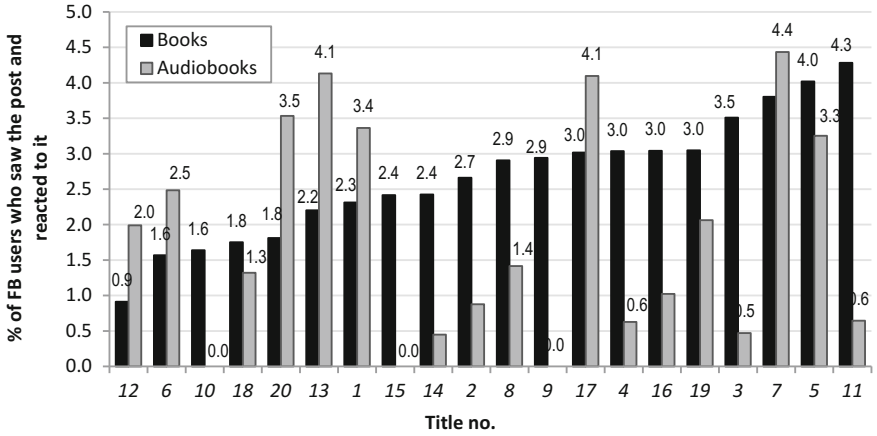

Fig. 5 Percentage of FB users who saw the posts and reacted to them—by liking, commenting, sharing, or clicking

audiobooks for individual titles are not correlated with each other, which is a rather unexpected result.

To sum up the results of the experiment, we can say that there were not significant differences in Facebook users' reaction to ads of audiobooks and books.

7 Conclusions

The paper discussed an experiment in which a real Facebook's fanpage of a Polish publishing house and bookstore was used to carry out a marketing campaign for books and audiobooks. Results of the experiment show that few viewers actively reacted to the ads and actions taken by them did not depend on the specific book titles. Generally one can say that the FB users' interest in the advertised books and audiobooks was similar—it is a promising result given that the audiobook market constitutes only about 2 % share of the whole book market.

An interesting result of the research is that the most popular products in the categories of books and audiobooks were different (only two of the twenty titles were comparatively popular in both categories). Moreover, the Pearson correlation calculated for customers' reaction to ads of books and audiobooks confirms that users reacted differently to different media with the same title. This suggests that besides the ad's content and title the kind of a medium (book or audiobook) matters for FB users.

Social media seem to be an appropriate tool for advertising innovative products like audiobooks, however a little bit higher reaction in terms of likes, comments, shares of users to book than audiobook ads (1.4 vs. 0.9 %, respectively) suggests that probably FB posts with audiobooks should contain another marketing message

than posts with books. One of possible solutions might be a presentation of multimedia files containing samples of audiobook reading, which would probably encourage more FB users to share the posts.

More experiments in real environment are needed to optimize the presented audiobook contents to encourage FB users to like, comment, or share posts. The possible continuation of our research is to conduct a comparative analysis of books' and audiobooks' ads effectiveness which will take into account not only organic results of post publishing on Facebook, but also paid ads, which usually attract a wider audience.

Acknowledgments This work was partially supported by the National Science Centre (NCN) in Poland under grant no. 2013/11/B/HS4/01061.

References

1. Nałęcz, M.: Poles do not read books. The alarming results of the National Library (the original title in Polish: Polacy nie czytają książek. Alarmujące wyniki badań Biblioteki Narodowej), Polska Metropolia Warszawska (2016)
2. Merriam-Webster Dictionary New Edition. Merriam-Webster, Inc., Springfield, MA (2016)
3. WhatIs.com. http://whatis.techtarget.com/definition/social-media
4. Helmrich, B.: Social Media for Business: 2016 Marketer's Guide, January (2016). http://www.businessnewsdaily.com/7832-social-media-for-business.html
5. Facebook for business. https://en-gb.facebook.com/business
6. Facebook in business (the original title in Polish: Facebook w biznesie). http://www.facebookdlabiznesu.pl
7. Gaber, H.R., Wright, L.T.: Fast-food advertising in social media. A case study on facebook in Egypt. JBRMR **9**(1), 52–63 (2014)
8. Vejačka, M.: Facebook advertising and its efficiency on the Slovak Market. E+M Ekonomie a Manage. **1**, 116–127 (2012)
9. Maurer, C., Wiegmann, R.: Effectiveness of Advertising on Social Network Sites. A Case Study on Facebook. Information and Communication Technologies in Tourism, Innsbruck, pp. 26–28, Austria, 2011
10. Lukka, V., James, P.T.: Attitudes toward facebook advertising. J. Manage. Mark. Res. **14**, 1–26 (2014)
11. Alsamydai, M.J., Khasawneh, M.H.: Antecedents and consequences of E-Jordanian consumer behaviour regarding facebook advertising. IJMBR **3**(4), 41–59 (2013)
12. Tucker, C.E: Social advertising: How advertising that explicitly promotes social influence can backfire. http://ssrn.com/abstract=1975897 (2012)
13. Dehghani, M., Tumer, M.: A research on effectiveness of facebook advertising on enhancing purchase intention of consumers. Comput. Hum. Behav. **49**, 597–600 (2015)
14. Lee, D., Hosanagar, K., Nair, H.S: The effect of advertising content on consumer engagement: evidence from facebook. http://ssrn.com/abstract=2290802 (2013)
15. Yaakop, A., Anuar, M.M., Omar, K.: Like it or not: issue of credibility in facebook advertising. Asian Soc. Sci. **9**(3), 154–163 (2013)
16. Stross, R.: Phonograph. the incredible talking machine. TIME June 23 (2010). http://content.time.com/time/specials/packages/article/0,28804,1999143_1999210_1999211,00.html
17. Newville, L.J.: Development of the phonograph at Alexander Graham Bell's volta laboratory: contributions from the museum of history and technology. United States National Museum Bull. **218**, 69–79 (1959)

18. V-Disc and Armed Forces Radio Catalogue, Blue Goose Publishers, St Louis. After: https://en.wikipedia.org/wiki/Gramophone_record
19. Rubery, M.: Introduction. Audiobooks, Literature, and Sound Studies, pp. 1–21. Routledge, New York, (2011)
20. Caedmon: Recreating the moment of inspiration. NPR Morning Edition, December 5, 2002. Retrieved: 6 Mar 2014
21. History of the Audio Book. http://www.silksoundbooks.com/history-of-audiobooks.html
22. Fenster, J.M.: How bing crosby brought you audiotape. invention & technology, fall (1994). After: https://en.wikipedia.org/wiki/Magnetic_tape
23. Rasen, E.: Compact discs: sound of the future. SPIN, http://www.spin.com/1985/05/compact-discs-sound-of-the-future
24. Have, I., Pedersen, B.S.: Sonic Mediatization of the book: affordances of the audiobook. MedieKultur **29**(54), 123–140 (2013)
25. Audiobook. https://en.wikipedia.org/wiki/Audiobook
26. Philips, D.: Talking Books. The encounter of literature and technology in the audio book. Convergence: The Int Journal of Res New Media Technol. **13**(3), 293–306 (2007)
27. Burkey, M.: Elevating the art of the audiobook: Deyan Institute of Voice Artistry & Technology. The booklist reader, 13 Mar 2014. http://www.booklistreader.com/2014/03/13/audiobooks/elevating-the-art-of-the-audiobook-deyan-institute-of-voice-artistry-technology
28. Kozlowski, M.: Global audiobook trends and statistics for 2016. Good E-Reader, 18 Dec 2015. http://goodereader.com/blog/interviews/global-audiobook-trends-and-statistics-for-2016
29. IPA's Global Publishing Statistics: Which Markets are Hot, Which are Not? The International Publishers Association, 13 Nov 2013. http://www.internationalpublishers.org/market-insights/data-and-statistics/155-ipa-s-global-publishing-statistics-which-markets-are-hot-which-are-not
30. Wischenbart, R.: Global trends in publishing 2014. http://www.pik.org.pl/upload/files/Global_Trends_in_Publishing_2014.pdf

Performance Analysis of SPA Web Systems

Wojciech Stępniak and Ziemowit Nowak

Abstract The effectiveness of methods accelerating the process of loading applications of SPA type, including the mechanisms offered by the HTTP/2 protocol, was analyzed. The performance of finished components supplied with AngularJS framework was researched. The study was conducted on a real SPA application. With the help of mini applications implementing the same functionality the performance of different designs offered by AngularJS was also compared. The study used Apache server, Chrome for Linux and Android and browser-perf and h2load tools.

Keywords Single page application · Performance · AngularJS · HTTP/2 · Browser-perf · H2load

1 Introduction

Single Page Application (SPA) is a web-based application that uses only one HTML page as the backbone for all application subpages. The interaction with the user of such an application is implemented using JavaScript, HTML and CSS. Unlike a traditional web application burdening primarily the server, SPA application works within the user's device. The characteristic feature of SPA application in relation to traditional web application is no need to reload the page in the course of the interaction. SPA applications resemble native software applications in their behavior, but do not have their own processes as they are started inside the browser process [1].

To create SPA applications developers use JavaScript frameworks based on the MVC pattern. As AngularJS is the undisputed leader among them, while studying

W. Stępniak (✉) · Z. Nowak
Wrocław University of Science and Technology, Wrocław, Poland
e-mail: ziemowit.nowak@pwr.edu.pl

Z. Nowak
e-mail: 192207@student.pwr.edu.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_19

the SPA application performance, the focus was on the mechanisms offered by this particular framework.

There are a lot of publications dealing with the issue of optimizing the loading time of traditional websites. From the book [2] it is evident that the subject has been studied for a relatively long time.

In 2015 Karabin and Nowak analyzed the performance frameworks AngularJS and Ember.js [3]. To the authors' knowledge, however, no studies have been performed of real downloading time of SPA application identifying what specific techniques, which do not require modifications to the existing source code, can accelerate the loading time of applications. It is even more difficult to find a work taking into account the new opportunities offered by HTTP/2, which is currently implemented in web servers and web browsers.

In the book [4], as well as on specialist blogs [5, 6], advice can be found on how to improve the performance of AngularJS application. Still, the authors of this chapter could not find any publications comparing the time of executing the commonly used ready-made AngularJS components, which are directives. The works usually focus on the examination of fragments of JavaScript code, rather than the directives which are provided by the AngularJS framework.

2 Methods to Optimize the Loading Time of SPA Applications

This section describes the selected approaches to accelerate the process of loading a website, in particular the SPA application.

Concatenation of resources. Concatenation is a technique of combining multiple CSS files, JavaScript into a single request. Reducing the total number of requests is the best method of optimization, regardless of the protocol or application used. However, if the reduction in the number of requests is not possible, then an alternative strategy to optimize HTTP/1.1 is joining multiple resources into one request [7]. In addition to the reduction in the number of tasks itself, the complete HTML document may be reduced due to the lower number of links to external resources.

Data compression. Data compression involves saving the original information using fewer bites. The size of text resources, such as HTML, CSS and JavaScript, can usually be reduced by 60–80 % using gzip compression. In contrast, images require more in-depth consideration [7]. For images there is lossy compression, which boils down to choosing the right format, determining the degree of compression, color palette and so on. Much depends on the subjective feeling of whether the image is still clear after applying compression.

Removal of unused CSS rules. If CSS styles delivered to the browser are not used by it, it means that the time needed for their transmission is an unnecessary expense. Furthermore, an increased number of CSS rules prolong the process of

developing the CSSOM model. Both the processes may slow launching a web application. There are tools to help detect unnecessary CSS rules. Such a tool is included in e.g. Chrome browser, prompting that a given CSS rule was not used on a specific page. There are also external tools such as grunt-uncss [8], which automatically, based on the static analysis of files, remove unnecessary rules from the CSS file.

Minification of .js files. The minification of files is a process of deleting unnecessary characters such as spaces, tabs, and blank lines, comments from CSS files or JavaScript—while maintaining the proper operation of the code. The recommended process before minification is to concatenate files of the same type. Optionally, the process also includes the replacement of variable names with their shorter forms [9].

Use of HTTP/2.0 protocol. As an optimization method the choice of protocol to transfer resources can also be given. The primary purpose of HTTP 2.0 is to reduce latency of applications by providing full multiplexing of tasks and responses, minimizing protocol overhead by using efficient compression of HTTP header fields, assigning priorities to tasks and pushing resources. Currently, HTTP 2.0 is at the stage of active formation [7].

The use of PUSH PROMISE mechanism. The PUSH PROMISE mechanism is to provide accompanying resources by HTTP/2 server in response to the browser. It means the possibility of sending additional resources through the server, before the browser requests them directly. Once connected, the client and server exchange SETTINGS frames, which may limit the maximum number of parallel streams in both directions. As a result, the customer can reduce the number of pushed streams or, setting them to zero, completely disable the pushing of resources [7].

3 AngularJS Constructions Affecting SPA Application Performance

In the books [4, 10] and [11] a few pitfalls associated with AngularJS performance were highlighted. It turns out that the conscious choice of an appropriate design, in particular the directive, can have a significant impact on the performance of SPA application in the browser environment.

Filtering collections. The filter included in the HTML code of the ng-repeat directive is started each time when a change has been noticed in the displayed collection. Perhaps in some cases, filtering data directly in the controller code will be a faster solution. The very filtering expression should be as simple as possible due to the potential ability to filter very large collections of elements.

Displaying collection items. Ng-repeat directive [12] creates an instance of a template for each element of the collection transferred as directive attribute. Each template instance is assigned its own range, where the loop variable received is set to the current collection element. It translates into the creation of a DOM element

for each of the items in the collection. When the collection is changed, those elements are removed from the DOM tree and re-created. Rendering a large number of items each time a collection is changed is a very expensive process. One solution to the problem is the use of *track-by expression*, which informs about the existing link between the DOM element and collection element. Thanks to this ng-repeat will not have to rebuild the already rendered DOM elements, even if the objects in the collection will be replaced with the new ones.

Using observers. Due to the time-consuming nature of performing $digest cycles, the book authors list the following advice, among others: observing the smallest model component possible, avoiding complicated observer expressions, limiting the number of observers. Unregistering unnecessary observers, when the range they observe is still needed and used by applications should also be considered.

One-time data binding. If in a given place of view there is a need to simultaneously display data after their downloading without changing their values in the DOM tree, one-time data binding introduced from version 1.3.x of AngularJS should then be used. "One-time expressions will stop recalculating once they are stable, which happens after the first digest if the expression result is a non-undefined value." [13] Together with the first data binding to view, the observers do not follow such an element any longer. Hence the subsequent $digest cycles do not have to check the bond again.

Hiding elements. There are many directives embedded in the AngularJS framework used to hide and show elements on the page, depending on a particular condition. Two of them, which are most commonly used, are ng-if and ng-show. These are the directives that belong to the category of modifying directives.

Directive ng-if [14] removes or re-creates a part of the DOM tree, based on the expression assigned to the directive as an attribute. If the result of the assigned expression is logical falsehood this element is removed from the DOM tree, otherwise a copy of the element is once again placed in the DOM tree.

Directive ng-show [15] shows or hides a given HTML element on the basis of the expression assigned as an ng-show directive attribute. The element is displayed or hidden by removing or adding a class .ng-hide on this element. .ng-hide class is predefined in the AngularJS framework and sets the display property CSS (display) to none value. If a given item is often hidden and then displayed, it is recommended to use the ng-show directive. If items from the DOM tree and observers associated with a given range must be removed, it is advisable to use ng-if [16].

Storing calculated function values. The technique referred to as memorizing is used to speed up applications for evoking functions that have already been evoked with a specified argument. It is expected that the result of the function will not be changed for a specific argument. This can be defined as the storage of already calculated function values for previously transferred arguments to the function [10].

Debounce ng-model. Debounce ng-model is a mechanism available in AngularJS from version 1.3. With it AngularJS can be told to wait for a certain period of time, not promoting changes from the view to the model. A situation when text is entered in the input control is a good example. Using debounce

ng-model loop shortcut activation can be delayed ($digest cycle) together with each character entered from the keyboard. Synchronization of data to the model may be delayed, e.g. by 500 ms, which will make the shortcut loop will run only once only after the text has been entered from the keyboard [17].

4 Research

The experiments were conducted using a computer (4 x i5-4210U 1.7 GHz) with a Linux (OpenSuse Tumbleweed 20160505). In order to conduct the research and collect the results, the computer had the following installed:

- Apache HTTP Server (2.4.20),
- Chrome web browser (51.0.2704.63),
- h2load tool (1.10.0),
- browser-perf tool (1.4.5),
- Selenium driver (2.53.0),
- Chrome Driver (2.21.371461).

At first, using h2load tool, the possibility of reducing the amount of data needed to start the application was examined, including the headers used by the HTTP protocol.

The other part of the research concerned the loading time of SPA application to the browser. It is the time it takes a browser to load an entire document with its associated resources from the moment of the first request, the so-called time to Load event occurrence—represented by the loadTime metric returned by the browser-perf tool. This experiment was carried out using Chrome browser (as described above) and Chrome on Android 5.0.1 (Huawei P8 Lite smartphone). Browser-perf obtains metrics from Selenium and Chrome Driver drivers.

For the purposes of researching the directives of AngularJS framework simple JavaScript scripts were written using browser-perf API. The scripts in an automated manner simulated the behavior of the user clicking the button actuating mechanisms related to the given directive after fully loading the application, and as a result of their actions returned JavaScript metric from Scripting group from Chrome time-line. The experiment was conducted only for Chrome for Linux.

Experiments related to the loading time of SPA application and the amount of transmitted data needed to load it, were conducted on a prototype of eKRK application (size: 4.37 MB) supporting the management of the studies curriculum according to the National Qualification Framework.

For the study of AngularJS directives for each of the experiment structures a separate mini-SPA application was built with a single controller using version 1.5.5 of AngularJS framework.

4.1 Description of Experiments of Loading Time of SPA Application

eKRK application was prepared in several versions:

Concatenation of resources—all JavaScript files needed to run the eKRK application were combined into one. In the same way CSS files were dealt with. For this task grunt-contrib-concat (1.0.1) of Grunt tool was used.

The removal of unnecessary CSS rules—based on the analysis of .html files with the help of grunt-uncss task (0.5.0) redundant rules were removed in the attached CSS file. Due to the rapidly emerging need for styles in the application, it was necessary to manually indicate rules which cannot be removed.

Minification of JavaScript code—already reduced concatenated resources with tasks grunt-contrib-uglify (0.2.4) by setting option *mangle: false.* Option *mangle: true* performs, in the course of code minification, the replacement of variable names to their shorter counterparts. Unfortunately, this process did not end up in the working code. This was due to the lack of maintenance of the right JavaScript code style by the developers.

For the purposes of research, an experimental version of the mod http2 module was used, available for Apache server version 2.4.20. The purpose of using this module was to check the effectiveness of the experimental implementation of HTTP/2 mechanisms in Apache 2.4.20 and supporting mechanisms in Chrome.

For each method configuration files for Apache were prepared, containing information on what .js and .css files are to be pushed with the index.html file.

The impact of gzip compression of HTML and CSS files was also examined for the time of application loading by the browser. For compression *mod deflate* Apache server module was used.

The research related to the amount needed to transfer the data was performed using the h2load tools. This tool was also used to gather information about the reported size of HTTP headers.

The experiment related to the loading time of the SPA application to the browser environment was carried out using browser-perf tool. Each such experiment was repeated ten times. The selection of such sample size was dictated by the compromise between the time of researching, and minimizing third factors. An interesting metric in this study, as returned by the browser-perf tool, was loadTime value expressed in milliseconds.

4.2 The Description of AngularJS Research Mechanisms

The way of data binding using ng-bind directive and {{expression}} was examined. The scenario involved displaying a certain number of elements, after their addition to the model, initiated by clicking the application button.

The time of the conditional display of elements using ng-if and ng-show was also examined. The scenario was to discover a certain number of text elements, after the user had clicked on the button.

Also the ng-repeat directive using the *track-by* construction and without it was checked. The scenario consisted of displaying a certain number of elements, and then after adding the same number to the collection, but of other objects and displaying all the elements.

During the study the focus was on the metric of JavaScript from Scripting group, i.e. the total execution time of JavaScript code. Each experiment was repeated ten times to minimize the impact of third factors. The selection of such sample size was dictated by the compromise between the time of experimenting, and minimizing third factors.

4.3 Findings

Figure 1 shows the relation between raw data size necessary to start the eKRK application and the method used to modify the data.

Figure 2 shows the relation between the size of the headers necessary during data transmission and the version of the protocol and the method used to modify the data. The techniques associated with minification and removing unnecessary CSS styles, following concatenation, do not affect the size of the headers, hence they were not included in Fig. 2. The results presented in both the charts can be regarded as deterministic as long as there are no repetitions associated with the packet loss. During the experiments on the local computer using the h2load tool there was no packet loss or transmission failure.

Figures 3 and 4 show the relation between the eKRK application loading time and the method used to modify those data and the version of the transmission protocol. Each experiment was repeated ten times. As already mentioned, the selection of such sample size was dictated by the compromise between the time of

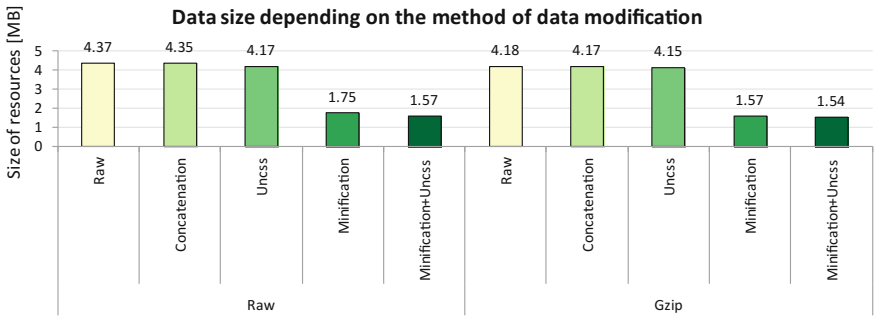

Fig. 1 Relation between raw data size necessary to start the eKRK application and the method used to modify the data

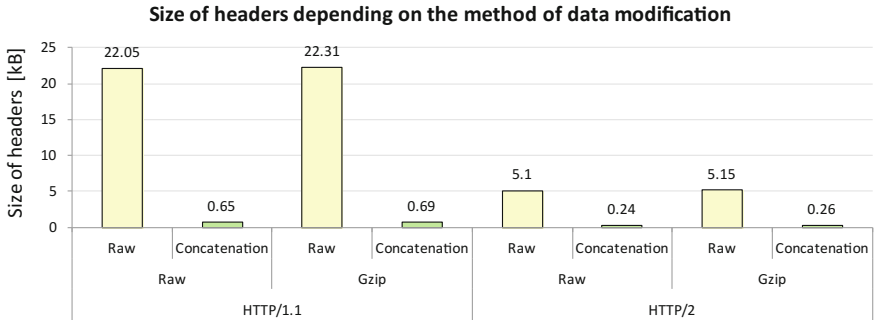

Fig. 2 Relation between the size of the headers necessary during eKRK data transmission and the version of the protocol and the method used to modify the data

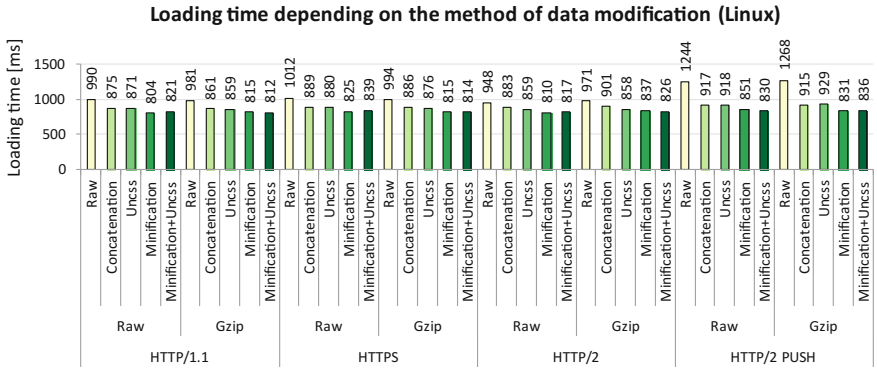

Fig. 3 Relation between the eKRK application loading time (Chrome on **Linux**) and the method used to modify those data and the version of the transmission protocol

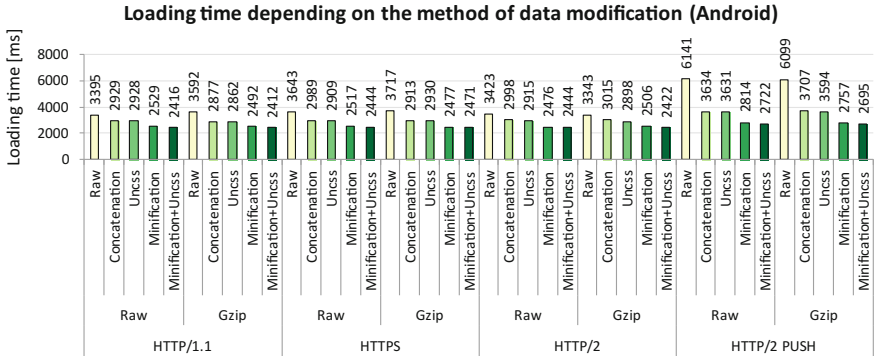

Fig. 4 Relation between the eKRK application loading time (Chrome on **Android**) and the method used to modify those data and the version of the transmission protocol

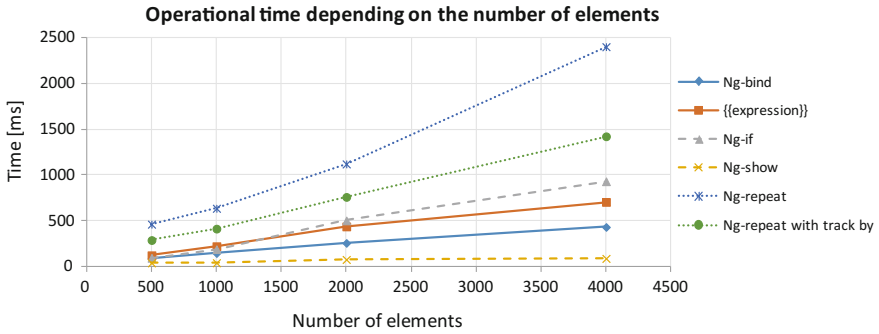

Fig. 5 Relation between operational time and number of elements for selected Angular JS constructions

experimenting, and minimizing third factors. With the statistical test by Shapiro-Wilk [18] the normal distribution of the obtained samples was investigated in environment R. For many samples the null hypothesis was rejected suggesting the normal distribution of the sample at the significance level of 0.05. Having to deal with the skewed distribution from the respective samples the median was determined.

Figure 5 presents a graph of:

- propagation time of changes from the model to the view based on a number of elements added to an empty collection using the ng-bind construction and {{expression}}
- ng-if and ng-show performance runtime depending on the discovery (presentation) of a certain number of elements that had previously been hidden,
- ng-repeat structure operational time using *track-by* and without it depending on the size of the collection.

Each result presented is the median of the 10-element sample with skewed distribution for which the variation coefficient does not exceed 3 %.

5 Discussion of the Results

5.1 Size of Transferred Resources and Headers

Analyzing the chart included in Fig. 1 it is clear that the greatest benefit was achieved as a result of minification of JavaScript files containing essential SPA application code. It is worth noting that in theory; even better results can be obtained by shortening variable names. The minification with shortening of names requires strict observance of rules imposed on JavaScript code and libraries from which it draws so that it could end up in success, i.e. running application. Examples

of such problems and suggested solutions are described in one of the tutorials for AngularJS [19]. Gzip compression, which was used for HTML and CSS files, also played an important role in reducing the data that needed to be transferred. It can be expected that the use of gzip compression for JavaScript files would also have had a similar effect, as confirmed by the observations of Matt Farina [20].

The positive results of concatenation, which should not affect the reduction of transmitted data, can be explained by a lower number of external attachments of CSS files in the HTML document itself. A similar explanation applies to references contained in HTML code to JavaScript files.

The total size of HTTP/1.1 and HTTP/2 protocol headers shown in Fig. 2 depends on the number of requests sent. The huge role played by the concatenation of resources can be seen. Techniques involving the minification and removing unnecessary CSS styles, after prior concatenation, did not affect the size of the headers; therefore they were not shown in the chart.

With the same number of requests the significant advantage of HTTP/2 protocol associated with the compression of headers can be seen. The size of headers sent in relation to the transferred resources is a few promiles, which can be considered negligible when transferring large files. It should be expected that in the case of downloading a large number of small files (less than one kilobyte) the application of protocol, which looks after the size of the headers will be important for the transmission rate, especially where there will be a long delay of network transmission.

5.2 Loading Time of SPA Application

The following observations relate to the research findings presented in Fig. 3. The PUSH PROMISE mechanism used to push all of the accompanying SPA application resources, contrary to expectations, did not help to shorten the loading time, despite the suggestion contained in [7]. The use of this mechanism much worsened application loading time. This is due to the fact that Chrome browser, although it noticed the PUSH PROMISE mechanism, requested the pushed resource again. This may result from the experimental phase of HTTP/2 implementation in Apache, or non-supporting it in the way expected by the browser. There are also voices suggesting that the mechanism is suitable for applications other than pushing related files, while its impact on the downloading time of resources in the implementation of H_2O web server is marginal [15]. According to Kazuho Oku, the benefits of PUSH PROMISE may be measurable only in the case of mobile networks [21].

The greatest benefits include reduced amount of transmitted JavaScript and CSS resources. No apparent difference was noted in the context of the loading time of SPA application between HTTP/1.1 and HTTP/2. No significant difference in the loading time of SPA application using gzip compression is noted. The authors

suppose that the greater advantage to use gzip compression can be achieved using cache of compressed resources within the Web server.

5.3 The Timing of the Construction of AngularJS Data Presentation

Based on the results shown in Fig. 4, it was observed that data binding using ng-bind directive brings better results than using {{expression}}, but the difference is noticeable only with a large number of bonds.

In turn, the time of performing the mechanisms related to the ng-show directive is up to ten times shorter than while using ng-if directive at 4000 components. As it can be seen, the manipulation of the CSSOM tree is much more efficient than the modification of the DOM tree.

Using the *track-by* construction in the ng-repeat directive can accelerate the execution of JavaScript code by over 40 % at 4000 elements.

5.4 Indications for Developers, Administrators, and Architects of Web System

Based on the study a list of recommended actions accelerating the process of loading the SPA application by the browser was formulated:

Minification of JavaScript code. The experiments revealed that the process of minification, regardless of the protocol used, was the most important factor in shortening the loading time of SPA application. If possible, an attempt should be made to shorten the names of variables during minification, bearing in mind, however, that it does not guarantee proper operation.

Removing unused CSS rules. A smaller number of CSS rules translate into the decrease in the amount of data sent over the network, and shorter construction of the CSSOM tree. The tools currently available, such as e.g. uncss, enable the automatic removal of unnecessary rules based on the static analysis of HTML files. Some CSS rules can be considered unnecessary by the tool. This is due to e.g. dynamic attachment of classes to DOM elements during the SPA runtime. Bearing in mind the imperfections of all the tools it should always be verified whether important CSS rules have not been removed.

Binding files. Regardless of the protocol used, concatenation of resources made a noticeable difference during the loading time of SPA application by the browser. While binding files, the relative paths occurring inside these files should be noted. They may need to be updated manually.

The above recommendations may be helpful to all SPA application developers, and those involved in their implementation.

6 Summary

The results of the study indicate that the technique accelerating the process of loading the SPA application the most is minification of JavaScript code, which significantly reduces the size of transferred resources. The removal of unnecessary CSS rules, despite the small difference in reducing CSS files, undoubtedly has a positive impact on accelerating the SPA application. A smaller number of CSS rules in the CSS file accelerate the construction of the CSSOM tree. The research revealed that the PUSH PROMISE mechanism of HTTP/2 protocol implemented in the mod http2 module of Apache web server slowed down the process of loading the SPA application and, at the present stage of its development, is not suitable to increase SPA application performance.

The only difference observed during experimenting, resulting from the use of HTTP/2 protocol, was header compression reducing their size at least fivefold. It can be expected that the greatest benefits of compressing headers in HTTP/2 will be visible in the transmission of small resources (<1 kB), whose size is not much bigger than that of HTTP/1.1 headers.

In the case of the directives provided with AngularJS framework, in the situation of frequent hiding and showing of elements, the ng-show directive should be used instead of ng-if. The duration time, in some of the cases studied, of the ng-show directive is up to ten times faster. In the publications containing recommendations regarding the performance of AngularJS the authors did not encounter an incentive to use ng-bind directive instead of {{expression}}. The research carried out showed that the use of ng-bind is a much more efficient way of data binding. In turn, the use of *track-by* construction for ng-repeat directive can not only shorten the duration of the JavaScript code, but also avoid re-rendering of DOM elements.

In view of the current, experimental implementation of HTTP/2 mechanisms by a few Web servers, it appears advisable to repeat the research, when the implementation is considered to be mature. A similar study should be carried out with regard to network delay, which could help determine the importance of individual techniques and protocol selection for mobile networks.

The team responsible for the AngularJS development, at the time of writing this chapter, is still working on the release of the official version 2, introducing new mechanisms. It would be interesting to examine the performance of the in-built constructions in the AngularJS 2.x framework after the release of the official version. Especially that Shawn McKay suggests that in terms of rendering AngularJS 2.0 framework in its alpha version is up to ten times faster than AngularJS 1.x [22].

References

1. Fink, G., Flatow, I.: Pro Single Page Application Development: Using Backbone.js and ASP. NET. Apress (2014)
2. King, A.: Speed Up Your Site: Web site Optimization. New Riders (2003)

3. Karabin, D., Nowak Z.: AngularJS vs. Ember.js—Performance Analysis Frameworks for SPA Web applications (in Polish). Polish Information Processing Society, Warsaw (2015)
4. Frisbie, M.: AngularJS Web Application Development Cookbook. Packt Publishing (2014)
5. Kras, A.: 11 Tips to improve AngularJS performance. http://www.alexkras.com/
6. Fundoo Solutions: Optimizing ng-repeat in AngularJS. http://www.befundoo.com/blog/
7. Grigorik, I.: High Performance Browser Networking: What Every Web Developer Should Know About Networking and Web Performance. O'Reilly Media (2013)
8. Osmani, A.: A grunt task for removing unused CSS from your projects. https://github.com/addyosmani/grunt-uncss
9. Bącik, Ł.: JS and CSS-Grunt files minification (in Polish). http://www.lukaszbacik.pl/blog-p53/
10. Lerner, A.: ng-book: The Complete Book on AngularJS. Fullstack.io (2013)
11. Seshadri, B., Green, S.: AngularJS. O'Reilly Media (2014)
12. AngularJS API reference: ngRepeat. https://docs.angularjs.org/api/ng
13. AngularJS Developer guide: Expressions, https://docs.angularjs.org/guide
14. AngularJS API reference: ngIf, https://docs.angularjs.org/api/ng
15. AngularJS API reference: ngShow. https://docs.angularjs.org/api/ng
16. Markiewicz, R.: Angular non-trivial performance hints. http://monterail.com/blog/2015/angular-non-trivial-performance-hints/
17. Ramesh, M.: Angular performance tips and tricks. http://blog.ideas2it.com/angular-performance-steps/
18. R Documentation: Shapiro-Wilk normality test. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/
19. Sevilleja, C.: Declaring AngularJS modules for minification. https://scotch.io/tutorials/
20. Farina, M.: Why minify JavaScript? https://engineeredweb.com/blog/why-minify-javascript/
21. Oku, K.: Performance improvements with HTTP/2 push and server-driven prioritization. http://blog.kazuhooku.com/2015/10/
22. McKay, S.: Comparing performance of Blaze, React, Angular-Meteor and Angular 2 with Meteor. http://info.meteor.com/blog/

Application of Spatial Econometrics Methods in the Analysis of WLAN Performance

Leszek Borzemski, Jakub Barański and Anna Kamińska-Chuchmała

Abstract This paper presents the spatial econometric modeling to performance prediction analysis of high-density client environments in higher education. According to our knowledge, these methods were not yet used in such analysis. Particular attention was devoted to SAR (Spatial Autoregressive Model) and SEM (Spatial Error Model) models, and their comparison with a classical non-spatial regression model. We have created models for two neighbor matrices to take into account different looks at distance definition in a 3D environment. The models were compared how well they predict the number of logged users which is considered as the WLAN performance index.

Keywords Wireless local area network · Performance models · Performance prediction · Spatial econometrics · Spatial autoregressive model · Spatial error model

1 Introduction

Econometrics traditionally refers to the statistical analysis of economic data of numerous economic phenomena [1–3]. Spatial econometrics (SE) covers econometrics by bearing in mind the possible effects related to the locations where data are collected. Econometrics and spatial econometric methods and models are employed not only in financial studies but in very different non-economic research and applications.

We apply methods of spatial econometrics methods to model network performance a wireless local area network (WLAN) at Wrocław Technical University

L. Borzemski (✉) · J. Barański · A. Kamińska-Chuchmała
Department of Computer Science, Faculty of Computer Science and Management, Wrocław University of Technology, Wrocław, Poland
e-mail: leszek.borzemski@pwr.edu.pl

A. Kamińska-Chuchmała
e-mail: anna.kaminska-chuchmala@pwr.edu.pl

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8_20

campus. Such spatial analysis may be important because regression models that exclude explicit specification of spatial effects, when they exist, can lead to inaccurate inferences about predictor variables.

In this paper, we are interested in the performance modeling of wireless computer local networks. Nowadays, Internet and systems that use it are already developed enough that an increasing number of users expect constant and reliable access to the network. More and more users have mobile devices equipped with wireless connectivity including wireless Local Area Networks (WLANs) which use the Wi-Fi technology with 2.4 and 5 gigahertz radio bands. Access to WLAN is via Access Points (APs) named hotspots. The users commonly want to have access to the Internet virtually all the time, especially when this is free access. APs can be installed inside and outside of buildings. Their locations are planned by practical experience, modeling using WiFi planner tools, site surveys, and spectrum analysis or satisfying other requirements and conditions. ASs can be independent of each other, or may be subject to central control. Usually, an AS has a range about 20 m indoors with things blocking or decreasing radio signals and a greater range outdoors. Moreover, in both configurations we usually may have multiple overlapping ASs areas but this overlapping is more considerate to the ultimate quality of WiFi access. The aim of this work is to do the performance analysis using spatial econometric methods in case of the academic wireless network named PWR-WiFi, at Wrocław University of Technology (WUT) campus. Specifically, we study the performance of indoor WiFi network installed in the B-4 building at WUT campus. Our network refers to high-density client environments. They have client densities greater than normal enterprise environments such as office environment.

Spatial econometric models are an extension of the classical econometric models. They may use GIS (Geographic Information System) data on the geographical location of the studied phenomenon. In these models the spatial effects: spatial dependence and spatial heterogeneity are included. It has been some 40 years [1] since first attempts at outlining the field of spatial econometrics (SE) were done. Nowadays SE plays a strong role in research and applications. The spatial pattern may result from a number of factors (e.g. economic, social, demographic, geographic, technical). Examples of spatial analysis are as follows: a spatial analysis of the socio-economic development of territorial units (e.g. unemployment, expenses, etc.), spatial price analysis (e.g. real estate), spatial analysis of the results of the baccalaureate. There are applications in crime analysis, health care, economic development process, and many others.

We also applied spatial econometrics to performance prediction problem in the case of Wide Area Network analysis and obtained satisfactory results [4, 5]. However, according to our knowledge, these methods were not yet used in performance prediction analysis of WLAN either in indoor or outdoor environments.

This contribution presents the application of spatial econometric models to performance prediction analysis of users logged to APs of the indoor WLAN located at the Wroclaw University of Technology at B-4 building. Data for this research has been collected from APs running under central controller in a campus computer network. The econometrics models were constructed and analyzed to

show their potentials in spatial performance analysis of indoor wireless local area networks. The analysis was performed using two spatial econometric methods, namely SAR (Spatial Autoregressive Model) and SEM (Spatial Error Model) methods, and compared to classical non-spatial regression.

The rest of the paper is structured in the following way. Firstly, we present the related work. Next, the experiment setup and data analysis are presented. Then, the spatial econometric analysis results are presented and discussed. Finally, the results are concluded.

2 Related Work

In this section, we describe the relevant work concerning the application of spatial econometrics in the analysis of the performance of computer networks, especially related to the important problem how to predict the system performance. The second our interest here is to show how other approaches may gain the knowledge and application to the performance prediction (forecast) problem as applied to the Web and Internet networks.

The performance analysis of computer networks has been the subject of several publications. For example, we can look at the performance prediction problem which can be related to various network characteristics, including RTT (Round Trip Time) and available file download throughput either at the TCP or HTTP layers for the Wide Area Networks (WANs). Our specific research direction in this regard is the use of these studies, methods and algorithms, whose primary use is in other areas of research. A machine learning approach to TCP throughput prediction is proposed in [6]. Yin et al. [7] presents highly accurate prediction models to provide accurate predictions for the optimal parallel stream number in Wide-area transfer of large data sets in high-bandwidth networks with speeds reaching 100 Gbps.

In our work, this problem has been studied using the data mining, and proposing a two-phase procedure based on clustering and classification algorithms to build a data throughput prediction [8–10]. We also fruitfully applied the transform regression and the Kolmogorov superposition theorem to TCP throughput prediction [11]. Next, our previous work dealt with the application of spatial-based methodologies originated from geostatistics: Turning Bands Method, Sequential Gaussian and kriging [12–15]. Recently we have added the spatial econometrics to WAN performance prediction [4, 5]. This paper is the first our contribution to performance prediction in wireless networks, specifically in interior WLANs. To the best of our knowledge, this work is the first such application of SE methods.

There are two main approaches dealing with performance problems in WLANs: (1) site survey based, (2) model (simulation) based. One of the most important problems considered in this area is the problem how to extend a network area by adding new APs, and how to plan their best placements. This problem considered in [16] is solved by employing a two-dimensional model of WiFi radio channels interference. Prediction of network performance of wireless mesh networks at the

level of individual flows in a small time granularity has been developed in [17]. The work proposes a measurement-based model-driven approach in which the actual network performance indexes (e.g. packet loss) were measured in current network conditions and infrastructure. These numbers were then used in a systematic modeling to predict the network performance.

The problem of planning and design of wireless network infrastructure was studied in [18]. The authors have proposed a rich set of features and analysis methods to inform building facility planning enabling studies of people's behavior in large building complexes utilizing solely measurements of WiFi signals from people's devices. The collection of large data sets of WiFi measurements was used not only to analyze the network's performance and usage but also to discover the density and flow of people within the building to better inform facility planning in large building complexes. Studies have been carried out in a large hospital complex with an area of more than 10 hectares. The data collected was from two weeks, during which approximately 18,000 different devices were registered.

In [19] the author analyzed the data from the same campus network, which is studied in this work but measurement data was taken in a different tracing period. Measurement data was collected from 12 indoor APs of PWR-WiFi. The aggregated number of users connected to all APs is presented in Fig. 1. Measurements were done every 5 min throughout the day during the period from 23 December 2014 to February 28, 2015. The academic community is a very particular environment for the use of a wireless network. At night, the campus is empty, as confirmed in Fig. 1, while the primary network traffic takes place from 6:00 to 21:00. Measurements have shown the activity through the working days, while at night and on weekends the

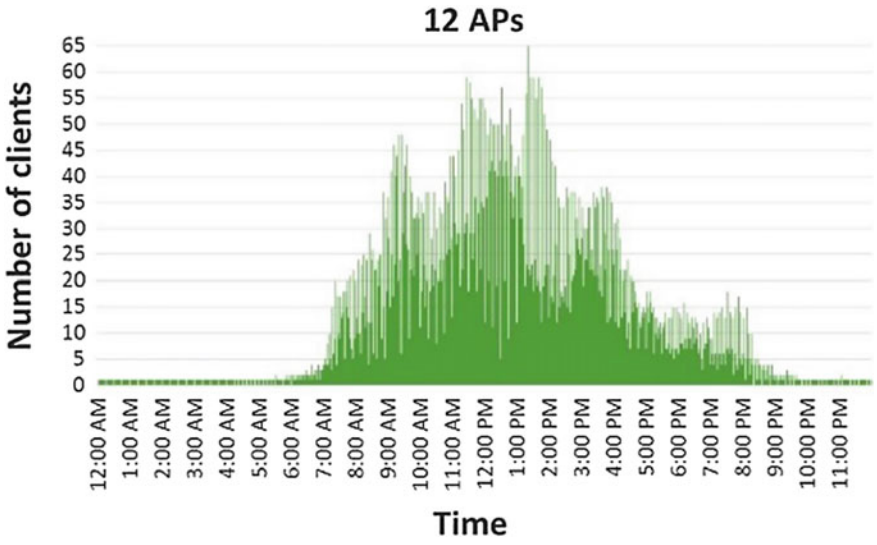

Fig. 1 The aggregated number of users connected to PWR-WiFi at Bldg. B4 on Thursday 29th January 2015. *Source* [16]

number of connected clients drops to zero. The measurements demonstrated that network capacity is sufficient. The limit of connected users is set to 100.

Paper [4] uses econometric models to predict time to download files to a host in TUW campus from different web servers in Europe. The same file was downloaded from 4 to 24 October 2013. During this period, the file has been downloaded each day four times at selected hours: 6:00, 12:00, 18:00 and 0:00. To predict the file downloading time three econometric approaches were used, namely linear regression, SAR, and SEM. In [5] we compared the econometric methods with the geostatistical methods and argued that presented econometrics methods to prediction Web server performance could be helpful in the spatial analysis of Internet characteristics.

3 Experiment Setup and Data Analysis

The data set used in the article was collected in the period 05.05.2014–01.06.2014, when it coincided in a given month, the largest number of working days and student activities. In order to reduce errors in the estimation of models, the days free from classes were eliminated. We collected information how many devices were connected to all ASs in the building at the given measurement time. These numbers were stored in Cisco wireless controller log file. The measurements were held in hour intervals between 7:00 and 21:00. This measurement time range was chosen because at night the classes do not take place and the number of connected users drops to zero. Building B4 has five floors and several rooms for laboratories, classes and lectures, including two big lecture halls with about 200 seats each. Space is used by two faculties which are among the most popular (overcrowded) faculties at the Wrocław University of Technology (WUT)—and with students having daily contact with modern computer techniques, fixed and mobile because they come from the computer science and mechanical engineering faculties. The room space is usually fully loaded all the working day, but in particular between 9:00 and 17:00 h. There are 11 Cisco's Aironet APs in the B4 building that together with other Aironet APs in the campus are under the central control of Cisco Wireless LAN Controller (WLC). WLC monitors and control access to WLAN and collects exploitation and maintaining data. The Lightweight Access Point Protocol (LWAPP) is the principal protocol used in Cisco's centralized WLAN architecture. It is used for the configuration and management of WLAN, tunneling WLAN client traffic to and from a centralized WLAN controller (WLC) within Ethernet frames UDP packets. It is needed to say that in building B4 there are also autonomously working APs that can be connected to PWR-WiFi and Eduroam WLANs. The users logged to these APs are not taken into account.

Figure 2 shows the locations of investigated APs in the coordinate system XYZ, whereas the APs' coordinates are given in Table 1. Data are expressed in a local coordinate system XYZ with an origin (0, 0)—it is aligned with a known real-world point at the campus. The local coordinate system is expressed in meters.

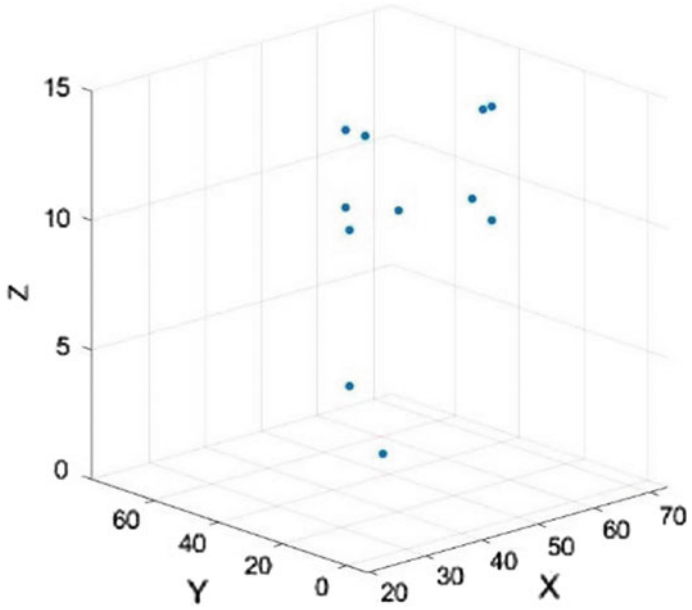

Fig. 2 Aerial view of APs in the coordinate system XYZ

Table 1 The coordinates XYZ of APs in B4 building

AP	X (m)	Y (m)	Z (m)
AP1-1	33	12	3
AP2-2	27	12	6
AP3-3	55.5	63	9
AP4-3	66	36	9
AP5-4	48	10.5	12
AP6-4	27	12	12
AP7-4	37.5	15	12
AP8-4	55.5	63	12
AP9-4	55.5	57	12
AP10-5	54	15	15
AP11-5	52.5	15	15

Three full-time hours were selected for the analysis in this work: 10:00, 12:00, 16:00. They correspond to three times of day that are the interesting for us: morning, midday and afternoon may characterize representative measurement data points. At these times there are the breaks between classes and lectures at which we have found by personal observation the students' habits that they usually massively use their smartphones, tablets, and laptops on the hallways and in the rooms. Therefore, these hours can be used to in the analysis of academic usage a network. The basic statistics of the data is presented in Table 2.

Table 2 Basic statistics of the total number of active connections

Hour	Minimum X_{min}	Maximum X_{max}	Average $\overline{X}$	Standard deviation S	Variation coefficient V (%)	Skewness coefficient G	Kurtosis coefficient K
7	0	5	0.66	0.99	1.50	1.83	3.51
8	0	30	4.93	5.83	1.18	1.79	3.29
9	0	42	7.31	7.26	0.99	1.67	3.48
10	0	61	12.35	12.71	1.03	1.61	2.41
11	0	52	11.77	10.61	0.90	1.14	0.77
12	0	72	13.10	13.62	1.04	1.62	2.52
13	0	53	11.62	10.98	0.95	1.26	0.99
14	0	70	12.34	13.06	1.06	1.59	2.42
15	0	45	8.99	9.00	1.00	1.54	2.31
16	0	56	10.46	11.78	1.13	1.72	2.67
17	0	43	6.76	7.60	1.12	1.73	3.33
18	0	40	6.67	7.98	1.20	1.53	1.97
19	0	30	3.80	5.50	1.45	2.18	5.02
20	0	20	2.40	3.96	1.65	2.24	4.84
21	0	20	2.40	3.96	1.65	2.24	4.84

Analyzing the data in Table 2, it could be claimed that there is a large difference between minimum and maximum values, which indicates the dispersion in the data.

4 Spatial Econometrics Analysis Results

4.1 Preliminaries

In econometrics, the basic regression modeling is a standard linear regression named Ordinary Least Squares (OLS). OLS purpose is to find a linear relationship between a dependent variable y and a set of explanatory variables X, $y = X\beta + \varepsilon$. The OLS estimates model's coefficient β by minimizing the sum of squared prediction errors, hence, least squares. Certain assumptions are made to determine β: the random error ε has a normal distribution with a mean zero, the random error terms are uncorrelated and have a constant variance. In the spatial analysis, in particular, in practical cases, these assumptions may not be always satisfied due to a spatial dependence (SD). SD dependence exists when a value observed in one location depends on the values observed at neighboring locations. These circumstances may impact the variables and error terms. The basic reason which is worth to research is that the SD exists due to important location and distance characteristics as well as some known or unknown spill-over effects.

When incorporating the spatial pattern in the analysis, one can use a *spatial autoregressive model* (SAR) and *spatial error model* (SEM). In the literature, we can find other spatial econometric models belonging to the families of SAR-type, SEM-type, or hybrid models. The fundamental difference between SAR and SEM models lies in the way spatial lags are used to reflect the spatial dependencies. They are based on (1) *spatial lag*—the dependent variable y in location i is affected by the independent variables in both locations, i and j, and (2) *spatial error*—the error terms across different spatial units are correlated. Spatial lag may model a diffusion problem—events in one location predict an increased likelihood of similar events at neighboring locations. The spatial error is indicative of omitted (spatially correlated) covariates that if left unattended would affect inference.

SAR model for the $n \times 1$ cross-sectional observations y is given by $y = \rho W_n y + X\beta + \varepsilon$, $\varepsilon \sim N[0, \sigma^2 I_n]$, ρ is autoregressive parameter. SEM model we define: $y = X\beta + \xi$, with $\xi = \lambda W_n \xi + \varepsilon$, where $\varepsilon \sim N[0, \sigma^2 I_n]$, λ is a parameter of model. In the SAR model, a lag of the dependent variable is included as an additional regressor. In the SEM model, a spatial lag of disturbances of the non-spatial model is used to account for spatial dependencies in the error structure. Accounting the importance of location and spatial dependency in the regression model determines the definition of the binary *neighborhood matrix R*. This matrix is used to bind a variable in one point of observation space for that variable elsewhere in the system. This is a square symmetric $R \times R$ matrix with (i, j) element equal to 1 if points i and j are neighbors of one another (or more generally, are spatially related),

and zero otherwise. By convention, the diagonal elements of this "spatial neighbors" matrix are set to zero. This matrix does not cover the distance between points. Therefore, to include the information about distances between points, we need to construct more powerful spatial neighborliness. This is done through the W matrix usually referred to as *spatial weights matrix*. Creating a distance one can assume the distance threshold d, and next if the actual distance between locations i and j is greater than d then the matrix element w_{ij} is set to 1, otherwise 0. In addition to specifying the adjacency also, it is needed to define the interaction in space.

In order to bring the variables to the mutual comparability, one should normalize the matrix. The most common transformation is called "row-standardization" in which the rows of the neighbor matrix are made to sum to unity. The individual elements of the normalized matrix of weights w_{ij}^* are calculated: $w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}}$. Then SAR model for the r-th observation point and K its neighbors takes the form: $y_r = \rho\left(\sum_{s=1}^{N} w_{rs} y_s\right) + \sum_{i=1}^{K} x_{ri}\beta_i + \varepsilon_r$. The SEM model can be rewritten in the similar way.

In this paper, the spatial analysis was made using GeoDa, the software developed at the Software of the Center for Spatial Data Science [20].

4.2 Neighbor Matrix

The first Law of Spatial Econometrics sounds: "*Everything is related to everything else, but near things are more related than distant things*." According to this law, we must specify what it means that the objects are closer or further. To determine this, we must define how the neighborhood and distance are measured. Such neighbor matrix is assumed a priori thus the quality of calculations in spatial analysis greatly depends on how it is defined. We define two neighbor matrices, namely the neighbor matrix W_d and W_k. W_d uses chordal distance measurement—this can be used at least for points within about thirty degrees of each other, whereas W_k uses the Euclidean distance for k nearest neighbors (we assume $k = 3$). Table 3 presents the AP's neighbors for both neighbor matrices W_d ($d^\circ = 30$) and W_k ($k = 3$).

4.3 Global Moran's I Statistics

Global Moran's I statistics (index) is a measure of spatial autocorrelation. It indicates not only the existence of spatial autocorrelation but also the degree of spatial autocorrelation. In the absence of spatial autocorrelation, the value of Moran's I index takes values close to 0. In the case $I < 0$, negative autocorrelation occurs, while in the case $I > 0$ there is a positive autocorrelation. The graphic

Table 3 AP's neighbors

AP	W_d (d° = 30)	W_k (k = 3)
	AP's neighbors	AP's neighbors
AP1-1	AP2-2, AP6-4, AP7-4	AP2-2, AP5-4, AP6-4, AP7-4, AP10-5, AP11-5
AP2-2	AP1-1, AP6-4, AP7-4	AP1-1, AP5-4, AP6-4, AP7-4, AP10-5, AP11-5
AP3-3	AP4-3, AP8-4, AP9-4	AP4-3, AP8-4, AP9-4
AP4-3	AP9-4, AP10-5, AP11-5	AP3-3, AP8-4, AP9-4, AP10-5, AP11-5
AP5-4	AP7-4, AP10-5, AP11-5	AP1-1, AP2-2, AP6-4, AP7-4, AP10-5, AP11-5
AP6-4	AP1-1, AP2-2, AP7-4	AP1-1, AP2-2, AP5-4, AP7-4, AP10-5, AP11-5
AP7-4	AP3-3, AP4-3, AP9-5	AP1-1, AP2-2, AP5-4, AP6-4, AP10-5, AP11-5
AP8-4	AP3-3, AP4-3, AP9-5	AP3-4, AP4-4, AP9-4
AP9-4	AP3-3, AP4-3, AP8-4	AP3-4, AP4-4, AP8-4
AP10-5	AP5-4, AP7-4, AP11-5	AP1-1, AP2-2, AP4-4, AP5-4, AP6-4, AP7-4, AP11-5
AP11-5	AP5-4, AP7-5, AP10-5	AP1-1, AP2-2, AP4-4, AP5-4, AP6-4, AP7-4, AP10-5

representation of spatial autocorrelation is given in a Moran's scatterplot. Scatter charts are calculated for the specific neighbor matrices—in our case, there are W_d and W_k matrices. The results depending on the neighbor matrix should be different. Where the results are similar regardless of the neighbor matrix, this shows the robustness of the results on the choice of the matrix. The charts with the global Moran's indexes for both neighbor matrices at 10:00, 12:00 and 16:00 h are presented in Fig. 3. Table 4 contains a summary of global Moran's indexes.

Moran's I index shows that the choice of the neighbor matrix is important - depending on the choice of the definition of the neighbor matrix, the Moran's index takes different values. Only in the case of the hour 12:00, the indexes are similar, which may attest to the robustness of the choice of neighbor matrix W. It should be noted that for the remaining hours the values of the Moran's statistics vary depending on W definition.

4.4 Model Comparison

The number of users in the network was measured in the period from 5.05.2014 to 1.06.2014 that is through of 27 days. Response (predictor) variable in the models will be the number of users on the t-th day, while the explanatory variables will be the number of users on (t-1)-th, (t-2)-th,…, n-th day, where $n < 12$. We treat data as time series. To estimate models, the working days were selected in which classes were held. They are the following: 5.04–9.04, 12.04–16.04, 19.04–23.04 and 26.04–30.04.2014.

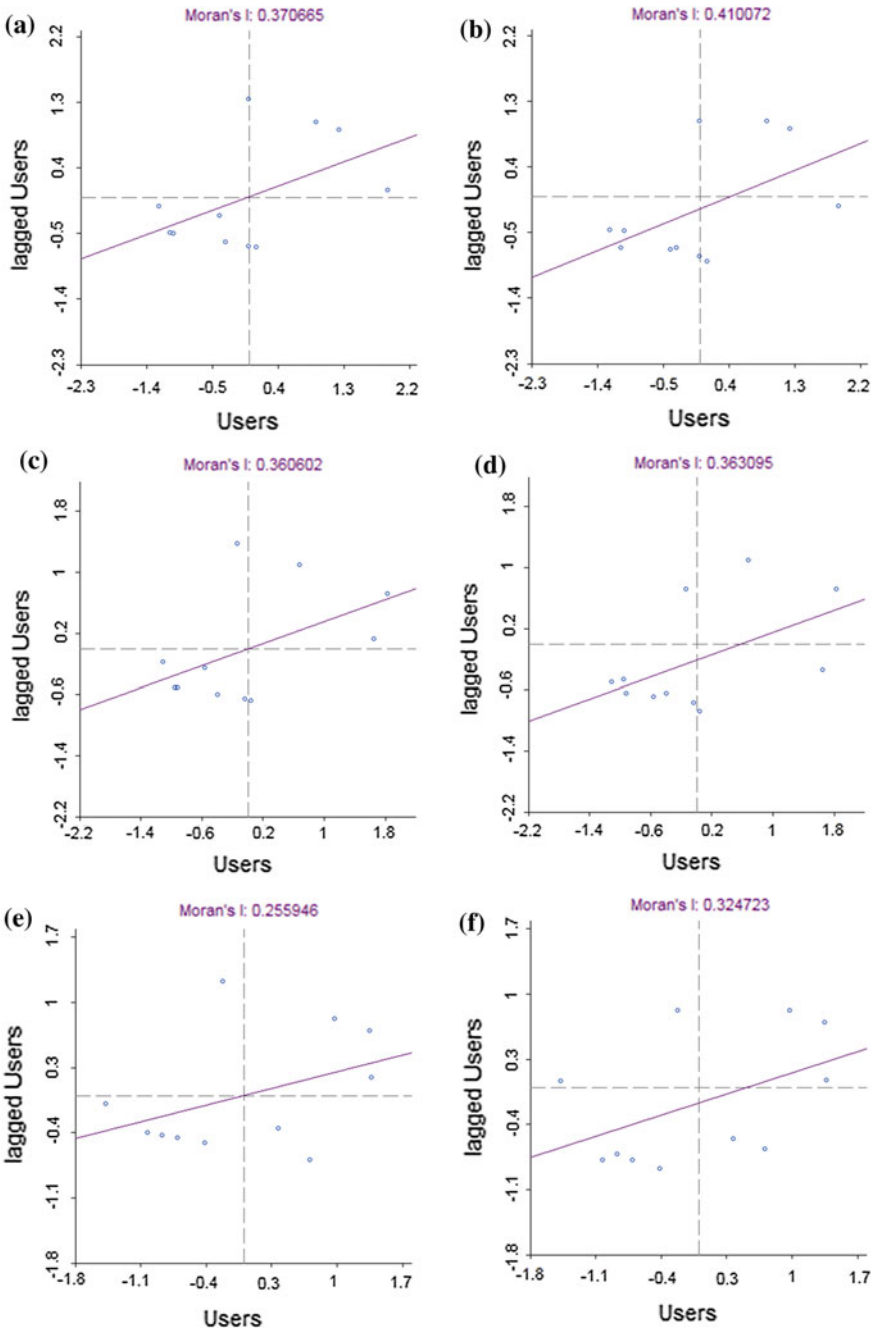

Fig. 3 Global Moran's *I* statistics charts at: 10:00—(**a**), (**b**), 12:00—(**c**), (**d**), 16:00—(**e**), (**f**), for W_d (charts on the *right*) and W_k (charts on the *right*)

Table 4 A summary of global Moran's statistics

	Moran's I index	
	Neighbor matrix	
	W_d	W_k
10:00	0.370665	0.410072
12:00	0.360602	0.363095
16:00	0.255946	0.324723

Table 5 The prediction error for neighbor matrix W_d

Hour	Model	Minimum $P_{\min}$	Maximum $P_{\max}$	The average $\overline{P}$
10:00	Linear regression	0.03	13.82	5.93
	SAR	0.08	14.54	6.24
	SEM	0.03	13.59	6.13
12:00	Linear regression	0.43	13.27	6.72
	SAR	1.08	14.13	7.27
	SEM	0.80	12.65	6.97
16:00	Linear regression	1.12	8.79	5.39
	SAR	0.95	9.03	5.44
	SEM	0.85	9.18	5.43

Table 6 The prediction error for neighbor matrix W_k

Hour	Model	Minimum $P_{\min}$	Maximum $P_{\max}$	The average $\overline{P}$
10:00	Linear regression	0.03	13.82	5.93
	SAR	2.60	14.09	7.51
	SEM	0.53	13.61	6.41
12:00	Linear regression	0.43	13.27	6.72
	SAR	0.59	15.50	10.59
	SEM	1.29	11.67	7.30
16:00	Linear regression	1.12	8.79	5.39
	SAR	0.66	9.16	5.65
	SEM	0.39	9.93	5.55

We compared the real number of users with the number that is predicted by using the developed econometric models. In the Tables 5 and 6, the predicted variable error P is presented for the hours that have been selected for the tests, and for appropriate neighbor matrix W_d (Table 5) and W_k (Table 6). The average of the real numbers of logged users was: at 10:00–12.35, at 12:00–13.10, and at 12:00–10.46 users.

The values of error for spatial models using matrix W_k are not much smaller than in the case of models for the matrix neighbor matrix W_d. Only in the case of SAR model for 12:00 h the difference is significant. The average error for the number of users decreased from 10.59 to 7.27. In the case of models created for 16:00 h, the average error is very similar in each of the three models, for both neighbor matrices. Errors for remaining hours differ from each other, the smallest error reaches a linear regression model, but it is not the spatial model. The average error in predictions between tested methods was no greater than 2 users for the average number of users about 10 to 14, excluding SAR model at 12:00 for neighbor matrix W_k.

5 Conclusion

We analyzed WLAN performance using spatial econometric methods. The application of SE in the field of WLANs is motivated by the questions how and where WLANs are built, how they work and how they are used. One of the questions is the question about indoor WLAN performance. Our work deals with the specific performance analysis, namely the performance prediction. The number of logged users defines the considered performance index. To our knowledge, it is the first such application of SE methodology in the literature.

SE proposes different models that in various ways take into account a spatial aspect. We used two of them, namely SAR and SAM models. We also compared the results of the WLAN performance prediction with a non-spatial linear regression model.

We have created models for two neighbor matrices to take into account different looks at distance definition in a 3D environment. The models were compared how well they predict the performance. The linear regression model achieved the overall best results taking into account the average prediction error $\overline{P}$, but the other models were not so far from the results of OLS. The difference is no more than 2 users, excluding SAR model at 12:00 for neighbor matrix W_k.

The prediction errors can be caused by different obstacles and interferences in the indoor infrastructure within buildings what may have an impact on signal spreading. Signal sent from the AP is affected by walls and other items located in the building. Hence, the application of SE in the outdoor WLAN could bring a better result, especially in the case of open space available to transmit radio waves. The AP's that are running in considered WLAN are equipped with the directional antennas, which are properly set up so that the signal is mainly propagated only in a certain direction where it is desired. Econometric models do not take into account such disorders. It is likely to reduce the errors if data was collected more frequently than every one hour. Next, it is possible that more grained monitoring would result with collecting better information about users' activities. This discussion formulates further research in employing SE methods in WLAN performance analysis.

References

1. Anselin, L.: Thirty years of spatial econometrics. Papers in Regional Science **89**(1), 3–25 (2010)
2. Anselin, L., Rey, S.J.: Spatial econometrics in an age of cyber GIScience. Int. J. Geogr. Inf. Sci. **26**(12), 2211–2226 (2012)
3. Arbia, G., Espa, G., Giuliani, D.: Dirty spatial econometrics. Ann. Reg. Sci. **56**(1), 177–189 (2016)
4. Borzemski, L., Kamińska-Chuchmała, A.: Web server's performance prediction with using spatial econometric methods. Rynek Energii **3**(112), 120–124 (2014)
5. Borzemski, L. Kamińska-Chuchmała, A.: Distributed web server's data performance processing with application of spatial econometrics models, A. Grzech et al. (eds.), Information Systems Architecture and Technology: AISC vol. 430, pp. 37–48, (2016)
6. Mirza, M., Sommers, J., Barford, P., Zhu, X.: A machine learning approach to TCP throughput prediction. IEEE/ACM Trans. Networking **18**(4), 1026–1039 (2010)
7. Yin, D., Yildirim, E., Kulasekaran, S., Ross, B., Kosar, T.: A data throughput prediction and optimization service for widely distributed many-task computing. IEEE Trans. Parallel Distrib. Syst. **22**(6), 899–909 (2011)
8. Borzemski, L.: The use data mining to predict web performance. Cyber. Syst. **37**(6), 587–608 (2006)
9. Borzemski, L.: Internet path behavior prediction via data mining: conceptual framework and case study. J. Univ. Comp. Sci. **13**(2), 287–316 (2007)
10. Borzemski, L., Kliber, M., Nowak, Z.: Using data mining algorithms in web performance prediction. Cyber. Syst. **40**(2), 176–187 (2009)
11. Borzemski, L., Starczewski, G.: Application of transfer regression to TCP throughput prediction. First Asian Conference on Intelligent Information and Database Systems (ACIIDS), IEEE Computer Society. 28–33 (2009)
12. Borzemski, L., Kamińska-Chuchmała, A.: Client-perceived web performance knowledge discovery through turning bands method. Cybern. Syst. Int. J. **43**(4), 354–368 (2012)
13. Borzemski, L., Kamińska-Chuchmała, A.: Spatio-temporal web performance forecasting with Sequential Gaussian Simulation method, Communications in Computer and Information Science, vol. 291, pp. 111–119. Springer, Berlin (2012)
14. Borzemski, L., Kamińska-Chuchmała, A.: Web performance forecasting with kriging method. In: Contemporary Challenges and Solutions in Applied Artificial Intelligence Studies in Computational Intelligence, vol. 489, pp. 149–154. Springer, Berlin (2013)
15. Borzemski, L., Kamińska-Chuchmała, A.: Distributed web systems performance forecasting using turning bands method. IEEE Trans. Industr. Inf. **9**(1), 254–261 (2013)
16. Gal, Z., Balla, T., Karsai, A. Sz.: On the WiFi interference analysis based on sensor network measurements, In: IEEE 11th International Symposium on Intelligent Systems and Informatics (SISY 2013) Conference Proceedings, Subotica, Serbia, 26–28 September (2013)
17. Yu, H., Zeng, K., Mohapatra, P.: Measurement-Based Short-Term Performance Prediction in Wireless Mesh Networks, Proc. of 20th International Computer Communications and Networks (ICCCN), pp. 1–6, (2011)
18. Prentow, T. P., Ruiz-Ruiz, A.J. Blunck, H. Stisen, A. Kjaegaard M.B.: Spatio-temporal facility utilization analysis from the exhaustive WiFi monitoring, Pervasive and Mobile Computing, pp. 305–316, (2015)
19. Kamińska-Chuchmała, A.: Performance analysis of access points of university wireless network. Rynek Energii **1**(122), 122–124 (2016)
20. GeoDa Center for Geospatial Analysis and Computation. https://geodacenter.asu.edu

Author Index

L. Borzemski et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part I*, Advances in Intelligent Systems and Computing 521, DOI 10.1007/978-3-319-46583-8