Biblioteka Informatyki
Szkół Wyższych

Information Systems Architecture and Technology

System Analysis Approach to the Design, Control and Decision Support

Library of Informatics of University Level Schools

Series of editions under the auspices
of the Ministry of Science and Higher Education

The ISAT series is devoted to the publication of original research books in the areas of contemporary computer and management sciences. Its aim is to show research progress and efficiently disseminate current results in these fields in a commonly edited printed form. The topical scope of ISAT spans the wide spectrum of informatics and management systems problems from fundamental theoretical topics to the fresh and new coming issues and applications introducing future research and development challenges.

The Library is a sequel to the series of books including Multidisciplinary Digital Systems, Techniques and Methods of Distributed Data Processing, as well as Problems of Designing, Implementation and Exploitation of Data Bases from 1986 to 1990.

Wrocław University of Technology

Information Systems Architecture and Technology

System Analysis Approach to the Design, Control and Decision Support

Editors
Jerzy Świątek
Leszek Borzemski
Adam Grzech
Zofia Wilimowska

Wrocław 2014

Publication partly supported by

Faculty of Computer Science and Management
Wrocław University of Technology

Project editor
Arkadiusz GÓRSKI

The book has been printed in the camera ready form

OFICYNA WYDAWNICZA POLITECHNIKI WROCŁAWSKIEJ
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław
http://www.oficwyd.pwr.wroc.pl;
e-mail: oficwyd@pwr.wroc.pl
zamawianie.ksiazek@pwr.wroc.pl

ISBN 978-83-7493-857-0

CONTENTS

PART 3. IMAGE PROCESSING AND PATTERN RECOGNITION

PART 4. EXPERT AND SOFTWARE SYSTEMS DESIGN

INTRODUCTION

All project, decisions and control algorithms are based on the knowledge about the plant under investigation. Systems approach to the computer aided design, control and decision support requires model of the investigated process. That's why models are so important in systems research. Investigation of object of the deferent nature (technical, economical, biomedical or computational) gives us many notifications about observed processes. Based on the collected knowledge, about investigated process the model of observed reality is proposed. The mathematical model gives precise plant description. Usually the relation between values characterizing process is given.

System analysis gives us the proper tools to create further decision about investigated plant based on the collected knowledge, and consequently based on the elaborated model. Base on the model the optimization, control and management task may be formulated. Base on the knowledge about the process the prediction or diagnosis may be proposed.

The above mentioned applications of different type tasks we can recognize in selected and revived chapters which have been divided into the following groups:

Part 1. *Business Process Optimization Platform in the Integrated Information Systems*
Part 2. *Knowledge Engineering and its Application in Decision Support Systems*
Part 3. *Image Processing and Pattern Recognition*
Part 4. *Expert and Software Systems Design*

The book provides an interesting representation of research in the area of system analysis in decision aided problems in proposed groups.

PART 1. BUSINESS PROCESS OPTIMIZATION PLATFORM IN THE INTEGRATED INFORMATION SYSTEMS

The Chapter 1 presents application of Business Process Optimization Methodology to select and develop proper analysis, planning and optimization methods for resource management in the information systems, dedicated to the domain of transport and logistics. The chapter addresses an innovative approach to the identification of decision making problems in transport companies, the description of these problems with the domain ontologies and the development of the formal models leading to the dedicated algorithms which are delivered as complex Web services and may be directly integrated into software products. In such a way we achieve an agile approach to the development of service-based software systems with the reusable components (services).

In the Chapter 2 the problem of asking relevant questions about the features describing concepts of interest basing on incomplete data. The main goal of this chapter is to present the novel approach that make use of Restricted Boltzmann Machines to deal with the issue of eliminating missing values o attributes. The proposed model makes use of reconstruction abilities of Restricted Boltzmann Machines to sample the most probable value to be imputed for each of the missing cases. The entire research is supported by preliminary experimental studies that examine the quality of the proposed approach comparing to the reference technique.

The main tasks of planning road transport include route planning, scheduling of drivers and scheduling of vehicles. Organization of transport involves addressing all these decision-making tasks simultaneously, in order to obtain the greatest possible profits from the operation of a transport company. However, attempting joint approach to these problems, causes that the models are too complex. In consequence the well-known algorithms become not useable and the new solutions are hard to be obtained. Thus, the most commonly used decomposition of the overall problem into a number of simpler optimization tasks. In the Chapter 3 the possibility of decomposing the problem into smaller is indicated. The exemplary models and algorithms solutions for some simple tasks are introduced.

In the Chapter 4 an outline of the Algorithm Composition Engine for solving optimization problems for domain-specific information systems is designed. The main goal of Algorithm Composition Engine is to provide a unified framework of description and implementation of optimization problems and their corresponding data structures and algorithms, to enable their reuse in building domain-specific information systems and to gather domain-knowledge in an ontology. The framework of composition of algorithms provides patterns of description and implementation of optimization methods and techniques, so that they can be hybridized into compound algorithm to collaboratively solve complex and hard optimization problems.

The Chapter 5 presents the main steps for the agile method of software development for resources management optimization for transportation companies. In transportation, the basic set of decision concerns schedules and routes. There are also other types of decisions as decisions concerning technical overviews, fleet modernization, etc. Very often these decisions are elaborated by an expert. However due to the scale of the problem and the dynamics of the business environment it becomes more and more hard to select the best or almost the best variant. This is why modern transportation companies require efficient decision supporting systems.

PART 2. KNOWLEDGE ENGINEERING AND ITS APPLICATION IN DECISION SUPPORT SYSTEMS

The aim of the Chapter 6 is to present the Constraint Programming modeling framework providing a methodology for Constraint Programming-based decision support systems' design allowing one to answer whether a given production order specified by its cost and completion time can be accepted in a given manufacturing system given by available production capability and/or what manufacturing systems capability guarantee completion of a given production order under assumed cost and time constraints.

In the Chapter 7 high frequency data processing and the use of complex event platform in combination with business rules approach are discussed. For such a high volume of data, it is suitable to use Complex Event Platform (CEP), because CEP allows for big data processing in real time. The improvement of decision making process under the condition of dynamical adaptation of the process on the fly is considered. The pattern recognition for detecting and predicting the trends in data by mining this information from historical data is used. After diagnosis the set of business rules are build according to which the process runs and control the process flow by defining the restrictions.

The Chapter 8 presents the project management problem, in which for the given project deadline and budget, under two-level uncertainty concerning duration of operations, the goal is to determine budget allocation maximizing certainty of the successful project completion, is considered. It is assumed that upper boundary time-resource function is given for each operation, with a fuzzy parameter inside. Such two-level uncertainty is recommended for use in real-life situations in which just rough estimations of operation durations are available. For typical upper boundary functions the optimization problem is nonlinear and non-convex, which disables application of classical optimization algorithms. This chapter presents a method of transforming the problem to the form allowing application of linear programming. Two ways of linearization are introduced and evaluated.

Effectiveness of company's management could be verified in various ways. In this context mainly profitability, market position and other factors are analyzed. In the Chapter 9 effectiveness of management is analyzed from the company's market value maximization point of view. Main goal of this chapter is to verify, whether choice of enterprise's legal form has got an influence on effectiveness of its assets management.

In the Chapter 10, attempt is made to verify if a purchase strategy which uses utility function to define the contribution of forward and spot contracts can effectively manage the purchase process on the Polish market. The changes in the Polish regulations large consumers can be participants in the energy market. Currently, according to the chosen strategy, it is possible to purchase energy within a tariff system from any trading company or directly from the market through energy exchange. In order to fully take advantage of the present conditions, consumers who decide to actively participate in market have to seek purchase strategies which on the one hand reduce cost and at the same time limit risks. The simulated results presented confirm the usefulness of the proposed methodology.

The Chapter 11 presents, for the building materials warehouse managers, the practical application of methods of mathematical programming as tools for computer-aided purchases organization. The problem is illustrated example of optimizing purchases for small warehouse.

PART 3. IMAGE PROCESSING AND PATTERN RECOGNITION

The Chapter 12 presents the results the possibilities of digital Hilbert optics applications to amplitude-phase images of dynamic scenes objects. The effectiveness is evaluated with digital Hilbert optics-images correlations and its' run-rotation movement dependences. As the main method for identification the maxima-correlation analysis into amplitude-phase images of anisotropic Hilbert transformed complex shape objects is used.

In the Chapter 13 the research area of data mining and its applications in medicine is described. The origins of data mining and its crucial features are shortly presented. The specificity of medicine as an application area for computer systems is discussed. Characteristic features of the medical data are investigated. Common problems in the area are also presented as well the strengths and capabilities of the data mining methods. Finally a set of modern applications of data mining designed for clinical use, are described and compared.

A novel hybrid algorithm for recognition of early stage of portal hypertensive gastropathy is proposed in the Chapter 14. First image preprocessing is described. Then disease symptoms characteristics are presented and hybrid algorithm scheme combining edge detection, Local Binary Patterns and local maxima clustering is shown. Finally the detailed description of these methods are provided. The parameters of the algorithm are also described with ranges used in tests and their best values (obtained empirically) are presented. The proposed algorithm is tested and compared to a few other algorithms showing it's comparable in terms of effectiveness in general case and a bit better than other ones in recognition of early stage of portal hypertensive gastropathy.

In the Chapter 15 endoscopic video analysis problems and artificial intelligence algorithms supporting diagnosis is discussed. One of the main problems of training and testing of those algorithms is that there is no simple and universal way of choosing most well trained algorithm after performing cross-validation, which is present in almost every artificial intelligence system. In this chapter, a method resolving this problem (at some circumstances) is proposed and examined in the task of recognizing cancer, healthy tissue, blurred frames and sharp frames on endoscopic videos by two exemplary artificial intelligence algorithms designed for this task, using neural networks and support vector machines. The results show that proposed method gives a little better results than the average algorithm after cross-validation.

In the Chapter 16 a problem of splitting data for *k*-fold cross-validation, where class proportions must be preserved, with additional constraint that data is divided into groups that cannot be split into different cross-validation sets is discussed. This problem often occurs in e.g. medical data processing, where data samples from one patient must be included in the same cross-validation set. As this problem is NP-complete, a heuristic anytime polynomial algorithm is proposed and described in the chapter. Also, it is experimentally compared to two other, simpler algorithms.

PART 4. EXPERT AND SOFTWARE SYSTEMS DESIGN

In the Chapter 17 the development and implementation of the concept of Virtual Engineering Object is described. A Virtual Engineering Object is a computerized real world representation of an engineering object. Virtual Engineering Object will act as a living representation of the object capable of adding, storing, improving and sharing knowledge through experience, in a way similar to an expert of that object. In this chapter, it is shown through test models how the concept of Virtual Engineering Object can be implanted with the Set of Experience Knowledge Structure and Decisional DNA. A test case study for three different drilling machines, drilling tools and the working holding devices is developed, to test and demonstrate the implementation of Virtual Engineering Object.

The aim of the Chapter 18 is to focus on visualization and practical aspect of presenting gathered knowledge in modern focused on end user way. The main goal of knowledge visualization is presentation purposes. It could be done in many ways using plain text techniques or graphical once for presenting facts. The knowledge visualization could also be a tool for data analysis. Depending on analysis type: explorative or confirmative. Explorative analysis seeks data presentation that allows finding hypotheses about the data. Confirmative analysis aims for confirmation or rejection of the hypotheses.

Passive testing is said to be uncontrolled. Although the obvious element of control (applying stimuli) is indeed absent, we argue that passive testing retains various other elements of control. They can fail, leading to false verdicts. Rather than proposing more robust ways of exercising control, we proceed by identifying the tacit elements of control and removing them altogether (so that they can no longer fail), at the cost of employing more complex testing algorithms. We argue that one of such elements of control is influencing, or even only assuming the characteristics of a test arrangement, and that relaxing control is in line with the current state of development of ICT systems. We further concentrate on the control assumption concerned with the placement of a passive tester within a distributed, asynchronous, message-passing system, in which messages in communication links experience delays. The Chapter 19 serves as a case study that illustrates a more general approach towards control.

In the Chapter 20 some properties of BPMN diagram are examine. The consistency, i.e. possibility to identify errors in requirements at the early stage of the development process and challenge due to a semi-formal nature of BPMN diagrams is verified. Moreover it is consider whether the BPMN diagram enables simultaneous modelling of the functionality, of the structure and of the behaviour of the target system model. Those properties should enable to develop consistent and complete models and thus to generate automatically complete workflow applications without any manual programming.

The Chapter 21 presents an algorithmic approach to improve robustness of linear-quadratic (LQ) optimal control strategies acting in network environment is presented. While providing smooth and efficient performance under nominal operating conditions, optimal controllers show sensitivity to parametric uncertainties and perturbations. In network control systems the situation aggravates due to the delay in feedback information exchange among the system elements and possibility of data loss. By using additional control information, the proposed algorithm increases the stability margin without downgrading dynamical properties of LQ optimal control solution. The robustness to networked-induced uncertainties is evaluated experimentally.

However there exist many systems that allow Users to find the route between given points on the map, their specific preferences are usually not taken into account. In the Chapter 22 an efficient way to handle multiple criteria with the use of the genetic algorithm that takes into account the reduction of the search space is presented. Conducted research shown that proposed algorithm performs better than algorithms known from the literature.

Wrocław, September 2014

Jerzy Świątek

PART 1

BUSINESS PROCESS OPTIMIZATION PLATFORM IN THE INTEGRATED INFORMATION SYSTEMS

0business process optimization, service-oriented systems
business process optimization methodology,
business problem optimization platform.

Krzysztof BRZOSTOWSKI*, Dariusz GĄSIOR*, Adam GRZECH*,
Krzysztof JUSZCZYSZYN*, Grzegorz KOŁACZEK*, Andrzej KOZIK**,
Radosław RUDEK***, Arkadiusz SŁAWEK*, Leopold SZCZUROWSKI*,
Paweł ŚWIĄTEK*

BUSINESS PROCESS OPTIMIZATION PLATFORM FOR INTEGRATED INFORMATION SYSTEMS

The paper aims to present a proposed attempt (Business Process Optimization Methodology) to select and develop proper analysis, planning and optimization methods for resource management in the information systems, dedicated to the domain of transport and logistics. We introduce the Business Process Optimization Platform (BPOP), which is an implementation of the proposed Business Process Optimization Methodology (BPOM) logic, and is currently in the development phase. The paper addresses an innovative approach to the identification of decision making problems in transport companies, the description of these problems with the domain ontologies and the development of the formal models leading to the dedicated algorithms which are delivered as complex Web services and may be directly integrated into software products. In such a way we achieve an agile approach to the development of service-based software systems with the reusable components (services). At the same time, we may quickly obtain the first results of the analysis and optimization of domain-specific processes.

1. INTRODUCTION

The subject of the proposed solution relates generally to the issue of the knowledge-based business process management using advanced computational techniques in the tasks of business process optimization in the transportation domain. The issue is addressed in two terms. The first is the Methodology, which is devoted to elaborate

* Faculty of Computer Science and Management, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland.
** Institute of Mathematics and Informatics, Opole University, Oleska 48 Str., 45-052 Opole, Poland.
*** Institute of Business Informatics, Wrocław University of Economics, Komandorska 118/120 Str., 53-345 Wrocław, Poland.

methods and tools allowing, mostly automatically, translation of business process descriptions into proper and adequate mathematical models, and further into domain-specific optimization tasks. The second is the framework (called the Platform), which is devoted to compose, mostly automatically, service-oriented application solving the discovered, domain-specific optimization tasks and extensively uses domain ontologies for knowledge processing.

Due to the independence of business rules management tools from the described processes, the same solution can be applied in different domain-specific management processes to solve different tasks and to address different purposes. This is achieved by distinguishing the functions of business processes from the decision making tasks, which are then associated with adequate formal models. Currently, one of the key unresolved issues and in the area of business computing is the problem of formulating a universal model of business rules. At the same time the number of decision problems and the optimization of operational (in particular in the area of transport systems) is huge, and the weight of their decisions – important for the efficiency of enterprises.

The main objective of the discussed approach is to propose flexible Methodology of selection of methods for analysis, planning and optimization purposes in management information systems in the field of transport and logistics. The general idea of the Methodology is based on assumption that many business process optimization tasks are based on the very similar mathematical models and that the possible, formulated basing on the mathematical models, optimization tasks may be solved using the same scope of algorithms.

A key research need addressed in the framework of the Methodology is the development and integration:

- methods of description, modeling, processing and optimization of business processes and the transport tasks being performed,
- algorithms for solving real-world optimization (decision making) problems,
- methods of flexible and adaptive composing adaptive of service-oriented making decision support systems in the context of a universal platform, which, in particular, will be used as a generator of domain-specific information systems.

Innovation research and anticipated results of the research involves a systematic approach to integrating the three, usually (as demonstrated by analysis of the state of the art) separated research areas related to the design, construction and implementation of integrated decision-making support systems. At the same time, it is planned to achieve innovative results in the form of the development of original techniques, methods and algorithms within each of these areas.

Achieving these expected outcomes requires the implementation of a number of research tasks relating to, inter alia, development of methods for the automatic composition algorithms to complex data and information processing, languages describing complex services with respect to their non-functional parameters and methods of translating business processes descriptions to the requirements addressing complex services supporting decision-making information systems. This is because of the necessity of individually designed algorithms to solve optimization problems in domain-specific decision-making support information systems, and the lack of methods for rapid prototyping of algorithms which solve unique optimization problems determined by specific business processes.

Known results on the task of developing methods for characterization, modeling , processing and optimization of resource management systems, passenger transport by road and rail point to numerous limitations and incompleteness of previously developed methods, including:

- Obtaining, scalable and open platform to generate domain-specific making-decision support information systems requires a tool to assist in the verification of models of business processes through incremental verification of the model and assessing the current [1–4].
- In many cases, existing models are not sufficiently precise (due to the use of simplifications) reflect the reality in the management of resources (including vehicles), scheduling problems (determination of timetables) and other related optimization problems (e.g. the inclusion of additional resources planning the location of bus stops, routes passes) take into account the additional assumptions and requirements (e.g. availability of vehicles associated with failures, reducing their efficiency, variable number of passengers, passengers who are disabled, etc.) [5, 6, 8, 9]. Due to the fact that, in reality, parameters of such problems may be changed, modeling must take into account the issues related to the resources availability and prediction of resources availability.
- Build effective mathematical models requires industrial research in the field of modeling and description of the processes occurring in the tasks related to the management of resources in transport systems, among others, to develop a list of the processes of business organizations to model business processes, obtaining data about the resources of the organization (transport, conditions of transport processes) as well as contextual information (used for infrastructure, the business environment of the organization) [7].

2. THE PROBLEM OF TRANSPORT PLANNING AND FRAMEWORK ARCHITECTURE

The terms used to in the proposed Methodology, issues of transport planning and optimization are formulated as follows (Fig. 1). The given business process, being an organization-specific set of rules, allows to identify requirements (criteria and constraints) necessary to formulate mathematical model and optimization problem (for example, set a new timetable). The given business process owner expects the optimal solution of the optimization tasks.

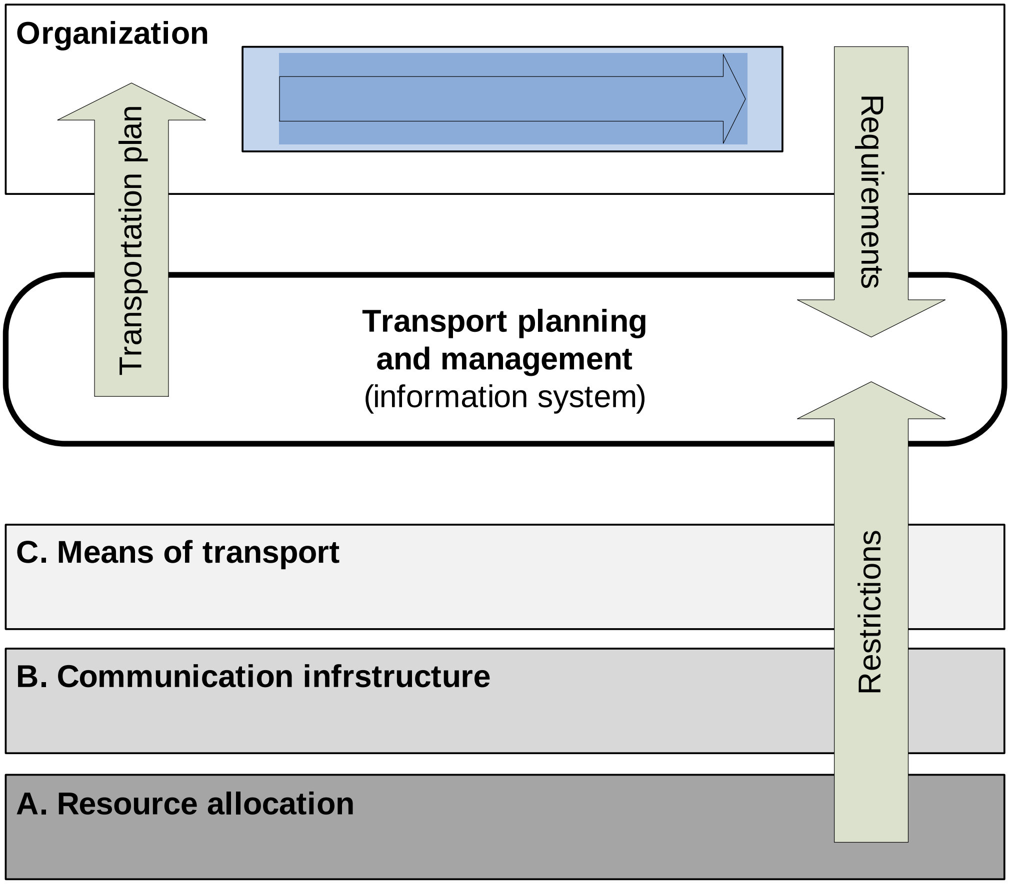


Fig. 1. Transport planning and optimization

Form of the resulting solution is determined by the means of transport, which are at the disposal of the organization, the available communication infrastructure, as well as the location of resources (people, goods, etc.) and relevant to the task on their transportation needs (Fig. 1). In this situation, an appropriate solution of the given optimization task depends on adequate representation the information system of knowledge about the requirements and conditions under which the transport processes are performed as well as on the appropriate definition of data sources and the use of dedicated optimization algorithms.

If additional requirement is the possibility of rapid prototyping solutions, dedicated to the organization and to provide reusability of selected components of the Platform (in particular implementing optimization algorithms), this leads to the selection of a suitable service architecture for the implementation of the Platform. In particular, the implementation of the components of the Platform as a Service allows to use attempts and improve approaches known in existing service-oriented systems:

- Languages for the description of services – to describe the services used are XML-based languages, in accordance with the recommendations of the World Wide Web Consortium (www.w3c.com), such as WSDL or OWL-S. It is also used domain-ontologies (dictionaries), allowing the description of the functionality and interfaces of services.
- Mechanisms for complex service composition – a composition of services currently requires the most operator intervention and is done in semi-automatic mode [10], and there are not available tools supporting the operator in the tasks of generating descriptions of services and management environments to enable the composition [11], in particular – tools integrating service composition and service execution engines.
- Automatic translation of business processes descriptions – a significant problem of direct translation of the business processes representation into the demands for services which implement the functions of the process is solved in part; usually for selected languages and with a limited range of applications [12]. There are not available systems being both complex solutions and offering required level of openness and flexibility, allowing them to be easily adapted to business process optimization needs [13].
- Mechanisms of adaptation and integration – the proposed Framework of services for process optimization purposes is expected to be the Framework where the above mentioned issues and proposed solutions are integrated in gain to obtain reconfiguration abilities in case of changes resulting from the changes in ongoing business processes (organization providing transport services).

The choice of service-oriented architecture makes possible to utilize widely used dedicated domain-specific dictionaries (ontologies), allowing for a consistent description of software components, ensure the compliance of the messages between them and enabling composition of services in complex processes based on pre-defined requirements.

3. METHODOLOGY OF FLEXIBLE CHOICE OF METHODS OF ANALYSIS, PLANNING AND OPTIMIZATION

The basic element that allows for the integration of the platform components, the use of appropriate data sources, selection of planning and optimization algorithms, and above all – an analysis of the organization's business process in a manner consistent with the objectives of enabling the use of an information system (frameworks) are ontologies – subject specific dictionaries containing terms and relationships that describe Platform components and the reality of the business process organization.

Among the mare:

- Organization Ontology – holding all concepts specific for the Organization being considered,
- Decision Task Ontology – with concepts needed to describe decisions such as vehicle routing problem, transport scheduling etc.
- Model Ontology, serving as a vocabulary for the description of formal models of the decision making tasks.

The proposed Methodology assumes the use of dictionaries available at the Platform for the description of Organization (transport company), the key decisions concerning transport tasks and the formal models of decision making problems (Fig. 2).

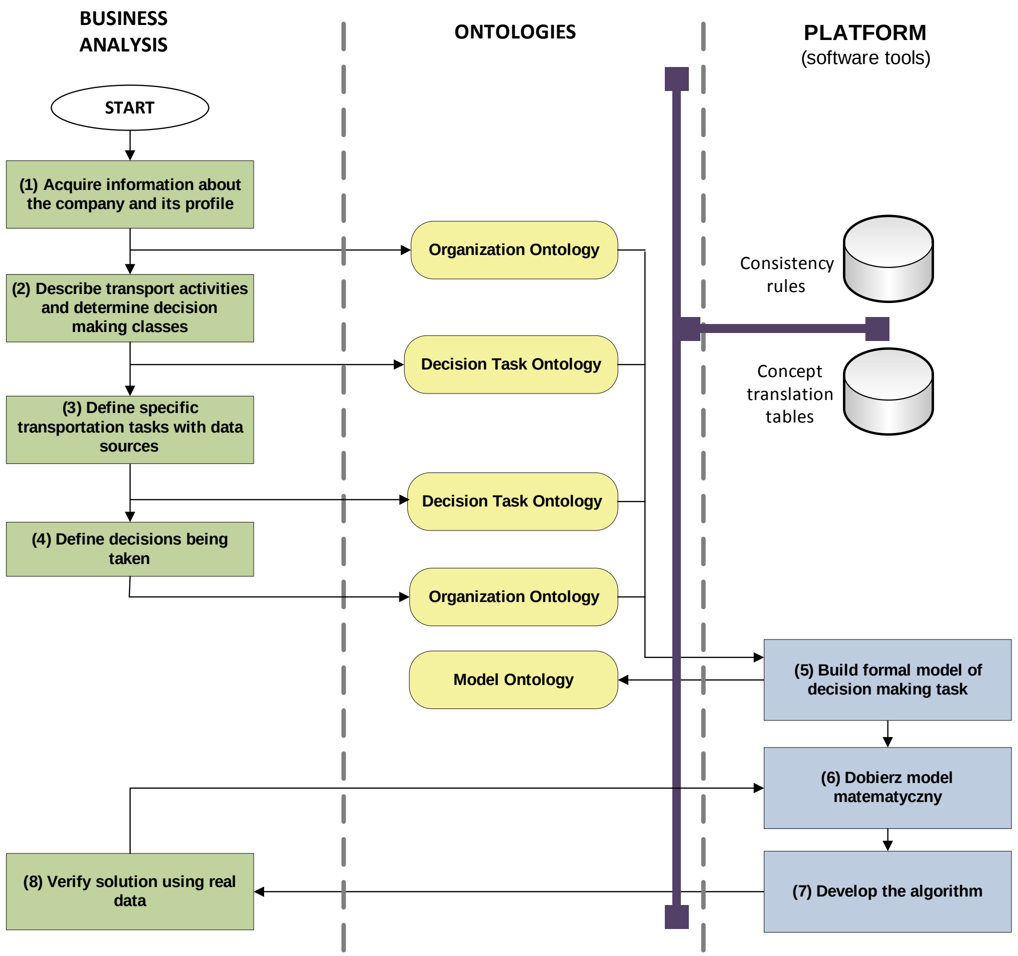


Fig. 2. Methodology of analysis, planning and optimization of transport tasks

The Methodology defines several steps, most of them involving the knowledge acquisition and extension of ontologies. The most important are:

(1) Preliminary interview: identifying Organization's profile, its resources and the transport tasks being performed.
(2) Determination and classification of key decision making tasks, met during realization of transport business processes.
(3) Detailed description of transport tasks.
(4) Definition of decision making tasks (in terms of criteria, constraints and parameters).
(5) Development of the formal models for identified decision making tasks.
(6) Development of mathematical model.
(7) Development of algorithm and/or algorithms for the mathematical model.
(8) Verification using real data which are acquired from the Organization and annotated with metadata from the Platform's ontologies.

As we see from the Fig. 2, most of the steps require adding appropriate concepts and relations (as well as their instances and attributes) to the ontologies. The final result – the algorithm and its implementation will allow to solve the identified decision making task under additional constraints (time, accuracy, computational and memory complexity, etc.). Such software is being delivered as a complex service, which consists of elementary services, among are the following services (linked together by the composition process in the Platform):

- computational,
- data access,
- data processing,
- data communication,
- user interface.

The above mentioned complex service is delivered by being part of the Platform runtime environment and is subject to monitoring and assessing the quality and efficiency of execution.

Linking the task of decision making task model identification with the selection of proper software components, available at the Platform in Software-as-a-Service (SaaS) mode, for the implementation of the process optimization is possible thanks to the original structure of dictionaries (ontologies) requires consistency between the ontological descriptions of the Organization, decisions and the models. In result, the proposed structure of dictionaries (Fig.3) is based on experience collected during previous works relating to service-oriented systems for Scientific Workflows information systems [14, 16, 17].

In order to maintain the consistency between the Ontologies, the Platform maintains repository of consistency rules as well as concept translation tables, which guarantee the consistency of the description of models, decision making tasks and the Organization (with its resources and business activities). Moreover, the software developed for the applied algorithms is, consistently, described using the metadata coming from the ontologies. Computing services, available at the discussed Platform, are devoted to solve optimization, analysis and planning problems, each of which corresponds to at least one mathematical problem (which may be formulated in different ways). The mathematical problem is solved by an appropriate algorithm, while maintaining the nature of the problem arising from constraints and criteria.

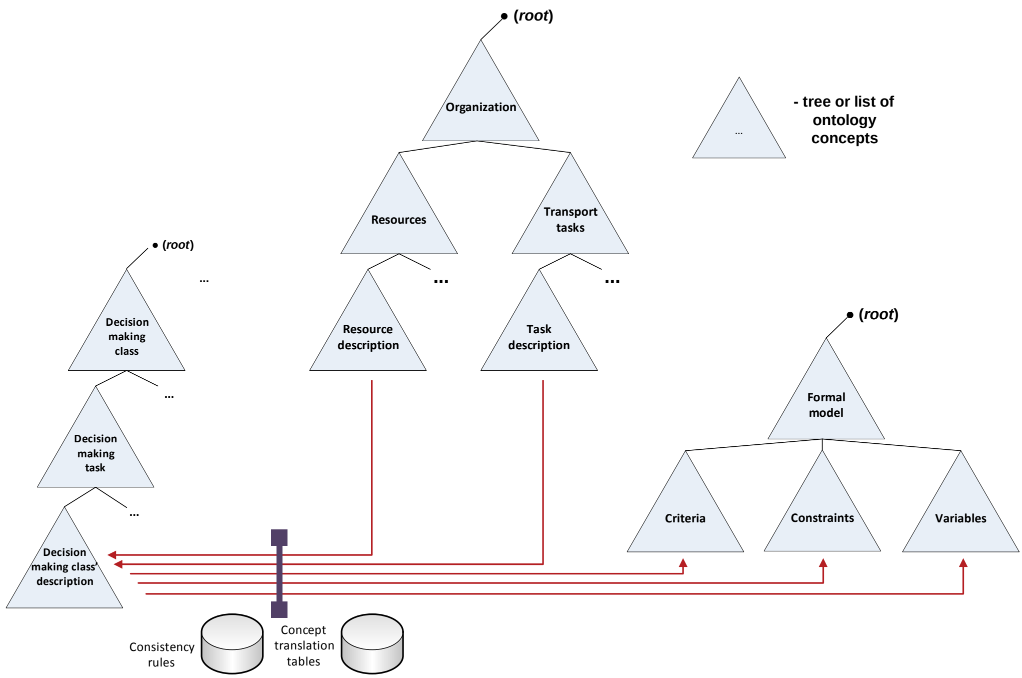


Fig. 3. Simplified structure of the Platform's ontologies

Because it is not expected that implemented computing components at the Platform will exactly match the needs of the organization and its specific (in terms of complexity, optimization problems and computational efficiency) optimization problems, an iterative approach is assumed in the proposed Methodology. Iterative approach gain is to deliver as soon as possible solution for recognized and identified optimization process and the obtained results are the starting point for further, if required, deeper investigation of the given decision making task and environment within which the process is performed. The further investigation is to extend transport task description and dictionaries in gain to recover detailed characteristics of the task, later represented in mathematical models and optimization tasks formulation.

CONLUSIONS

The proposed approach, involving the innovative use of a Platform which offers dedicated services along with domain-specific dictionaries and Methodology for selection analysis, planning and optimization methods for domain-specific business processes in the field of transport. The Platform is an integrated approach to the business process modeling, optimization problems formulation and software prototyping.

The Methodology is implemented as logic of Platforms for business processes optimization, the functionality of which should lead to increased efficiency and lower costs of each stage of the software life cycle (i.e. phases: requirements analysis, design, implementation, testing, deployment, maintenance and adaptation of information systems, cost reduction the various stages of manufacture of domain information systems).

The proposed Methodology allows to speed up development and reduce the cost of software development at all stages of the software life cycle: requirements analysis, design, implementation, validation, implementation, maintenance and product development.

It should be emphasized that special attention is paid to the possibility of demonstrating the results of the optimization process at an early stage and their use in subsequent phases of software development.

Taking into account that for the forthcoming Platform for all the above results will be used together, as part developed Methodology, a very important factor is the occurrence of innovation and the use of synergies between them. The proposed Platform will also have the functionality of the unknown in the current market solutions in the field of manufacturing of complex information systems, and allowing, among other things:

- automatic planning, adaptation and optimization of the logical architecture of domain information systems
- automatic adaptation of the modular system for the purpose of supporting the implementation of processes for which the system was not designed,
- automatic composition of software components available functionality defined in the domain-specific business processes.

ACKNOWLEDGEMENTS

The research presented in this paper waspartially supportedby the Polish Ministry of Science and Higher Education and the European Union within the European Regional Development Fund, Grant No. POIG.01.03.01-02-079/12 and within European Social Fund.

REFERENCES

[1] MAROUANE K., HOUARI S., MOUNIR B., *Model Transformation as an Optimization Problem*, MoDELS 2008, LNCS 5301, 2008, 159–173.

[2] LIN Y., *A Model Transformation Approach To Automated Model Evolution*, The University of Alabama at Birmingham, 2007.

[3] MENS T., Van GORP P., *A Taxonomy of Model Transformation*, Electronic Notes in Theoretical Computer Science, 2006, 152, 125–142.

[4] TISI M., JOUAULT F., FRATERNALI P., CERI S., BEZIVIN J., *On the Use of Higher-Order Model Transformations*. Proceedings of the 5th European Conference on Model Driven Architecture – Foundations and Applications, 2009, 18–33.

[5] CEDER A., *Public-transport vehicle scheduling with multi vehicle type*, Transportation Research Part C: Emerging Technologies, 2011, 19, 485–497.

[6] KHISTY C.J., LALL B.K., *Transportation Engineering: An Introduction*, Prentice Hall, 2003.

[7] MENS T., Van GORP P., *A Taxonomy of Model Transformation*, Electronic Notes in Theoretical Computer Science, 2006, 152, 125–142.

[8] RUDEK A., RUDEK R., *A note on optimization in deteriorating systems using scheduling problems with the aging effect and resource allocation models*, Computers and Mathematics with Applications, 2001, 62, 1870–1878.

[9] ZÄPFEL G., BÖGL M., Multi-period vehicle routing and crew scheduling with outsourcing options. International Journal of Production Economics, 2008, 113, 980–996.

[10] AGARWAL V. et. al., *Synthy: A system for end to end composition of web services*, World Wide Web Conference, 2005, 3(4), 311–339.

[11] PONNEKANTI S.R., FOX A., *SWORD: A developer toolkit for Web service composition*, 11th World Wide Web Conference, 2002, 97–103.

[12] HACKMANN G., GILL C., ROMAN G., *Extending BPEL for interoperable pervasive computing*, IEEE International Conference on Pervasive Services, 2007, pp. 204–213.

[13] SWIATEK P., STELMACH P., PRUSIEWICZ A., JUSZCZYSZYN K., *Service Composition in Knowledge-based SOA Systems*, New Generation Comput., 2002, 30(2–3), 165–188.

[14] GRZECH A., JUSZCZYSZYN K., KOŁACZEK G., KWIATKOWSKI J., SOBECKI J., ŚWIĄTEK P., WASILEWSKI A., *Specifications and Deployment of SOA Business Applications within a Configurable Framework Provided as a Service*, Studies in Computational Intelligence, Springer, 2013, 7–71.

[15] XUBO Fei, SHIYONG Lu, *A Dataflow-Based Scientific Workflow Composition Framework*, IEEE T. Services Computing, 2012, 5(1), 45–58.

[16] DIAMANTINI C., POTENA D., STORTI E., *Ontology-Driven KDD Process Composition, Advances in Intelligent Data Analysis VIII*, Lecture Notes in Computer Science, 2009, Vol. 5772, 285–296.

[17] OLIVEIRA D. et. al., *Ontology-based Semi-automatic Workflow Composition*, JIDM, 2012, 3(1), 61–72.

business analysis,
automatic questionnaire generation,
machine learning

Maciej ZIĘBA*, Jakub M. TOMCZAK*,
Krzysztof BRZOSTOWSKI*

ASKING RIGHT QUESTIONS BASING ON INCOMPLETE DATA USING RESTRICTED BOLTZMANN MACHINES

This paper focuses on the problem of asking relevant questions about the features describing concepts of interest basing on incomplete data. The main goal of this work is the novel approach that make use of Restricted Boltzmann Machines (RBM) to deal with the issue of eliminating missing values o attributes. The proposed model makes use of reconstruction abilities of RBM to sample the most probable value to be imputed for each of the missing cases. The entire research is supported by preliminary experimental studies that examine the quality of the proposed approach comparing to the reference technique.

1. INTRODUCTION

Clear understanding of consumer's needs is one of the crucial element of business analysis. Proper analysis of business process is indispensable step towards finding and solving the business problems. One of the well-known methods to discover consumer's needs is an interview which makes use of questionnaires with predetermined questions. The drawback of this solution is a disability to ask consecutive questions based on previous answers. As a consequence, some of the prepared questions can be unsuitable for some of the scenarios. Moreover, the ordering of the questions is crucial for eliminating redundant questions in the considered context. Selecting proper questions demands gradual domain knowledge acquired by a business analyst. Mentioned limitations can be overcame by implementing

* Wrocław University of Technology, The Institute of Computer Science, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland.

algorithm to design adaptive questionnaires. Such solution has ability to select proper question based on obtained answer. Additionally, sequence of questions is not fixed and it can be changed depending on answers.

The issue of selecting the proper set of questions was considered in various research fields. One of the example is 20 questions game [1]. The essence of this game is to figure out what word one person has in mind by asking as few questions as possible. In classical version of this game only binary questions are allowed (i.e. yes/no). Obviously there are some variants of this game in which higher level of uncertainty is acceptable. The problem of 20 questions game is not only considered as scientific problem. Some of designed algorithms are applied in game such as *Akinator* [2] (the goal is to guess the name of a famous character), *20Q A.I.* [3] (refers to the general objects) or *Winston* [4] (aims at discovering animals).

The problem of selecting proper question taking into account previous answers is crucial in the process of information retrieval on the Internet [5,6]. Information retrieval can be demanding task when the user cannot describe the desired term exactly. To deal with that issue the set of supporting questions can be asked to the user to construct the reasonable hints to identify the forgotten term.

Another example, strictly related to the business analysis, is acquisition of information for customer's requirements specification [7]. Such information can be collected during business interview. It is essential, that the sequence of questions asked during interviewing should be well-aimed and asked in right order. It helps to understand customer needs quickly and adequately.

As it was mentioned and the beginning, one of the common technique used during such business meetings is questionnaire with predefined questions. Usually, the analyst – a person who interviewing – is well prepared to the business meeting. However, he might not be prepared for all possible answers or topics and might not be familiar with all of the problems from the domain. Hence the idea of the system to support the analyst during business conversation.

Let us consider the example of business meeting in the company operating in the transportation domain. The goal of such meeting is to identify the problems solved in this company to provide various services for the customers. Some of this problem can be optimize and the role of the analyst is to find these problems that can be improved. For instance, one of the rendering service is related to the problem of scheduling bus drivers [8]. It is clear that it can be made in various way but some of the solutions are better than others in the point of view a performance index. The performance index usually is related to company's profits.

Most of the problems occurs in the transportation companies are well-addressed and described in the literature. Moreover, various solutions for these problems are proposed as well. It is worth stressing that each model for optimization problems are well-defined by the set of characteristic features. It means that the analyst ask the questions that are referred to these features. Such technique helps to identify the proper optimization model accommodated to the problems solved in the interested transportation company. In Figure 1 the process of interviewing is illustrated.

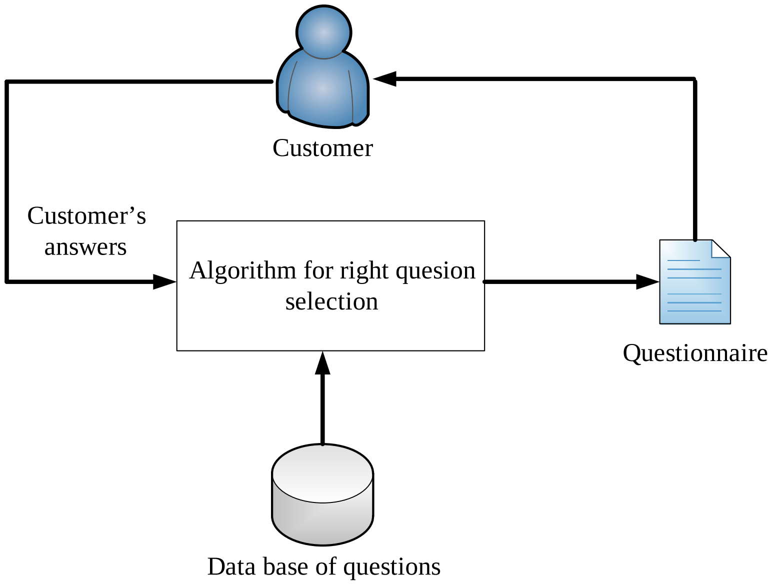


Fig. 1. The process of interviewing with selection of right questions

The presented applications indicate importance of the problem of choosing the right questions. More formally, the problem can be described in the following way. Assume that there are some concepts (e.g. products, services, animals, famous people, etc.) and each of the concept is described by the vector of common binary features. If the value of the i feature equals 1 then we say that the object has this property, and 0 – otherwise. For instance, the cat has the feature "mammal" equal 1. The main issue of the problem is to select a sequence of the features values to be questioned about that will help to discover the right concept. Additionally, total number of selected features should be minimized.

Several methods are presented in the literature to solve the problem of selecting proper features to be questioned about. One of them is to represent the problem as the classification problem with the individual class label for each of the concepts, in which a classifier is constructed, e.g., *C4.5* decision tree [9]. This group of classification methods has ability to put only relevant features in the nodes, while edges represent decision variants and class labels are stored in the leafs of the tree. As a consequence, the process of discovering features is performed by starting from the root that contains the most informative feature and moving downwards on the selected path until one of the leafs is reached.

A different approach to solve the stated problem is described in [6] which makes use of semantic network together with Concept Description Vectors (CDF) [10] to search for the right questions. Each feature in the network is represented by a node and each relation between two features is described with an edge. Each edge is described by a CDF that contains two values denoting support of the connection and the confidence of knowledge. The questions are selected basing on the CDFs of the features in relation with the actually considered feature.

One of the recent works concentrates on the application of Restricted Boltzmann Machines (RBM) to approximate the distribution over the features describing concepts, which is further used to formulate proper sequence of questions [11]. The RBMs are widely used in many applications including: features extraction [12], classification [13] and collaborative filtering [14]. In the considered application authors exploit the reconstruction abilities of RBM to identify the most relevant features. They represent the concepts by the sparse binary vectors of the features and they train the RBM in the unsupervised mode. Next, they construct an evolving random process to find the most probable features to be asked for. Finally, they identify the closest concept basing on the answers to the stated questions.

The main drawback of this approach is that it assumes that each of the concepts is described by complete vector of features. Practically it means, that for selected concept each of the features are known. Such situation is rarely observed in practice, especially if the values of the features describing the concept are being assigned by a human that has not complete knowledge about all concepts. Moreover, an additional concept can be included in the data and the vector of features should be extended by incorporating additional attributes that are characteristic for this additional component. As a consequence, the values of the new features for the existing concept are unknown and should be imputed either by a human or by an algorithm. If the imputation is performed by a machine this issue is known in literature as missing values problem [15].

In this paper we propose a novel method of imputing unknown values that also relays on RBM model. The main idea of this approach is to construct RBM model assuming incompleteness of the data and further put the most probable values using the trained model. While completing the missing cases by receiving the real values gathered during the interview the RBM is updated with actual data and the missing cases are refilled by the updated model. In this approach the reconstruction abilities of RBM are used to set the most probable values for the missing cases.

The paper is organized as follows. In Section 2 authors present reference approaches that are used to deal with missing values problems. Section 3 contains description of the RBM-based imputation technique. Section 4 gathers some preliminary results of experimental studies. The paper is summarized by some conclusions in Section 5.

2. RELATED WORKS

Various techniques are applied to solve the missing values issue [15]: eliminating cases with missing values, imputation or estimation of missing values, using model-based procedures, adjusting machine learning approaches. In first two groups of the methods the problem of missing values is handled externally, either by eliminating examples (or attributes) containing missing values, or by completing missing cases with the most probable value or by the application of some other imputation techniques [16]. In contrast, solutions from the third group model the probability density function of the input data (complete and incomplete cases), which is further used to classify new instances basing on Bayes decision theory. In the last group, the classifier is designed for handling incomplete input data by modifying typical training procedures. In this group we can distinguish techniques, that are used to construct decision trees [17], Support Vector Machines [18] and ensemble classifiers [19] directly from incomplete data.

3. RESTRICTED BOLTZMAN MACHINES FOR MISSING VALUES IMPUTATION

RBM is bipartite Markov Random Field in which visible and hidden units can be distinguished. Practically it means that only connections between units from different layers are observed. The joint distribution of hidden and visible units is a Gibbs distribution given by the formula:

$$p(\mathbf{x}, \mathbf{h} \mid \boldsymbol{\theta}) = \frac{1}{Z} \exp(-E(\mathbf{x}, \mathbf{h} \mid \boldsymbol{\theta})) \tag{1}$$

where the energy function is defined as follows:

$$E(\mathbf{x}, \mathbf{h} \mid \boldsymbol{\theta}) = -\mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{h} - \mathbf{b}^{\mathrm{T}}\mathbf{x} - \mathbf{a}^{\mathrm{T}}\mathbf{h} \tag{2}$$

Vector $\mathbf{x} \in \{0,1\}^D$ represents visible units, vector $\mathbf{h} \in \{0,1\}^M$ stays behind hidden units, $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{a}, \mathbf{b}\}$ gathers together all parameters of the model (**W** is weight matrix, **a** and **b**are bias vectors for hidden and visible units respectively) and Z is the normalization constant depending on $\boldsymbol{\theta}$.

Since there are no there are no connections among the units within the same layer, i.e., no visible to visible, or hidden to hidden connection, the visible units are conditionally independent given the hidden units and vice versa:

$$p(x_i = 1 | \mathbf{W}, \mathbf{b}) = \text{sigm}(\mathbf{W}_i \mathbf{h} + b_i) \tag{3}$$

$$p(h_j = 1 | \mathbf{W}, \mathbf{a}) = \text{sigm}(\mathbf{W}_j \mathbf{x} + a_i) \tag{4}$$

where $\mathbf{W}_i$ is i-th raw of the weight matrix, $\mathbf{W}_j$ is j-th column of the weight matrix and is the sigmoid function:

$$\text{sigm}(a) = \frac{1}{1 + \exp(-a)} \tag{5}$$

Unfortunately, in order to learn parameters $\boldsymbol{\theta}$ gradient-based optimization methods cannot be directly applied because exact gradient calculation is intractable. Fortunately, we can adopt Contrastive Divergence algorithm which approximates exact gradient using sampling methods [12].

The proposed model was used to deal with the missing values issue in the context of asking the right questions. Assume that we have given the data representing the collection of concepts, $\mathbf{X}_N = \{\mathbf{x}_n\}_{n=1}^N$. Each element in $\mathbf{X}_N$ represents separate concept that is characterized by D features. If the complete vector of features ($\mathbf{x}_n \in \{0,1\}^D$ for each $n \in \{1, ..., N\}$) the procedure of selecting right questions described in [11] can be applied. In the other case some imputation procedure should be performed to deal with that issue.

In this paper we propose the novel technique for completing unknown values of features. Assume, that for unknown values we assign 0.5, $\mathbf{x}_n \in \{0, 0.5, 1\}^D$ for each $n \in \{1, ..., N\}$. It can be observed, that such vector does not fit to visible layer of the RBM model, because of additional possible value 0.5. However, instead of operating on binary vectors of the hidden layer we can operate directly on the probabilities given by (3). Practically it means, that for unknown value of the feature the probability given by (3) is equal 0.5, if the concept is not characterized by the feature the probability value is equal 0 and is equal 1 in the remaining case. Taking under consideration this assumption we can train the RBM model on incomplete data.

The procedure of imputing missing values can be described in the following steps:

1. Train the RBM model with incomplete data $\mathbf{X}_N = \{\mathbf{x}_n\}_{n=1}^N$, where $\mathbf{x}_n \in \{0, 0.5, 1\}^D$ for each $n \in \{1, ..., N\}$.

2. For each $\mathbf{x}_n$ in $\mathbf{X}_N$ sample $\mathbf{h}_n$ from distribution given by equation (4). Parameters **W** and **a** were estimated during training RBM model in Step 1.
3. For each $\mathbf{h}_n$ sample $\bar{\mathbf{x}}_n$ from distribution given by equation (3). Parameters **W** and **a** were estimated during training RBM model in Step 1.
4. For each $n \in \{1, ..., N\}$ create $\tilde{\mathbf{x}}_n$ that takes on *j*-th position the *j*-th value of $\bar{\mathbf{x}}_n$ if *j*-th value of $\mathbf{x}_n$ is equal 0.5 and *j*-th value of $\mathbf{x}_n$ otherwise.

As a consequence, $\tilde{\mathbf{X}}_N = \{\tilde{\mathbf{x}}_n\}_{n=1}^N$ contains only complete components that inherit certain values from $\mathbf{X}_N$ and missing values are imputed by sampling the probable values from RBM model. Further, the procedure of selecting right questions is performed starting with imputed data $\tilde{\mathbf{X}}_N$. This time the RBM model is trained on complete data and the sampling procedure is performed to find the sequence of relevant questions. This procedure is described in details in [12].

During asking the questions about the features we can observe, that some of the unknown values can be discovered in the given answers. In such situation the data $\mathbf{X}_N$ should be completed with the discovered values and the imputation process should be run once again using updated dataset $\mathbf{X}_N$.

4. THE EXPERIMENTAL STUDIES

In this section we aim at evaluating empirically the performance of the proposed method of imputing missing values. First, we take under consideration well-known Zoo dataset available in UCI Machine Learning Repository [20]. Each of the instances in the considered dataset represents one of the 101 animal species. Each animal is described by a vector of 28 binary features. The dataset was used in [12] to examine the performance of the method of selecting relevant features to identify the animal basing on complete data. In this section we examine the quality of imputing missing values of the animals' features using RBM model.

The methodology of the experiment is as follows. We randomly create missing cases by covering some percentage of values in Zoo dataset. Next, we examine the reconstruction capabilities of the model by examining the reconstruction error, which is defined by the ratio between correctly assigned values and total number of imputed cases. The entire experiment is repeated 100 times for different percentage of missing cases. We compared the performance of our solution with the results achieved by benchmark method that assigns most time observed (MTO) value for each attribute.

The results of preliminary experimental studies are presented in Fig. 2. The reconstruction error was calculated for percentages of missing values starting from 1% to

20% of unknown cases. It can be observed that for RBM-based approach the reconstruction error is growing when the percentage of missing values increases, while for the reference MTO model the level of reconstruction error does not depend on the number of unknown values of attributes. As a consequence for low number of unknown values the RBM-based imputation significantly outperforms the reference approach. The more complete data is stored in the RBM model, the higher is the accuracy of predicting the missing cases. For high (over 20 %) percentage of missing data the MTO gained better result than RBM imputation technique. The reason for that is the dominance of zeros in the training data. As a consequence, the MTO suggests to complete missing values with 0 much more often than with 1.

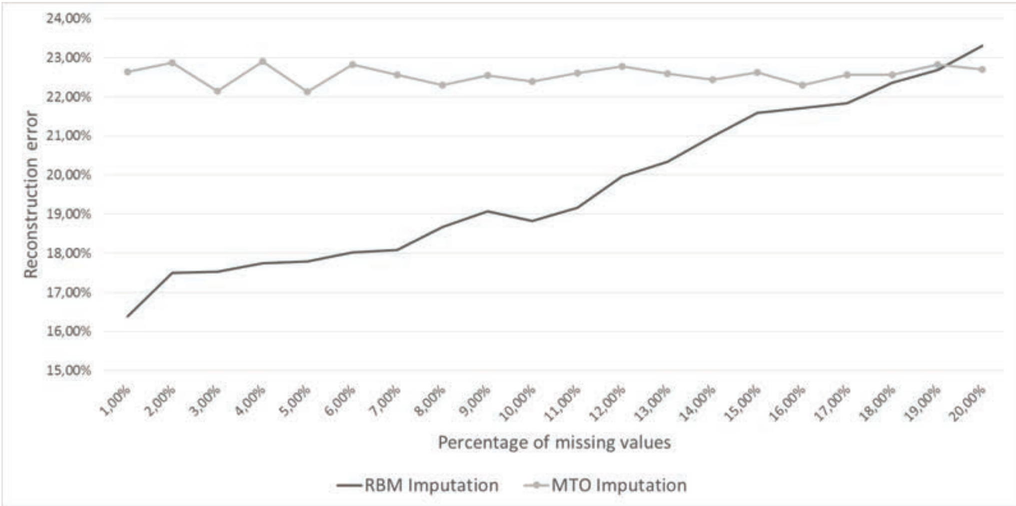


Fig. 2. The chart presents reconstruction error values for different percentage of missing values considering RBM-based and Most Time Observed (MTO) imputation

5. CONCLUSIONS AND FUTURE WORKS

This paper concentrates on the problem of asking relevant questions about the features describing concepts of interest basing on incomplete data. The main contribution of this work is the application of Restricted Boltzmann Machines to the problem of eliminating missing values o attributes. The proposed model makes use of RBM to sample the most probable value to be imputed for each of the missing cases. Preliminary results show, that RBM-based imputation technique performs significantly better than one of the most popular reference approaches.

For the future works authors are going to consider the use case in which the answers for some questions are uncertain or incorrect. Further, they are going to handle the issue of noisy data. Finally, they are going to incorporate some relations among the features (hierarchy, ontologies) to accelerate the procedure of detecting the concept basing on the sequence of answers.

ACKNOWLEDGEMENTS

The research presented in this paper was partially supported by the Polish Ministry of Science and Higher Education and the European Union within the European Regional Development Fund, Grant No. POIG.01.03.01-02-079/12 and within European Social Fund.

REFERENCES

[1] DUCH W., SZYMAŃSKI J., SARNATOWICZ T., *Concept description vectors and the 20 question game*, [in:] *Intelligent Information Processing and Web Mining*, Springer, 2005, 41–50.
[2] ELOKENCE, *Akinator*, URL http://en.akinator.com/, last access 07.2013.
[3] 20Q.net Inc. Q20 A. I, URL http://www.20q.net/, last access 07.2013.
[4] WINSTON, http://kask.eti.pg.gda.pl/winston/20q, last access 07.2013.
[5] DUCH W., SZYMAŃSKI J., *Semantic web: Asking the right questions*. Proceedings of the 7 International Conference on Information and Management Sciences, 2008, 1–8.
[6] RZENIEWICZ J., SZYMAŃSKI J., DUCH W., *Adaptive algorithm for interactive question-based search*, [in:] *Intelligent Information Processing VI*, Springer, 2012, 186–195.
[7] WANG M., ZENG Y., *Asking the right questions to elicit product requirements*, International Journal of Computer Integrated Manufacturing, 22(4), 2009, 283–298.
[8] LOURENÇO H., PAIXÃO J., PORTUGAL R., *Multiobjective metaheuristics for the bus driver scheduling problem*, Transportation Science, 2001, 35(3), 331–343.
[9] QUINLAN J.R., *C4.5: Programs for machine learning*, Machine Learning, 1994, 16, 235–240.
[10] SZYMAŃSKI J., DUCH W., *Information retrieval with semantic memory model*, Cognitive Systems Research, 2012, 14(1), 84–100.
[11] ZIĘBA M., TOMCZAK J. M., BRZOSTOWSKI K. *Selecting right questions with Restricted Boltzmann Machines*, ICSEng, 2014 (in press).
[12] HINTON G., *A practical guide to training Restricted Boltzmann Machines*, [in:] *Neural Networks: Tricks of the Trade*, Springer, 2012, 599–619.
[13] LAROCHELLE H., BENGIO Y., *Classification using discriminative Restricted Boltzmann Machines*. Proceedings of the 25th International Conference on Machine Learning, ACM, 2008, 536–543.
[14] SALAKHUTDINOV R., MNIH A., HINTON G., *Restricted Boltzmann Machines for collaborative filtering*. In: Proceedings of the 24th International Conference on Machine learning, ACM, 2007, 791–798.
[15] GARCÍA-LAENCINA P.J., SANCHO-GÓMEZ J.L., FIGUEIRAS-VIDAL A.R., *Pattern classification with missing data: a review*, Neural Computing and Applications, 2010, 19(2), 263–282.
[16] GARG L., DAUWELS J., EARNEST A., PANG L., *Tensor based methods for handling missing data in quality-of-life questionnaires*, Journal of Biomedical and Health Informatics, 2013.

[17] ZHANG S., QIN Z., LING C.X., SHENG S., *“Missing is useful”: missing values in cost-sensitive decision trees*, IEEE Transactions on Knowledge and Data Engineering, 2005, 17(12), 1689–1693.
[18] PELCKMANS K., De BRABANTER J., SUYKENS J.A., De MOOR B., *Handling missing values in support vector machine classifiers*, Neural Networks, 2005, 18(5), 684–692.
[19] NANNI L., LUMINI A., BRAHNAM S., *A classifier ensemble approach for the missing feature problem*, Artificial Intelligence in Medicine, 2012, 55(1), 37–50.
[20] BACHE K., LICHMAN M., *UCI machine learning repository*, last access 07.2013.

transportation system,
vehicle routing problem, route planning,
driver scheduling

Dariusz GĄSIOR*, Jerzy JÓZEFCZYK*,
Grzegorz FILCEK*, Maciej HOJDA*,
Donat ORSKI*, Magdalena TUROWSKA*,
Piotr DĄBROWSKI*

MODELS AND ALGORITHMS FOR ROAD TRANSPORT PROBLEMS

The main tasks of planning road transport include route planning, scheduling of drivers and scheduling of vehicles. Organization of transport involves addressing all these decision-making tasks simultaneously, in order to obtain the greatest possible profits from the operation of a transport company. However, attempting joint approach to these problems, causes that the models are too complex. In consequence the well-known algorithms become not useable and the new solutions are hard to be obtained. Thus, the most commonly used decomposition of the overall problem into a number of simpler optimization tasks. In this paper, the possibility of decomposing the problem into smaller is indicated. The exemplary models and algorithms solutions for some simple tasks are introduced.

1. INTRODUCTION

The purpose of the transport company is the realization of orders consisting of transporting people or goods from given sites to specified locations with respect to the additional constraints (i.a. related to the driving time or law regulations). Having certain amount of material and human resources, the company need to plan their work to achieve their objectives, e.g. to maximize profit.

Thus, decision-making tasks for transport companies include number of allocation and scheduling problems, in particular route planning, vehicle scheduling and drivers scheduling problems.

---

* Wrocław University of Technology, Institute of Computer Science, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland.

In this paper, we will focus only on tasks related to the road transportation. Moreover we restrict our considerations to the transportation of people problem. We explain how the mathematical models for such a problems may be developed and we indicate solution algorithms, which may be applied to solve the considered problems.

2. TRANSPORT PLANNING PROBLEMS

The three fundamental decision-making problems for road transport may be distinguished:

a) planning of routes and timetables [1], [2],
b) material resource allocation problem (vehicle assigning) [3], [4],
c) human resource allocation problem (drivers scheduling) [5], [6].

All these tasks may be considered separately. However, they are strongly related, which is depicted in Fig. 1. For example a solution of route planning is the input data for bus assignment task and drivers scheduling. On the other hand, the solution of the resource allocation problems allows to determine the real cost of routes. In consequence, such decomposition does not allow obtaining an optimal solution of transport planning problem for a company.

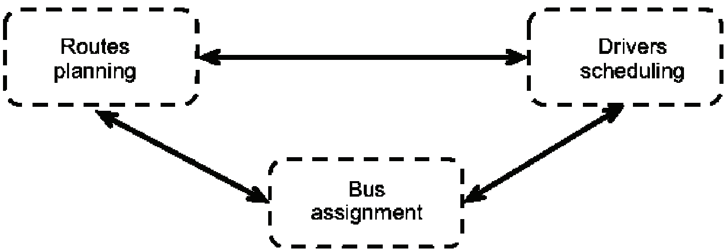


Fig. 1. Relationship between the main transport planning problems

Still, usually all these problems are extremely complex. Thus, either the mathematical models for them are either too complicated which makes them useless or requires some simplification or further decomposition.

Let us consider the drivers scheduling problem in EU countries. There are number of regulations, which must be taken into account. There are limitations for driving time during day, week, and month. The appropriate spare time for drivers must be assigned.

When these requirements are not met, the company must pay high penalties. Moreover, each firm has its own rules, which must be also taken into account.

As a result, the following decomposition of the drivers scheduling problem into two subproblems is very convenient:

a) planning the workday for drivers and allocation of material resources,

b) scheduling of work monthly drivers (based on the daily plans).

In this paper, we present the possible model for the exemplary people transportation problem.

4. MODELS AND ALGORITHMS FOR TRANSPORTATION OF PEOPLE PROBLEMS

Assume that, the company transports people from their origin locations to the destination. Each person has defined the time when he must arrive and the longest time he may spend in the bus. Such a rides repeats periodically.

Thus, the problem is to make such a timetable and to allocate resources in such way that all passengers reach their destination within the prescribed time limit and the legal restrictions are preserved.

One may decompose this problem into sequence of the simpler optimization task. The first one consists in finding courses that compose valid timetable. The second one consists in finding driver's workday plans, which are composed of courses to be traversed. The third one consist of assigning workday plans to particular drivers for the given time horizon (usually one month).

4.1. ROUTE PLANNING PROBLEM

The route planning problem consists in finding sequences of locations and times at which they are visited by the vehicle, so all passengers are enabled to travel according their requirements and the company achieves the best possible value of its objective (e.g. the highest profit, the lowest cost).

Once, we want to formulate such a problem as an optimization task, we must introduce mathematical model to describe road map, demands, decisions which are to be made and their relationships. The proposed model is based on well-known approaches to the vehicle routing problems with time-windows, since the problem may be classified as an open pick-up and delivery vehicle routing problem with time windows and transhipments [7], [8].

The road map may be represented as a directed multi-weighted multi-graph as it is depicted in Fig. 2. Each node (v^{I}_i) represents some location on the map and there may be more than one road ($e^{\mathrm{I}}_{i,j,m}$) between each pair of locations (v^{I}_i, v^{I}_j). Moreover, each road may be described by more than one parameters, i.e. distance ($d^{\mathrm{I}}_{i,j,m}$), travel time ($t^{\mathrm{I}}_{i,j,m}$), travel cost ($c^{\mathrm{I}}_{i,j,m}$).

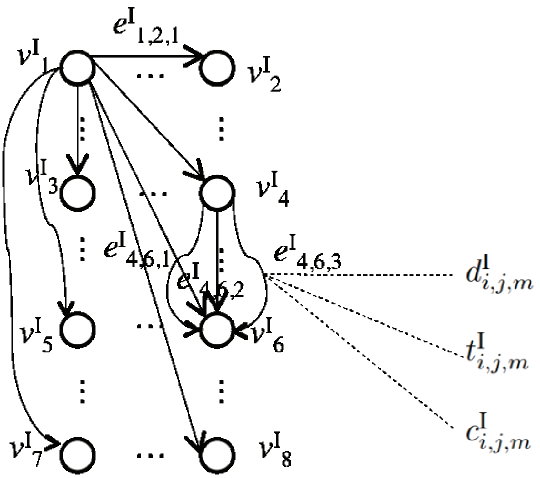


Fig. 2. Road map as a multigraph

Each transport demand (indexed with n) is characterized by: the origin location ($v^{I}_{o,n}$), the earliest departure time ($t^{I}_{o,n,\min}$), the latest departure time ($t^{I}_{o,n,\max}$), the destination location ($v^{I}_{d,n}$), the earliest arrival time ($t^{I}_{d,n,\min}$), the latest arrival time ($t^{I}_{d,n,\max}$). Furthermore, in some cases the requirements concerning travel comfort should also be taken into the consideration, e.g. maximal travel time or maximal number of transfers.

The exemplary set of transport demands is presented in the Fig. 3.

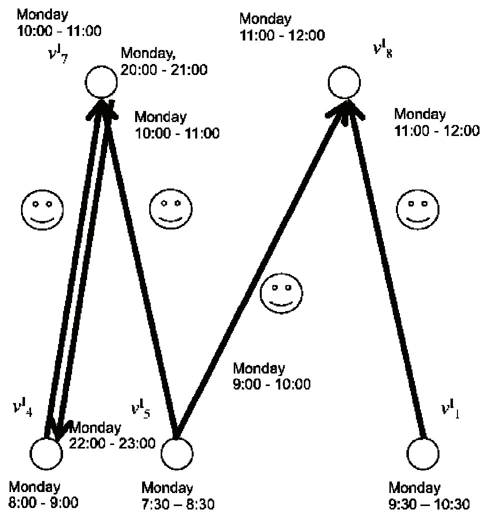


Fig. 3. Sample transport requests

The decision, which is to be made, consists of the set of courses. Each course (indexed with *k*) is defined by the sequence of nodes (locations), which are visited, edges (roads), which are used to travel between two subsequent nodes ($z^{\mathrm{I}}_{i,j,m,k} = 1$ indicate that *m*-th edge is used during *k*-th course between nodes *i* and *j*) and arrival times $t^{\mathrm{I}}_{k,i}$ related to all nodes, which are to be visited. Furthermore, the trip plan for each passenger ($\hat{y}^{\mathrm{I}}_{n,k,i,j} = 1$ indicate that nth passenger travel with *k*-th course between nodes *i* and *j*) has also to be determined since one wants to verify if all passengers may travel according to their requirements. The decision variables and their relations to the road map and transport demands are illustrated in Fig. 4.

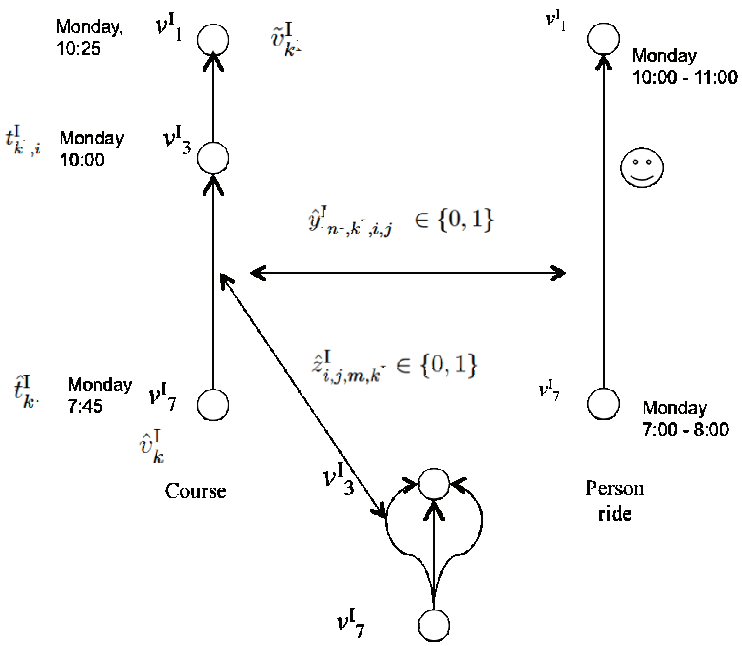


Fig. 4. The part of the exemplary solution
with relationship to the road map and transport demands

The route planning may be formulated as a linear integer programming problem. Furthermore, it may be seen as a specific class of vehicle routing problem, so the appropriate algorithms may be adapted to solve the problem, e.g. exact algorithms [9], approximate algorithms [10], heuristics [11], genetic [12], etc.

Once, the set of courses is determined. The most important is to schedule drivers. As it was mentioned, we propose to decompose this problem into two subproblem. First one consist in determining driver's workday plans, i.e. such a sequence of courses which may be performed by one driver during one day of work and neither legal restrictions nor company regulations are broken.

4.2. DRIVER'S WORKDAY PLANNING PROBLEM

Planning driver's workdays consist in grouping courses into a sequence in such a way that there is enough time to drive from the last location of previous course to the first location of the next course and the driver has a break if it is needed. The total time of such workday plan may not exceed maximal working time defined by the legal restrictions or company rules. Additionally, there is enough time to go back to the depot and go from the depot if it is required (e.g. at the beginning and at the end of workday). The set of driver's working day plans should allow maximizing (or minimizing) company's objective (e.g. maximizing profit or minimizing costs). The idea of composing courses into workday plans is summarize in Fig. 5.

Each course (indexed with i) is characterized with its start time (t^{II}_i) and duration time (τ^{II}_i). The decision variable x^{II}_{il} reflects if i-th course to the l-th workday plan. Moreover, the appropriate type of vehicle must be assigned to the workday plan, since there is particular vehicle capacity needed to serve given courses. So, the y^{II}_{ik} variable determines if k-th bus type is assigned to the i-th course.

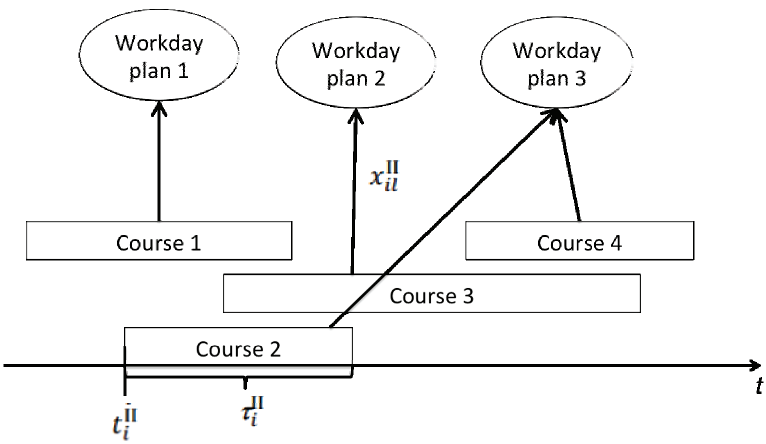


Fig. 5. The idea of the composition of workday plans

The example of workday plan is presented in Fig. 6. The particular bus type is assigned to serve course number 2 and number 4. It starts in depot number 1 and goes to serve course number 2, then directly go to the beginning of the course number 4 and finally it goes back to the depot number 1.

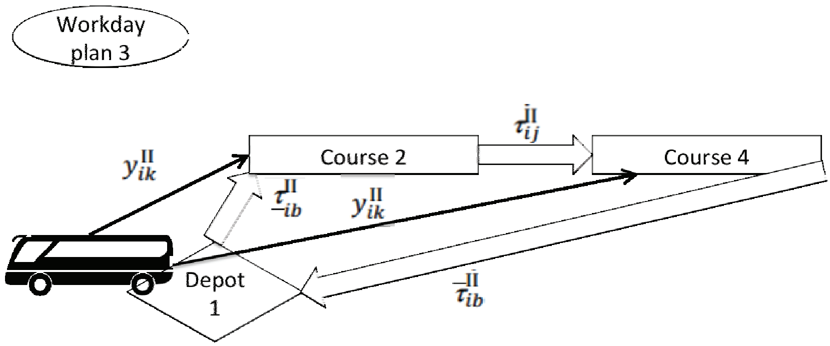


Fig. 6. The example of workday plan

The considered problem may be formulated as a binary linear programming problem. It may be also viewed as a scheduling problem with moving executors [13] or it may be transformed to the multiple traveling salesman problem (mTSP) [14]. Consequently, the well-known solution algorithms may be directly applied to such a problem. One may use an exact method ([15], [16]) to determine optimal workday plans set. Since, such a problem will be solved rarely, so some of these methods are acceptable even though they are not time efficient. If the solution is needed to be obtained quickly, the well-known approximate algorithms [17] or heuristic [18] (including e.g. genetic algorithms [19]) may be applied.

4.3. DRIVER SCHEDULING PROBLEM

Having the workday plans and set of drivers, one may schedule the work for the given time horizon (week, month, year). Taking into consideration drivers' vacations, illness, obligatory breaks, one should find such an assignment of workday plans to drivers that all courses in the given time horizon will be realized and neither legal restrictions nor company rules are violated. The previous schedule must be also taken into an account, since it affects, when the obligatory breaks for the driver must be made. The assignment should be the best possible, i.e. giving the best possible value of objective (e.g. maximize profit or minimize costs).

So, we assume that a time sample is Δt (e.g. 15 minutes) and the initial moment for schedule is κ. For each driver d and for each time sample n in the time horizon $N(\kappa)$, we decide if it should be assigned to the jth workday plan or have a spare time. This is determined with variable $x^{III}_{d,j,n}(\kappa)$. If the driver obtain the spare time, it must be indicated what type of spare time it is, i.e. obligatory break, vacations, etc. We use variable $y^{III}_{d,n,l}(\kappa)$ for such a purpose, where lth is the index of free time type. The idea of this model is presented in Fig. 7.

4.3. DRIVER SCHEDULING PROBLEM

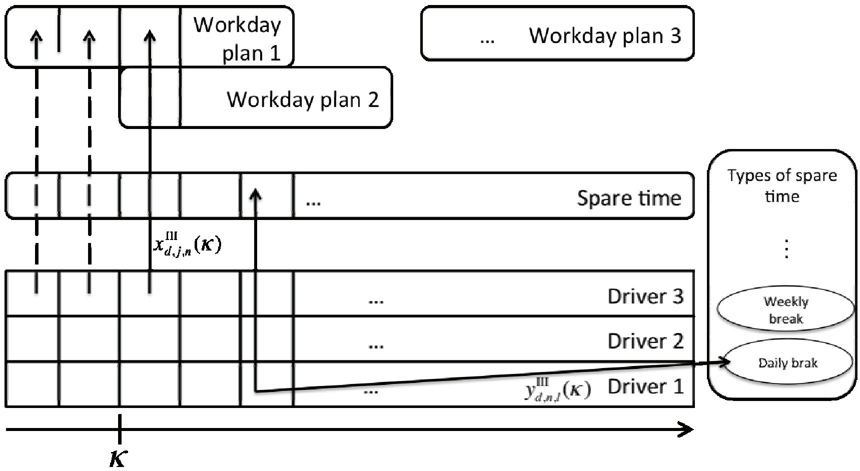


Fig. 7. The illustration of the model for driver scheduling problem

All regulations should be expressed in terms of mathematical equations. It makes the problem very complex. For instance modelling all the legal restrictions for Polish drivers needs adding over 80 linear constraints. Even though it can be still given as a linear integer programming.

To solve such a problem a number of methods were proposed, e.g. genetic algorithms [20], tree search algorithm [21] and other heuristics [22].

Some models and algorithms are even dedicated for a driver scheduling problem in European Union, while the specific regulations have to be incorporated [23].

5. CONCLUSIONS

In this paper, the main decision making tasks for a road transport have been introduced. It was discussed how the mathematical model are developed for such problems. We indicated that all these problems might be formulated as an integer programming problem. However, in practical application, when all constraints are taken into account, the problems are NP-hard. We indicated the algorithms, which may be used to solve these problems.

As it was mentioned, usually all introduced problems are solved separately. However, undoubtedly, solving them jointly may significantly improve the transportation company's profits (for more details see [24], [25]).

ACKNOWLEDGMENTS

The research presented in this paper was partially supported by the European Union within the European Regional Development Fund program Number POIG.01.03.01-02-079/12.

REFERENCES

[1] CHRISTOFIDES N., *The vehicle routing problem*, RAIRO-Operations Research-Recherche Opérationnelle, 1976, 10(1), 55–70.
[2] TOTH P., VIGO D., *The vehicle routing problem*, Siam, 2001.
[3] FOSTER B.A., RYAN D.M., *An integer programming approach to the vehicle scheduling problem*, Operational Research Quarterly, 1976, 367–384.
[4] SALZBORN F., *Optimum bus scheduling*, Transportation Science, 1972, 6(2), 137–148.
[5] WREN A., ROUSSEAU J.-M., *Bus driver scheduling – an overview*, Springer, Berlin–Heidelberg, 1995.
[6] FORES S., PROLL L., WREN A., *An improved ILP system for driver scheduling*, [in:] *Computer-Aided Transit Scheduling*, Springer Berlin–Heidelberg, 1999, 43–61.
[7] KUMAR S.N., PANNEERSELVAM R., *A survey on the vehicle routing problem and its variants*, Intelligent Information Management, 2012, 4(66).
[8] CORDEAU J.-F., DESAULNIERS G., DESROSIERS J., SOLOMON M.M., SOUMIS F., *VRP with time windows*, [in:] *The vehicle routing problem*, 2001, 157–193.
[9] BALDACCI R., HADJICONSANTINOU E., MINGOZZI A., *An exact algorithm for the capacitated vehicle routing problem based on a two-commodity network flow formulation*, Operations Research, 2004, 52(5), 723–738.
[10] LAPORTE G., *The vehicle routing problem: An overview of exact and approximate algorithms*, European Journal of Operational Research, 1992, 59(3), 345–358.
[11] LAPORTE G., GENDREAU M., POTVIN J.Y., SEMET F., *Classical and modern heuristics for the vehicle routing problem*, International transactions in operational research, 7(4–5), 2000, 285–300.
[12] BAKER B. M., AYECHEW M.A., *A genetic algorithm for the vehicle routing problem*, Computers and Operations Research, 2003, Vol. 30, No. 5, 787–800.
[13] JÓZEFCZYK J., *Scheduling tasks on moving executors to minimise the maximum lateness*, European Journal of Operational Research, 2001, 131(1), 171–187.
[14] BEKTAS T., *The multiple traveling salesman problem: an overview of formulations and solution procedures*, Omega, 2006, 34(3), 209–219.
[15] PADBERG M., RINALDI G., *A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems*, SIAM Review, 1991, 33(1), 60–100.
[16] GROMICHO J., PAIXAO J., BRONCO I., *Exact solution of multiple traveling salesman problems*, Combinatorial Optimization, Springer, Berlin–Heidelberg, 1992, 291–292.
[17] ARORA S., *Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems*, Journal of the ACM (JACM), 1998, 45(5), 753–782.
[18] LIN S., KERNINGHAM B.W., *An effective heuristic algorithm for the traveling-salesman problem*, Operations Research, 1973, 21(2), 498–516.
[19] GREFENSTETTE J., GOPAL R., ROSMAITA B., Van GUCHT D., *Genetic algorithms for the traveling salesman problem*. Proceedings of the First International Conference on Genetic Algorithms and Their Applications, Lawrence Erlbaum, New Jersey, 1985, 160–168.
[20] DIAS T.G., De SOUSA J.P., CUNHA J.F., *Genetic algorithms for the bus driver scheduling problem: a case study*, Journal of the Operational Research Society, 2002, 53(3), 324–335.

[21] BEASLEY J.E., CAO B., *A tree search algorithm for the crew scheduling problem*, European Journal of Operational Research, 1996, 94(3), 517–526.
[22] MARTELLO S., TOTH P., *A heuristic approach to the bus driver scheduling problem*, European Journal of Operational Research, 1986, 24(1), 106–117.
[23] GOEL A., *Truck driver scheduling in the European Union*, Transportation science, 2010, 44(4), 429–441.
[24] RAFF S., *Routing and scheduling of vehicles and crews: The state of the art*, Computers and Operations Research, 1983, 10(2), 63–211.
[25] POTVIN J.-Y., ROUSSEAU J.-M., *A parallel route building algorithm for the vehicle routing and scheduling problem with time windows*, European Journal of Operational Research, 1993, 66(3), 331–340.

optimization problem, hybrid algorithm, hyper-heuristic, Vehicle Routing Problem

Andrzej KOZIK*, Maciej DRWAL**,
Adam GONCZAREK**

ALGORITHMS COMPOSITION ENGINE FOR DOMAIN-SPECIFIC INFORMATION SYSTEMS

In this paper an outline of the Algorithm Composition Engine (ACE) for solving optimization problems for domain-specific information systems is designed. The main goal of ACE is to provide a unified framework of description and implementation of optimization problems and their corresponding data structures and algorithms, to enable their reuse in building domain-specific information systems and to gather domain-knowledge in an onthology. The framework of composition of algorithms provides patterns of description and implementation of optimization methods and techniques, so that they can be hybridized into compound algorithm to collaboratively solve complex and hard optimization problems.

1. INTRODUCTION

In this paper an outline of the Algorithm Composition Engine (ACE) for solving optimization problems for domain-specific information systems is designed. The premise of the ACE is that, for a given domain (e.g., transportation domain), an algorithm solving a new optimization problem can be built (semi-)automatically based on previously solved problems and on the gathered knowledge about the considered domain.

The main goal of ACE is to provide a unified framework of description and implementation of optimization problems and their corresponding data structures and algorithms, to enable their reuse in building domain-specific information systems. The way of composition of algorithms in ACE enables easy utilization of parallel programming techniques and computations aided by GPUs. The ACE provides two meth-

* Institute of Mathematics and Informatics, Opole University, pl. Kopernika 11a, 45-040 Opole, Poland.
** Institute of Informatics, Wroclaw University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland.

ods of algorithm composition – automatic, that algorithmically searches the space of structures of algorithms composition, and manual, enabling visual design of compound algorithm with the use of Graphical User Interface. Both ways are required for rapid prototyping of algorithms.

The presented engine is composed of two components: *on-demand* algorithm construction and a framework of composition of compound algorithms. In the further part of the paper we give an outline of each ACE component together with its application examples.

2. ON-DEMAND ALGORITHM CONSTRUCTION

The first ACE component is a framework of *on-demand* algorithm construction. The general idea of this framework is that each optimization problem from the considered domain is classified and precisely described using *terms* defined in the ontology. We assume, that terms are words or sentences given in a natural language and formulated by e.g., business analytics. A single concept of the domain can be described either by a single term, which can be a longer description, or can be a composition (sequence) of other terms appearing in the ontology. Terms are in one-to-one correspondence with mathematical modeling, so that each problem description given in the natural language can be translated (equivalently) into a mathematical model or programming language. Then, based on similarities with problems tackled so far and on the knowledge maintained in the domain ontology, a new problem can be defined with a mixture of existing terms, and in consequence, an algorithm can be automatically build or selected from an algorithm portfolio if it already exists.

The framework considers that each problem is stated as a description of *instance* (input data), *solution*, *constraints* and *objective function* using terms defined in the domain-ontology.

The instance describes the environment of the problem and given resources. Consider the case of transportation domain and the Vehicle Routing Problem with Paired Pickup and Delivery (VRPPPD). It can describe road-map with cities and roads between them. For each road are given its parameters, as length, travel cost, seasonal load, etc. It can also include location of parking lots, where the trucks can have a break if necessary. Instance describes a fleet of the transportation company and its characteristics. For example, it describes its capacity (constant, changeable (in the run on in depot only), cost of operation with respect to road parameters from a road-map, and its type, which dictate some driving restrictions (limit on drive-time, permitted days in some countries, etc). The instance describes freights to be served by a company's fleet. This description includes its source and destination on a map, times (or time-windows in some cases) of pickup and delivery (which are often negotiable), and capacity characterization (multidimensional).

The solution describes a schedule that solves a given problem. In the case of VRPPPD, it includes, for each truck, a list of cities it passes together with times of arrival and departure. For each such city the solution include a list of picked and delivered freights. Note, on this basis various parameters describing a solution can be computed, and a solution can be evaluated by an objective function.

The constraints describe restrictions that are to be met by a solution. The obvious ensure that each freight is served exactly by a single truck, and picked and delivered only once. Others ensure trucks capacities limits are not exceeded and time windows are met. For the case of heavy trucks, ride-time limits, forced by EU or local law, are to be definitely satisfied. In the case of transportation of handicapped or elderly persons passenger inconvenience must be considered. The transportation of perishable goods also requires maximum ride-time limits. Another typical constraints appearing in transportation are LIFO constraints, i.e., an additional restriction is placed on the loading of the trucks: at any delivery location, the item being delivered must be the item most recently picked up. This scheme reduces the loading and unloading times at delivery locations because there is no need to temporarily unload items other than the ones that should be dropped off. Sometimes the freight are exclusive (full-truck-load), i.e., no other customer can be visited between a pickup and its associated delivery location.

Note, that presented organization of instance, solution and constraints enables easy extension of characteristics by adding new fields into the description or new constraints by adding terms.

In the ACE framework each description (term) is mapped (via ontology) into corresponding data structures (input parameters and solution variables) and functions (constraints and objectives), stored in the framework as pieces of source code. Those pieces are then combined into complete *classes* of considered programming language, and then used within universal algorithms (metaheuristics and mathematical programming solvers) sharing common interfaces. The framework thereby provides patterns of description and implementation of certain concepts and elements of the problems such that they can be reused and interchanged between different algorithms. The translation flow between natural and programming languages is presented in Fig. 1.

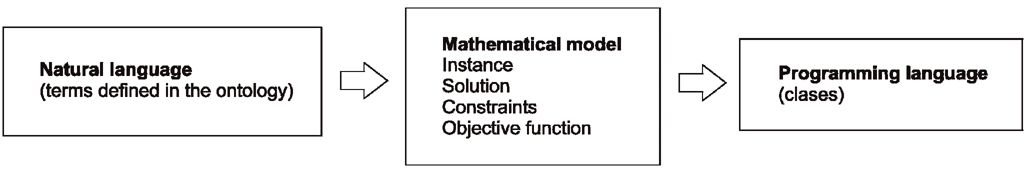


Fig. 1. Construction of optimization problem description and representation

However, the above defined classes (especially solution) is inconvenient to be used as a solution space for algorithms to solve such stated optimization problem. To this end, ACE extends such problem description by *representation*, which spans a solution space for combinatorial or mathematical programming algorithms. The representation class is defined (constructed) as a composition of basic data structures (vectors, permutations, lists, etc.), which populated, correspond to a solution of the given problem. Next, given the principles of algorithms operation, the representation class is extended with:

- translators – which code a problem solution into given representation and decode internal representation into solution, for which, in turn, satisfiability of constraints is verified and objective function is evaluated;
- operators – which perform moves from defined neighborhoods or crossover operations, used with modern metaheuristics;
- internal constraints – which can define additional constraints (if needed) imposed by the representation (usually, representation is chosen in such a way, that it discards constraints of the considered problem, as it guarantees them to be always satisfied).

For the considered VRPPPD, the representation can be a partition of the set of duplicated freights (one copy corresponds to pickup and second for delivery), where each partition subset represents freights attached to a corresponding truck. The truck serves attached freights (visits corresponding pickup and delivery locations) according to a permutation of its subset. Obviously, additional internal constraints include that each duplicated pair is attached to the same truck and, within each pair, pickup is before delivery (precedence constraint). Such representation enables to construct a variety of moves and neighborhoods, which can be efficiently explored by modern hybrid-metaheuristics.

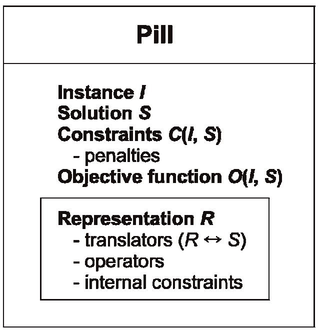


Fig. 2. Design of the Pill object

The classes corresponding to instance, solution, representation, constraints and objective are gathered into so-called *Pill* object, presented in Fig. 2, which encapsulates all the problem details and description, and instantiates an concrete input instance to solve and contains a solution (coded in representation) to be manipulated by optimization algorithms. Observe, in the form of Pill object we encompassed all problem information and functionality needed by algorithms designed to solve that problem. Especially Pill object defines all methods needed by algorithms to operate.

The framework provides implementations of various fundamental combinatorial data structures, matrices, graphs, etc. to be used as input data and solution variables, and many methods of multiobjective problem solving (mostly aggregation of several objectives into single criterion function), as well as general methods of tackling constrained problems by algorithms designed for unconstrained optimization (methods incorporating penalty functions of various types into criterion function).

3. COMPOSITION OF COMPOUND ALGORITHMS

The second ACE component is a framework of composition of compound algorithms, combining various methods and techniques, cooperating in solving complex optimization problems.

The framework is responsible for generating algorithm fulfilling given non-functional requirements, as algorithm performance (run-time and quality of delivered solution). The framework integrates various methods and techniques, that so far functioned separately:

- methods of management of the problem complexity [2] – methods of problem decomposition (primal, dual) into sub-problems (with problem relaxation) and into sub-domains of variables (eg., discrete and continuous parts) and methods of merging solutions of sub-problems (hierarchical, coordinate descent, etc.);
- methods of management of the scale of the problem - the instance decomposition methods, based on the divide and conquer approach (recursive division, multilevel improvement [2] and POPMUSIC [1]), aimed to efficiently limit the size of the solution space;
- methods of management of algorithms – methods that manage a portfolio of various algorithms (often with complementary properties) to design of so called hybrid metaheuristics [2], aimed to combine advantages of different approaches to solve optimization problems;
- methods of management of the structure and elements of compound algorithms, mostly using concept of hyper-heuristic [4] to algorithm selection, configuration and tuning (on-line and off-line) [5].

Within the framework, methods are stored as fragments of source code, that are exchangeable and reconfigurable (eg., different implementations of annealing schedule for simulated annealing [3], etc.). Moreover, they support multilevel nesting – they are defined as sequences of parts of prescribed functionalities (each implemented differently by many source code fragments), that can be intertwined (hybrydized) with other algorithms (from predefined subset). Hyper-heuristic concepts are implemented either as plugins of compound algorithms, aimed to on-line tuning of nested algorithms, or as a part of ACE responsible for automatically build and evaluate a structure of compound algorithms, automatic selection of nested algorithms and their off-line configuration.

The framework of composition of compound algorithms provides patterns of description and implementation of optimization methods and techniques, so that they can be nested and hybridized into compound algorithm to collaboratively solve complex and hard optimization problems. The framework can also handle hand-crafted algorithms, dedicated to the particular problem, if they are only written according to defined interface (can accommodate objects of defined pattern).

The essential assumption of ACE is that each algorithm defined the framework processes a single Pill object (see, Fig. 3) or many Pill objects of the same type. The first case is a *type P* of algorithm, the second is a type *mP*. Each algorithm performs manipulations on the contained solution either with use of methods predefined in the Pill object (e.g., universal metaheuristics [3], which use *move* and *crossover* operators) or in a specialized form (e.g., hand-crafted algorithms). This enables the ACE to include universal algorithms (e.g. metaheuristics) and to perform hybridization [2].

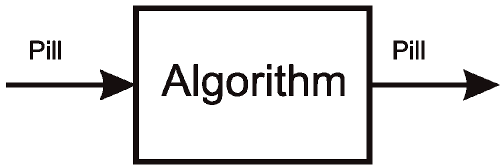


Fig. 3. Universal view of the algorithm in ACE

4. ALGORITHMS HYBRIDIZATION IN ACE

Based on the literature survey on hybrid algorithms, we described common methods of algorithm hybridization, which manage a portfolio of various algorithms aimed to combine advantages of different approaches to solve hard optimization problems. Such methods often have complementary properties as intensification or diversification of the solution search process, different trade-offs on feasibility/cost of solutions and performance (mostly time complexity). We classified such methods into four types:

- serial composition of algorithms,
- nesting of algorithms,
- interchange of algorithms parts,
- complex hybridization.

4.1. SERIAL COMPOSITION OF ALGORITHMS

Since each algorithm in the ACE framework is designed so that it processes a Pill object (or multiple Pill objects), algorithms can be composed into a sequence. The only limitation is that algorithms are processing the same Pill object and are either universal algorithms or specialized in the processing of the given Pill type. In the case of Pills for the same optimization problem, which differ in internal representations, algorithms sequence must be intertwined by *adapters*. Adapter, presented in Fig. 4, takes a source Pill *P* and generates its corresponding Pill *P'* with other representation, using included translators.

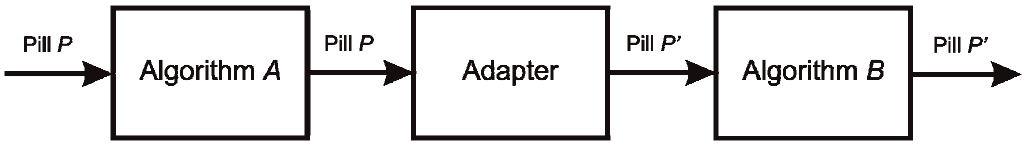


Fig. 4. Adapter between sequence of algorithms

4.2. NESTING OF ALGORITHMS

In the presented ACE framework, an algorithm is represented as a sequence of individual operations performed on the Pill object. Each such operation is precisely described and stored as a source-code fragment in the corresponding database. Nesting of algorithms is to insert into a sequence of master algorithm operations an execution of the slave algorithm which additionally performs a sequence of operations on a Pill solution. On the basis of literature survey, it follows that such hybridization is performed only in predefined places in the operation sequence of the algorithm, e.g. descent search algorithm is intertwined with evolutionary algorithm after forming a new population in *memetic search* algorithm [3]. Therefore, in the ACE framework, each algorithm precisely defines places in its operations sequence that can be intertwined with a slave algorithm. This is done either in a form of *delegates* or as a distinguished lines of source-code. Example of algorithm nesting id presented in Fig. 5.

4.3. INTERCHANGE OF ALGORITHMS PARTS

In the presented framework each algorithm and its operations have described functionality. Therefore each operation of the algorithm can be interchanged, given the new operation performs the same function. For example initial temperature selection, annealing schedule or stopping condition in simulated annealing can be interchanged in many ways. Similarly, the whole algorithms can be changed if they can operate on the same Pill object.

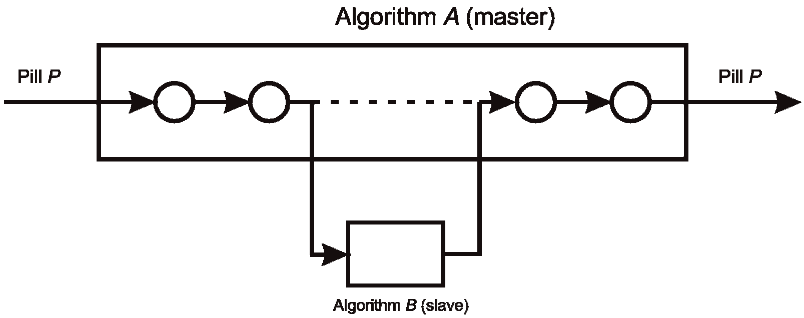


Fig. 5. Nesting of algorithms

4.4. COMPLEX HYBRIDIZATION

The complex hybridization methods are beyond the scope of the aforementioned cases and involve the use of a decomposition and parallelization methods. In the proposed framework implementation of this type of hybridization is supported by the so-called. functional blocks. Below we describe functional blocks of the ACE.

The P → MP → P functional block, presented in Fig. 6, implements the functionality of cloning the input Pill object into a multi-pill object (mP) with a width set as a parameter. Then, the code is executing a delegate (slave algorithm) algorithm of type mP (e.g., an evolutionary algorithm). To the original Pill object the best of solutions present in the vector is cloned, but other aggregation function are also possible.

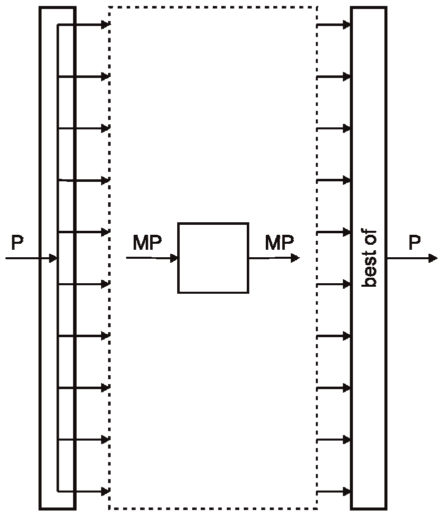


Fig. 6. The P → MP → P functional block

The mP functional block, presented in Fig. 7, implements the functionality of the mP algorithm in such a way that for each individual Pill object of the multi-pill vector a separately defined delegate (potentially different algorithm) is performed. This block is used to implement the memetic algorithms, i.e., to start local search algorithm for each solution of the population [3].

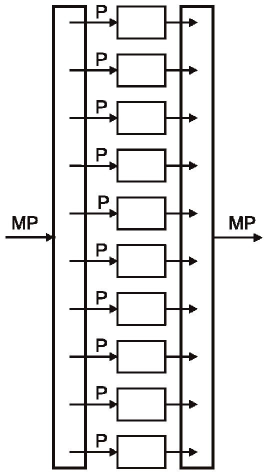


Fig. 7. The mP functional block

The R/R' functional block is implemented as a pair of blocks that perform problem relaxation (block R) and reproduce the solution of the problem on the basis of the solution of the relaxed problem (block R′).

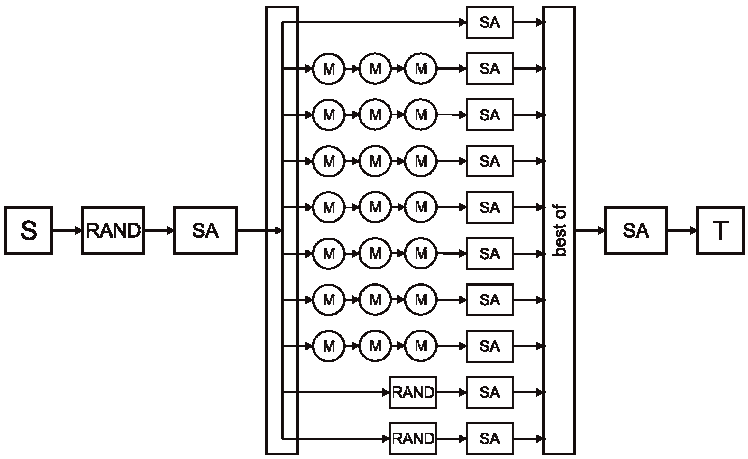


Fig. 8. An example of compound algorithm

An example of the compound algorithm composed in ACE is presented in Fig. 8. The algorithm starts in an S node, generates a random solution by a specialized algorithm and then perform a simulated annealing algorithm. Such obtained solution is cloned to perform parallel processing by different algorithms (ranging from simple random moves, to more advanced simulated annealing). Best of such obtained solutions is once again improved by simulated annealing and returned in the T node.

5. CONCLUSIONS

Based on the observation, that algorithm composition shares the idea with SOA service composition, we implemented ACE such that compound algorithms are realized in Service Oriented Architecture of Algorithms (SOAoA), in which each fundamental algorithm is implemented as a service (AaaS), and compound algorithms are composite services hosted by an orchestration engine. This provides scalability to the system and introduces additional hyper-heuristic layer to the engine, as service selection can act as an algorithm selection step.

ACKNOWLEDGMENTS

The research presented in this paper was partially supported by the Polish Ministry of Science and Higher Education and the European Union within the European Regional Development Fund, Grant No. POIG.01.03.01-02-079/12.

REFERENCES

[1] ALVIM A., TAILLARD E., *POPMUSIC for the world location routing problem*, EURO Journal on Transportation and Logistics, 2013, to appear.
[2] BLUM C., AGUILERA M., ROLI A., SAMPELS M., *Hybrid Metaheuristics. An Emerging Approach to Optimization*, Studies in Computational Intelligence, 2008, Vol. 114.
[3] BLUM C., ROLI A., *Metaheuristics in combinatorial optimization: Overview and conceptual comparison*, ACM Computing Surveys, 2003, Vol. 35, No. 3, 268–308.
[4] BURKE E.K., GENDREAU M., HYDE M., KENDALL G., OCHOA G., OZCAN E., QU R. *Hyper-heuristics: a survey of the state of the art*, Journal of the Operational Research Society, 2013, Vol. 64, 1695–1724.
[5] HUTTER F., HAMADI Y., HOOS H.H., LEYTON-BROWN K, *Performance prediction and automated tuning of randomized and parametric algorithms*, Lecture Notes in Computer Science, Principles and Practice of Constraint Programming, 2006, Vol. 4204, 213–228.

business process optimization, service-oriented systems,
business process optimization methodology,
business problem optimization platform

Krzysztof JUSZCZYSZYN, Grzegorz KOŁACZEK,
Paweł ŚWIĄTEK, Adam GRZECH*

AGILE METHOD OF SOFTWARE DEVELOPMENT FOR RESOURCES MANAGEMENT OPTIMIZATION PROCESSES IN TRANSPORTATION

One of the most crucial element for final success in any type of business is resource management. It is also true for transportation companies. In transportation, the basic set of decision concerns schedules and routes. There are also other types of decisions as decisions concerning technical overviews, fleet modernization, etc. Very often these decisions are elaborated by an expert. However due to the scale of the problem and the dynamics of the business environment it becomes more and more hard to select the best or almost the best variant. This is why modern transportation companies require efficient decision supporting systems. The paper presents the main steps for the agile method of software development for resources management optimization. The method has been elaborated and is evaluated during collaboration with several transportation companies.

1. METHOD FOR RESOURCE MANAGEMENT OPTIMIZATION

According to the dictionary definition, method is a standardized approach to solving problems [1]. In the case of this work, which focuses on issues related to decision-making support systems in transportation organizations, the overall context of the method is presented in the following context.

In an organization, decisions are made for the acceptance of orders and allocation of resources to handle accepted orders. Decisions are made within the decision-making process, which should be identified and described for each organization and for each type of decision. The aim of the Method is to produce methods and tools to

* Faculty of Computer Science and Management, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland, e-mail: krzysztof.juszczyszn@pwr.edu.pl, grzegorz.kolaczek@pwr. edu.pl, pawel.swiatek@pwr.edu.pl, adam.grzech@pwr.edu.pl

support decision-making [2]. The tools which will be developed are intended to support decision making by solving optimization problems. Hence, the purpose of Method is:

- identification and modeling decision-making process for accepting orders and allocation of resources for the handling of orders;
- identification and modeling orders carried out by the organization;
- identification and modeling resources used to handle orders;
- modeling of individual tasks of making decisions about accepting orders and the allocation of resources to execute orders, including:
  - types of events that enforce the need for decision-making,
  - data to be taken into account when making decisions for each of the types of events,
  - quality evaluation of each decision;
- mapping of individual tasks of making decisions on formally defined optimization problems whose solution will support decision-making in the organization.

1.1. MAIN BENEFITS OF THE PROPOSED METHOD

The most characteristic elements of a developed Method and features which differentiate the approach proposed to other solutions are:

- agility – a set of appropriate methods and tools is delivered to make fast and accurate identification of the relevant decision-making processes for the organization,
- knowledge accumulation – knowledge derived from solved previous problems is collected and made available in a systematic way to analyze further organizations,
- service orientation – the use of SOA paradigm makes it possible to improve the implementation phase and integration of software solutions created with the existing and previously used in the systems used by organization.

The key benefits of application of the proposed Method in the process of software production for the analysis, planning and optimization of resource management processes of the organization are:

- decrease of time required for the development, implementation and adaptation of the system by allowing the use of standard and specialized libraries processes and algorithms dedicated to the areas of transport management,
- cost reduction of implementation of the system by simplifying and speeding up the analysis of the requirements necessary to develop or adapt the system,
- optimization of management processes and resources utilization through the selection of appropriate optimization methods dedicated to specific customer requirements,

- simulation of the impact of selection different evaluation criteria to optimize utilization of resources
- integration of optimization modules with functionality offered by monolithic systems through the use of service-oriented paradigm (Service Oriented Architecture) and enterprise data bus (Enterprise Service Bus).

2. THE METHOD OF IDENTIFYING AND ANALYZING THE OPTIMIZATION REQUIREMENTS OF THE ORGANIZATION

In the Method the organization is described in the way where at the beginning a stream of transportation orders is defined and then, the organization develops a plan how these orders will be handled and finally the organization assigns its resources in such way that it is possible to handle the selected orders. Handling orders is defined as a completion of at least one so-called Transport Task (TT), which has been shown in Fig. 1.

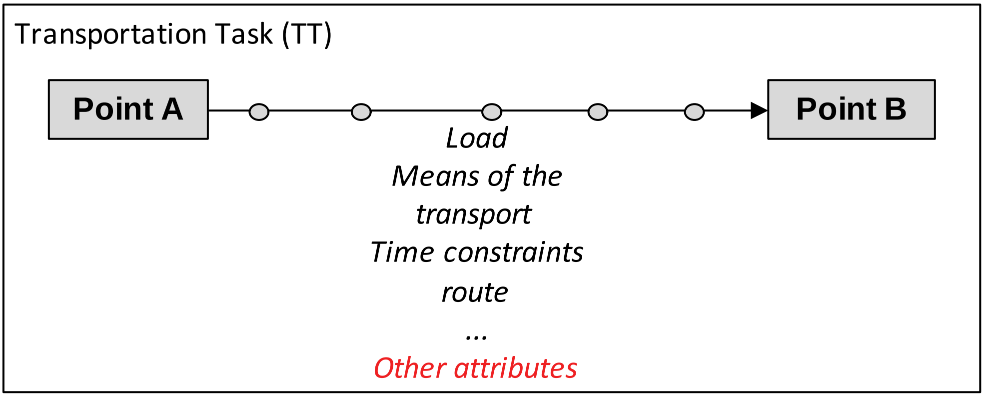


Fig. 1. Transportation Task

TT is defined as the act of transportation of cargo with application of some type of the means of transport along (fixed or not) route between the required locations. It has been assumed, moreover, that the handling orders on many goods and location can be represented by a set of TT.

In practice, a complete description of TT needs to specify, its attributes relevant to the orders carried out by the Organization (operators of means of transport, time constraints, etc.).

At any time we deal with a situation in which the organization carries out some (active) orders, performing numerous TT. It is also possible that while some TT is carried out a new orders appear.

Development of a plan describing how the accepted orders will be handled, requires a number of decisions concerning the (possible categories, based on past interviews of analyst):

- acceptance or rejection of orders;
- decision about splitting or consolidation of some orders;
- route design;
- resource allocation (means of transport, their operators, etc.);
- determining timetables;
- response to random events, e.g., closing section of the route, an accident on the highway and associated delays, failure of means of transportation, unavailability of operators (illness, vacation), delays caused by contractors, delays caused by the operator of transport infrastructure, ...;
- ..., etc.

In general, these decisions are not independent, in addition we do not know in what order shall be taken in the organization and what are their priorities. The final product of decisions is set of TT, the implementation of which is equivalent to handling selected orders. One of the main result of the Method implementation should be exact description how to model this decision-making process of allocation of resources.

2.1. DICTIONARIES

The Method defines four dictionaries to describe any organization:

- Organization Dictionary (OD),
- Decision Making Tasks Dictionary (DMTD),
- Models Dictionary (MD),
- Domain Dictionary (DD).

The scope and purpose of each above mentioned dictionary is as follows:

1. Organization Dictionary contains terms to describe:
   - served by the organization transport tasks (TT) and classes of these tasks,
   - resources used by the organization in the implementation of transport tasks,
   - orders for transportation tasks supported by the organization,
   - terms appropriate for the description of the above.
2. Decision Making Tasks Dictionary (DMTD) – contains terms to describe:
   - list of classes related to decision-making processes,
   - terms appropriate for the description of the above mentioned classes of tasks.

3. Models Dictionary (MD) includes terms to describe:
   - formal models of decision-making tasks and related:
     - criteria
     - constraints
     - variables
     - related terms
4. Domain Dictionary (DD)
   The purpose of the Domain Dictionary is to gather knowledge about the field of transport, derived from previous cases and can be characterized by the following proprieties:
   - structure corresponding to the structure of Organization Dictionary (transport tasks, orders, resources),
   - the hint system allows the analyst to identify the cases in which it is possible to use existing structures in the description of the new organization.

2.2. THE ROLE OF THE ANALYST

The main task of the analyst is to conduct an interview and complete the above defined dictionaries with corresponding knowledge about the organization. This step should be done in such a way that it will be possible to define a formal model of the selected decision making tasks. The model will be used in defining of the optimization problem and the information from the interview should allow development of an appropriate solution algorithm.

In parallel, analyst may recognize a broader business context of the problem under consideration, and fill up corresponding dictionaries. However, the priority is to identify the decision-making process, and to choose the most important (from the point of view of the organization) decision-making task or tasks.

Where the described organization is not the first case when the Method has been applied, there is the possibility that the identified transport tasks, and decision-making tasks descriptions already exist in the repository of knowledge (dictionaries, in particular DD). In this case, analyst at each stage of analysis performs the following steps:

- verify that the selected term is appropriate for the organization that is currently under consideration (if not - define a new and update dictionaries or change the definition of the term)
- verify that the terms occurring in the sub-tree of the selected term are appropriate for the organization under consideration; if appropriate - follow for each term the previous step and remove improper terms.

In the most convenient situation, the above procedure will allow the use of previously developed models, algorithms and code for a new customer.

2.3. PROCEDURE

Dictionaries described in previous section are used to identify, formulate and formally define the tasks of decision making in organizations by following procedure, which is an important element of the Method (Fig. 2).

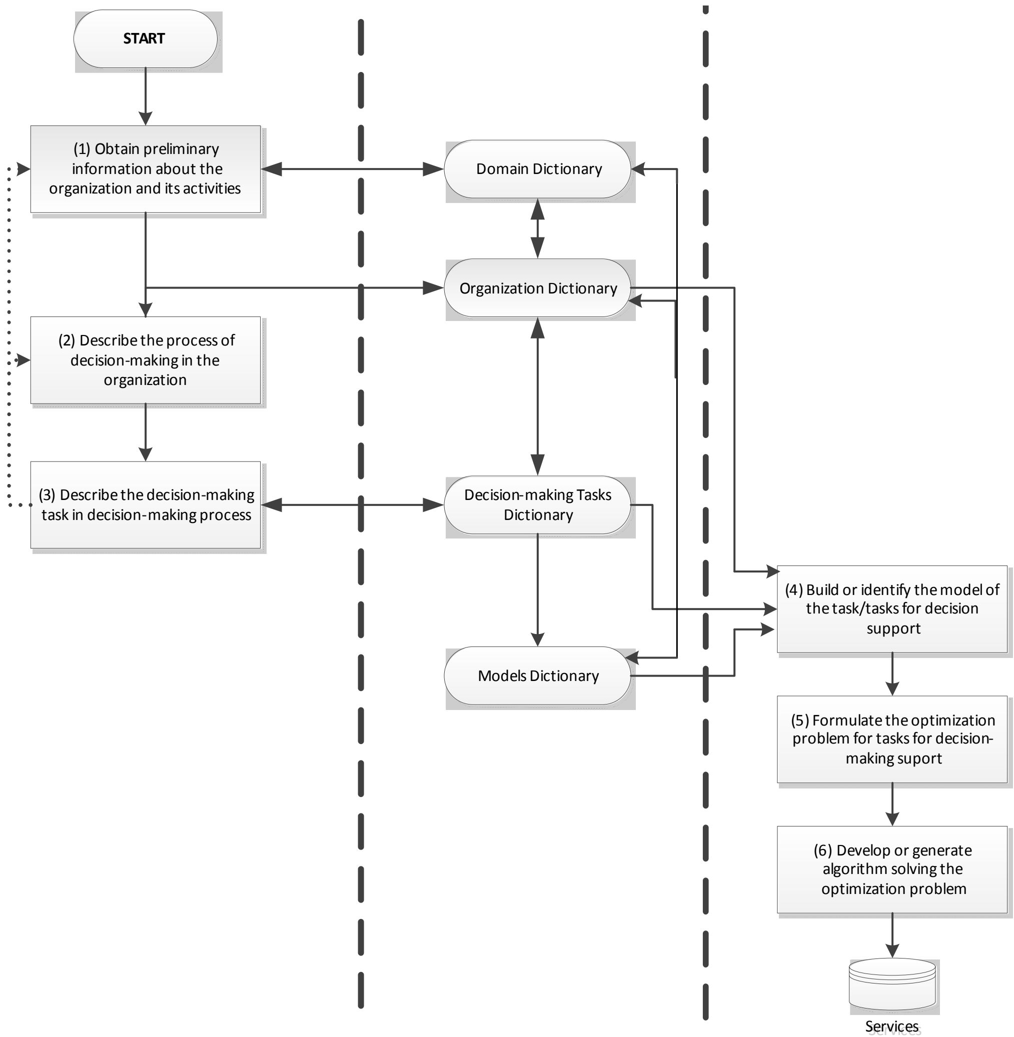


Fig. 2. Identification of optimization requirements

Step 1. Obtain preliminary information about the organization and its activities
In this step, the analyst is trying to pre-identify on the basis of publicly available sources the scope of transport activities of the Organization (Fig. 3).

Using the collected information analyst determines a direction of direct interview and if available, the basic structures of dictionaries are filled with relevant data and used in further steps to identify optimization problems important for the organization.

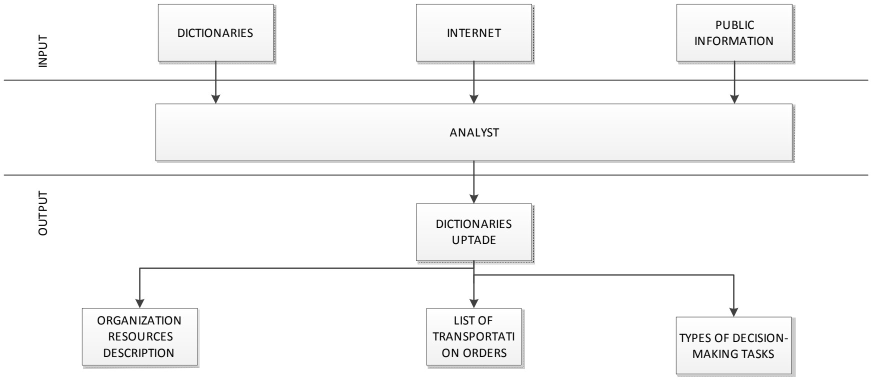


Fig. 3. Step 1 – Input/Output

Step 2. Describe the process of decision-making in the Organization

This step is performed by an analyst with application of dedicated tools supporting the interview process. The first goal of analytics is to verify the validity of pre-selected requirements of organization during the first step (Fig. 4). Then, the analyst should identify the most important elements of the decision making process and to fill dictionaries with the knowledge of the transport orders and resources of an organization. This knowledge comes from the information and documents provided by employees of the organization.

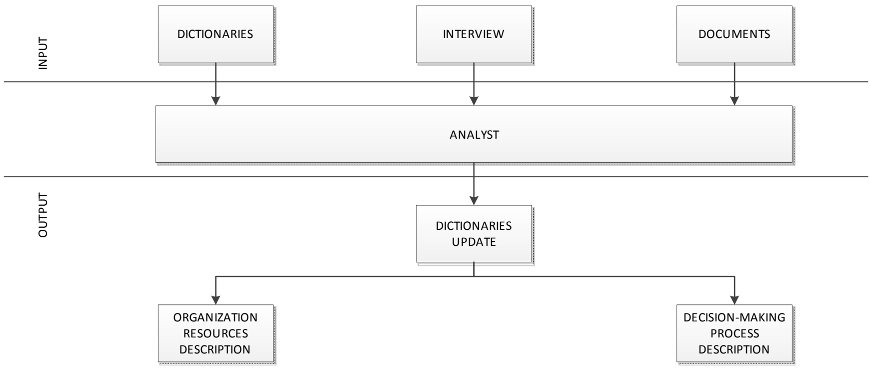


Fig. 4. Step 2 – Input/Output

Step 3. Describe the various decision-making tasks in decision making
The third step leads to completing gathering of knowledge of the decision-making tasks identified in the previous step. Tools supporting the work of the analyst will make possible an assessment of the completeness and consistency of the collected knowledge in dictionaries. In particular, the analyst gets feedback indicating whether introduced the results of the analysis enable the transition to the next step of the Method. While the next step is to select/develop an appropriate model for decision support. Decision-making tasks must be described by specifying their inputs, outputs, criteria, constraints, variables and parameters (Fig. 5).

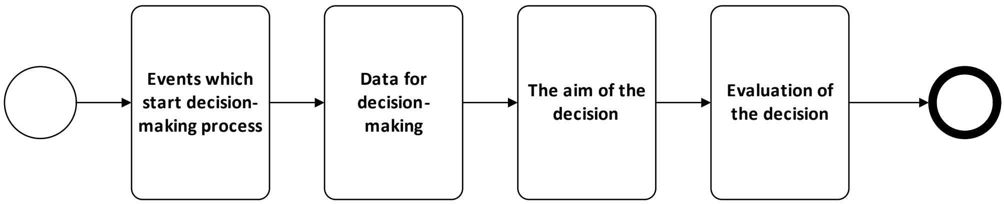


Fig. 5. Decision making process

Step 4. Build or identify the model of the task / tasks for decision support
This step is performed after the steps performed by the analyst. In this step, the specialized tools allow verification if it is possible to apply previously developed models of decision support tasks to provide the solution

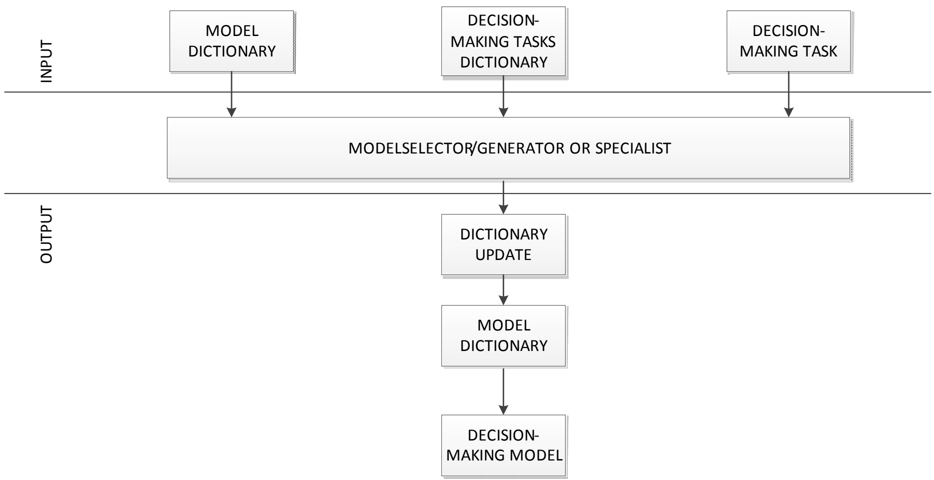


Fig. 6. Step 4 – Input/Output

to the optimization problems of the currently analyzed organization (Fig. 6). In case of unavailability of the appropriate model, tools allow to determine whether it is possible to automatically generate such a model, or whether it is necessary to involve a specialist. If it necessary, the specialist will have to define adequate for the organization model using the results of work of the analyst made available in the form of dictionaries and repositories.

Step 5. Formulate the optimization problem for tasks / activities support decision--making

In the fifth step of the Method, optimization problem is formulated for the selected tasks of decision support. The optimization problem is formulated automatically using dictionaries and repositories, or by an expert (Fig. 7).

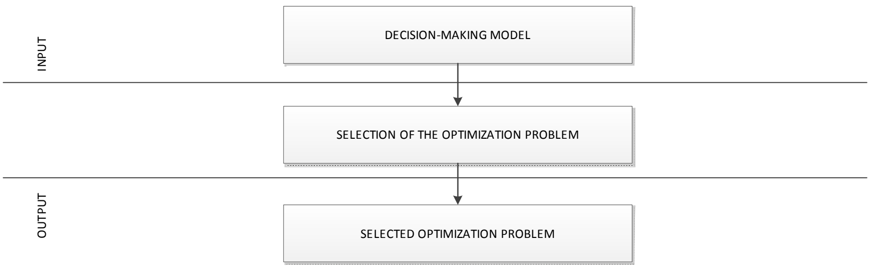


Fig. 7. Step 5 – Input/Output

Step 6. Develop or generate algorithm which solves the optimization problem

In this step of the Method the corresponding tools tries to generate an algorithm which solves the selected optimization problem. In the absence of auto-generated algorithm, the algorithm should be proposed by an expert (Fig. 8).

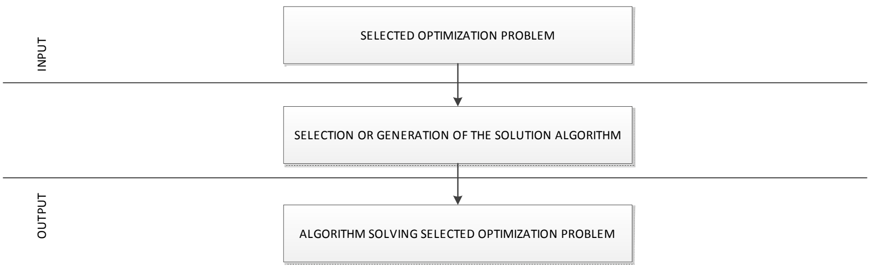


Fig. 8. Step 6 – Input/Output

3. CONCLUSIONS

The paper presents the overview of the key features of agile Method of software development for resources management optimization processes in transportation. The key benefits of application of the proposed Method in the process of software development have been characterized. Also the brief information about the main steps leading from the organization requirements identification till provision of the solution algorithm has been presented. As the described Method is still under development and finally will be evaluated in cooperation with ten selected transportation organizations, the final details of the Method will be presented in the nearest future.

ACKNOWLEDGEMENTS

The research presented in this paper was partially supported by the Polish Ministry of Science and Higher Education and the European Union within the European Regional Development Fund, Grant No. POIG.01.03.01-02-079/12 and within European Social Fund.

REFERENCES

[1] BRZOSTOWSKI K., DRAPAŁA J., GRZECH A., ŚWIĄTEK P., *Adaptive decision support system for automatic physical effort plan generation: data-driven approach*, Cybernetics and Systems, 2013, Vol. 44, No. 2/3, 204–221.

[2] ŚWIĄTEK P., STELMACH P., PRUSIEWICZ A., JUSZCZYSZYN K., *Service composition in knowledge-based SOA systems*, New Generation Computing, 2012, Vol. 30, No. 2/3, 165–188.

PART 2

KNOWLEDGE ENGINEERING AND ITS APPLICATION IN DECISION SUPPORT SYSTEMS

multi-project planning, modeling,
constraint programming,
project portfolio rescheduling,
decision support system

Izabela NIELSEN*, Robert WÓJCIK**,
Grzegorz BOCEWICZ***, Zbigniew BANASZAK****

TOWARDS CONSTRAINT PROGRAMMING DRIVEN METHODOLOGY FOR ONLINE MULTI-PROJECT PLANNING AND CONTROL

Declarative approaches to systems and/or process modeling promise a high degree of flexibility. Constraint Programming (CP) is an emergent software technology for declarative description and effective for solving large combinatorial problems especially so in areas of integrated production planning. In that context, CP can be considered as a well-suited framework for development of decision making software supporting small and medium size enterprises in the course of multi-product, i.e. multi project-like, production. The considered problem regards finding a computationally effective approach aimed at simultaneous routing and allocation as well as batching and scheduling of a new project subject to constraints imposed by a multi-project environment. The aim of the paper is to present the CP modeling framework providing a methodology for CP-based decision support systems' design allowing one to answer whether a given production order specified by its cost and completion time can be accepted in a given manufacturing system given by available production capability and/or what manufacturing systems capability guarantee completion of a given production order under assumed cost and time constraints.

---

* Department of Mechanical and Manufacturing Engineering, Aalborg University, Aalborg Denmark, e-mail: izabela@m-tech.aau.dk

** Institute of Computer Engineering, Control and Robotics, Wrocław University of Technology, Wrocław, Poland, e-mail: robert.wojcik@pwr.wroc.pl

*** Department of Electronics and Computer Science, Koszalin University of Technology, Koszalin, Poland, e-mail: bocewicz@ie.tu.koszalin.pl

**** Department of Business Informatics, Warsaw University of Technology, Warsaw, Poland, e-mail: Z.Banaszak@wz.pw.edu.pl

1. INTRODUCTION

Current manufacturing environment can be characterized in terms of many factors but the key one for companies confronting the challenge of remaining competitive in an era of globalization is undoubtedly the capability of fast and accurate decision making, especially so in the project management domain. Most companies, particularly small and medium sized enterprises (SMEs) have to manage various projects sharing a pool of constrained resources and taking into account various objectives at the same time. It is known from surveys conducted (Lova et al., 2000), that about 80% of companies have to deal with multiple projects, what corresponds to the other data stating that about 90% of all projects occur in a multi-project context.

Since the project management problems belong to a class of NP-complete ones, new methods and techniques aimed at real-life constraints imposing on-line decision making are of great importance. Such methods enhancing an on-line project management, and supporting a manager in the course of decision making, e.g. in the course of evaluation whether a new order can be accepted to be processed in a multi-project environment of a manufacturing system at hand or not, could be included into Decision Support Systems (DSSs) tools integrated into standard project management software. The main objective of a DSS aimed at multi-product production flow planning is the coordination of processes and activities related to work order processing, i.e., the transportation, inventory management, warehousing and production. In other words the goal is to achieve a well-synchronized behavior of dynamically interacting components, where the right quantity of the right material is provided in the right place at the right time.

In that context, the aim is to provide an effective methodology combining genetic algorithms (GA), Constraints Programming (CP) (encompassing different types of criteria: stochastic/deterministic and financial/due date) and fuzzy sets theory for multi-project planning and online control which mainly focuses on scheduling and rescheduling activities and allocating resource for multiple projects. The methodology furthermore contains a unified solution search engine cooperating with CP to reduce the computational time in finding solutions.

In multi-project planning the main focus is on deciding on a schedule for all activities of projects and allocating resource in order to finish projects before their deadlines. One of our objectives is to propose a method that allows generating a schedule for set of orders execution with resource allocation for a given period of time, guaranteeing a solution meeting a set of enterprise specific goals. Furthermore, due to the fact that unpredicted circumstances frequently happen during the execution of orders, it is required to control them in an online mode in order to quickly react to these circumstances. Among the activities of online control, rescheduling the activities of multiple projects and reallocating resources are critical. Therefore, another of our objectives is to develop a method for rescheduling the project portfolio and reallocating resources with the consideration of budget, cash flow, resource capacity, new projects, etc.

In practice, the generated schedules are subject to disturbances due to *uncertain data* (*fuzzy data*) such as the duration of activities and *random (stochastic) events* e.g. unforeseen situations and disturbances influencing the order specification (usually described by a probability density function). This makes it critical during the execution to predict in real time how one activity influences others with regards to resource consumption, estimated deadline and financial liquidity.

A method enabling the minimization of the impacts of these disturbances employs the proposed *reverse planning problem* concept. Reverse problems become critical, when the direct planning approach represented as a direct planning problem fails to find a solution meeting all constraints. Reverse problems address this issue by establishing which levels the constraints should as a minimum have to satisfy e.g. a maximum completion time of a project. In the current state, financial matters are not integrated in neither direct nor reverse project planning methodologies.

The remainder of the paper is organized as follows: Section 2 provides an overview of related works. Section 3 presents the problem formulation, and a new methodology aimed at online planning and control under uncertain conditions is presented in Section 4. Concluding remarks are presented in Section 5

2. RELATED WORK

The scheduling of multi-stage batch processes has received significant attention from researchers in the process systems engineering community. Existing methods assume that routing and allocation as well as batching and scheduling decisions are made independently, i.e. each production order is treated as an activity network and is assigned to processing units, and then divided into a number of batches (batching), and sequenced (scheduling).

In recent years, focus has come to be on the Resource-Constrained Project Scheduling Problem (RCPSP) which involves scheduling project activities subjected to temporal and resource constraints, while minimizing the total project duration. Two different approaches have been taken in solving the RCPSP. The first approach includes exact algorithms, which produce optimum solutions. The second approach is comprised of algorithms that provide admissible (or suboptimal) solutions refer to [3-5] and [7-10] These all demonstrate good results for the limited objective of resource allocation, but fail to truly support decision makers on all stages of project execution.

Very limited work focuses on the joint technological processes, transportation routing and financial [1]. Furthermore, there is another aspect of the addressed problem, namely multi-criteria decision making under uncertain conditions. Fuzzy multi-criteria decision making is primarily adopted for selecting, evaluating and ranking of alternative solutions to problems [11]. Studies conducted so far on declarative models implemented in fuzzy sets framework [2], show that the proposed concept of *reverse* projects portfolio planning provides a promising alternative to commercially available ones. Therefore the new methods and techniques addressing the impact of real-life constraints on online decision making are of great importance [5]. To do this in a way compatible with real life settings necessitate the use of stochastic and fuzzy logic frameworks [8]. The fuzzy model of project portfolio online control can be specified as a declarative one and then implemented using constraint programming techniques and finally implemented as a decision support system [3].

Regardless of its character and scope of business activities a modern enterprise, has to build a project-driven development strategy in order to respond to challenges imposed by growing complexity and globalization. Managers need to be able to utilize a modern DSS as to undertake optimal business decisions in further strategic perspective of enterprise operations. In this context this contribution covers various issues of decision making while providing the concept of Constraint Programming (CP) as well as modeling and designing of decision support tools aimed at management in SMEs and/or the associated Extended Enterprise [5].

3. PROBLEM FORMULATION

A production system is given in which a set of projects (the production orders portfolio) has to be executed. The state of the system encompasses its resources allocation to projects planned for execution. However, some parameters describing the system and project portfolio are uncertain (fuzzy), e.g. durations of activities.

Traditionally-stated multi-criteria planning problems formulated as direct ones address standard questions as: is it possible to undertake the given project portfolio under a given resource availability while guaranteeing disturbance-free execution of activities? Such a formulation, however, may cause rejecting projects, which could actually be approved by the system if a satisfactory solution could be found by changing the levels of the constraints.

Therefore, an alternative problem statement formulated as a reverse planning problem can be as follows: Which values of the system parameters guarantee that the set of orders will be completed while giving a certain set of values for performance indexes? In case of any new event, caused e.g. by including a new project, a new system state has to be considered to determine a new project portfolio schedule. In that context, the proposed approach involves solving both direct and reverse problems for systems where project portfolios (specified by fuzzy data) change over time as a result of random occurring events.

Proposed methodology should provide the following contributions:

- a method for follow up planning and online control subject to financial, time and resource capacity constraints, in an uncertain multi-project environment,
- an approach solving the reverse problem for project portfolio planning by integrating stochastic, fuzzy logic and the constraint programming methods.

It seems obvious, that not all behaviors (functionalities) are reachable under constraints imposed by a given system's structure. The similar observation concerns the system's behavior that can be achieved in systems possessing specific structural constraints. So, since system constraints determine its behavior, both the system structure and the desired behavior have to be considered simultaneously. In that context, our contribution provides a discussion of some solubility issues concerning structural properties providing conditions guaranteeing assumed system behavior (straight problem formulation) as well as behavioral requirements imposing conditions that have to be satisfied by system structure (reverse problem formulation).

4. METHODOLOGY

Proposed methodology is divided into two stages namely the *initial planning* and *online control stages* as shown in Fig. 1.

The first stage problem can be formulated as a multi-project scheduling with resource allocation problem. This is an NP-complete problem and hence a genetic algorithms (GAs) based approach is proposed for solution purposes. Following the procedure of GA [6], a set of solutions is generated and the chromosomes are represented for the scheduling of the tasks of projects. A chromosome includes a set of sub-strings where each sub-string is the schedule of tasks for a project. A proposed heuristic algorithm is executed for each chromosome in order to allocate the resources for each task. The total cost is calculated for each chromosome and it is used to evaluate the fitness of the chromosome. After this GA operations including crossover, mutation, and selection are executed to generate the new generations that contains a new set of chromosomes. The procedure is repeated until one of termination criteria (e.g. maximum number of generations) is satisfied [6]. The solution is determined as the best one found so far from this GA-based algorithm. The result from the first stage will then be used to form the initial state for the second stage as indicated by the red arrow in Fig. 1.

In the second stage, the solution for a state is valid only for the particular system's state guaranteeing the conditions determining production orders execution are constant. However, the current system state changes when a new event appears, for example there is not enough resource to finish a certain project before its due date or the deadline. Therefore, there is a requirement for orders rescheduling and resource reallocating. The solution has to be found quickly to be able to react to the frequent disturbances in daily activity. The idea of the *online control* is presented in Fig. 1. Project portfolio rescheduling and resource reallocating take into account the stochastic characteristics of the new processes submission.

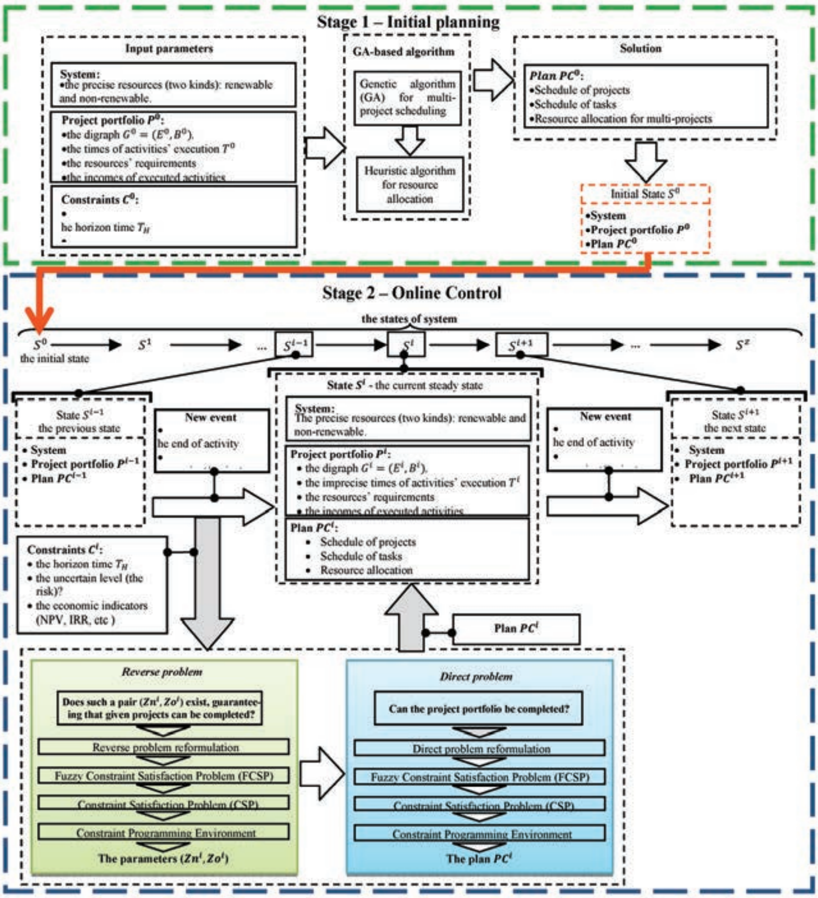


Fig. 1. The idea of scheduling and rescheduling

When an unforeseen event occurs, it can make the current schedule infeasible. Thus, it is necessary to reschedule project portfolio and to reallocate resource in online mode.To be able to achieve this requires solving the problem in two steps: first, as a reverse problem and then as a direct problem. The reverse problem is formulated to establish the range of values of parameters guaranteeing a feasible plan exists. Therefore, the result from the reverse problem will guarantee finding a feasible solution in the direct problem and significantly reduce computational time.

The solution of the reverse problem will be used as input parameter for the direct problem which aims to find a new plan for projects with minimum cost. Due to the high complexity of planning problems it is assumed that *reverse/direct problems* are represented in constraint satisfaction formalism to give the current state of the system.

The advantages of using constraint programming technique are: 1) it reduces the search space and therefore, it can find a feasible solution in the short time which is required for online control; 2) it can be implemented in standard software such as ILOG, PROLOG [2]. Therefore the problems considered are stated in a new form of constraint satisfaction problems CSP, i.e., Fuzzy Constraint Satisfaction Problem (FCSP). Unfortunately standard CP platforms are not able to cope with fuzzy decision variables. For this reason a FCSP transformation to standard CSP form is proposed. FCSP is defined as a triple: fuzzy decision variables, e.g. duration of activities, their fuzzy domains (families of the membership functions), and fuzzy constrains (fuzzy relations) a definition of their relationships representing for instance financial liquidity, etc.

In the proposed approach each fuzzy decision variable is described by a set of precise parameters (discrete alpha-cuts) and each fuzzy constraint is represented by the set of constraints describing relationships between parameters of fuzzy variables. This kind of representation allows solving FCSP as a standard CSP composed of parameters of fuzzy variables and their constraints.

Our approach using the novel form of fuzzy variables representation (*the discrete α-cuts*) enables this and finally enables us to *solve reverse/direct problems* (modeled in terms of FCSP) while using standard CP environments. Therefore, the approach assumes the portfolio rescheduling takes place at states, where FCSP (representing *reverse/direct problem* of the considered *system*) is transformed to CSP and then solved using CP-based techniques. Its distinct advantage is that it separates the problem statement and its resolution methods. In addition, integrating the cash flows and the resource allocations with data describing the stochastic nature of possible disturbances are considered in the online control.

The above solution procedures will be integrated in prototype software shown in Fig. 2.

An interface module is built for interaction with users. This module allows users to input data and indicates data as well as results of the multi-project planning and control problem see Fig. 3.

The data module is used to read/write data from/to the database. The procedure to find the initial plan is built in the initial planning module (IP module in Fig. 2). The procedures to solve the direct and reverse problems are built in the online control module (OC module in Fig. 2).

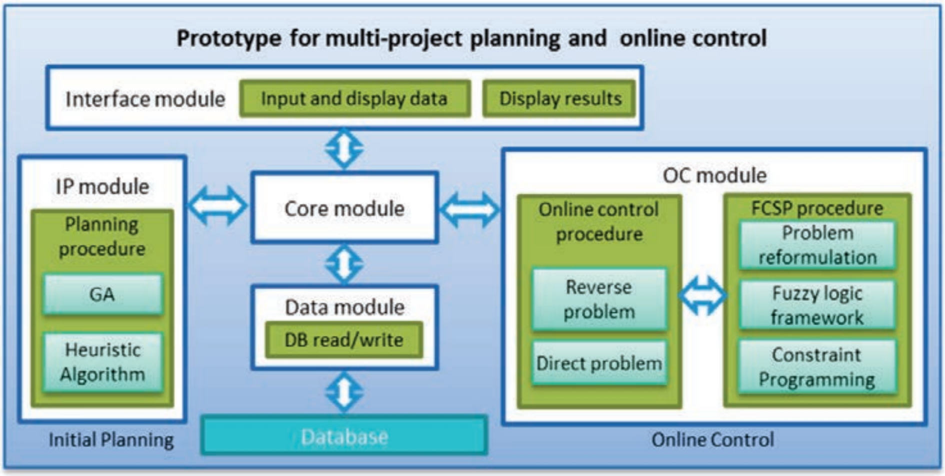


Fig. 2. Description of prototype

A core module is built to link the interface, data, IP and OC modules. The prototype software works as follow: all required data is input to the database. Following this data is read from databases through the data module and the IP module is executed to find the initial schedule and form the initial state S^0 (Fig. 1).

The solution of this is shown by the interface module and saved to the database by the data module. When there is an uncertain situation, user input to the system and OC module is initiated. The prototype software will continue until the end of the planning horizon.

Idea standing behind of interactive interface module employs a navigation multi-board concept shown in Fig. 3. Its solution assumes hybridization of *Drag and Drop*, *Touch Screen Panel*, and *Virtual Table* technologies.

The menu composed of a set of tabs and folders allows one to specify parameters and decision variables describing both enterprise's capability (e.g., following from its structure and possible ways of work flows organization), and requirements imposed by production orders at hand (determining for instance the batches size, production cycles, work-in-progress, and so on).

Dependent on the kind of decision problem considered the relevant tabs are selected and structured on the board as to encompass the one of the following problems formulation:

- a straight planning problem (e.g., Is it possible to undertake the given project portfolio under a given resource availability while guaranteeing disturbance-free execution of activities?),
- a reverse planning problem (e.g., Which values of system parameters guarantee that the project portfolio at hand will be completed while following constraints assumed performance indexes values?).

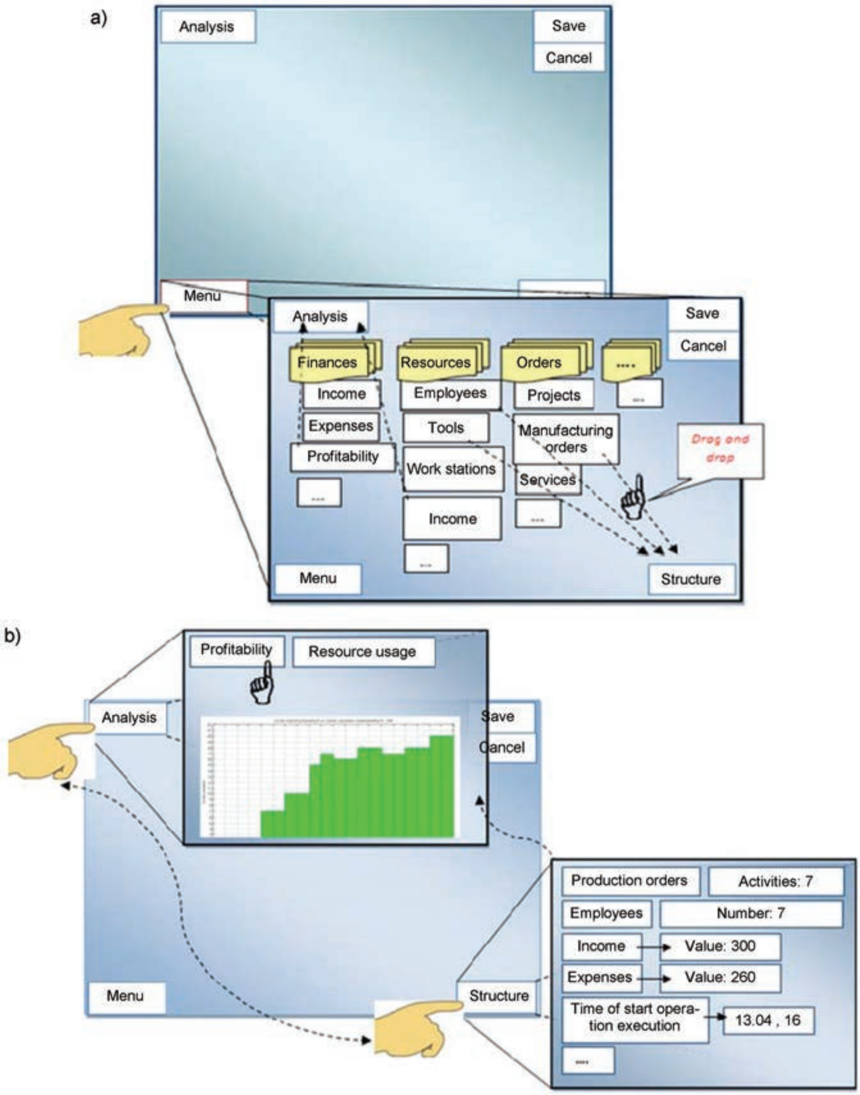


Fig. 3. Structure of Multi-board interface configuration;
exemplary structure of folders and/or directories menu (a), decision problem formulation (b)

Note, that in the course of interactive solution searching any change in parameters describing:

- an enterprise – results in different values of criteria matching-up production orders requirements,
- the criteria specifying production orders requirements – results in suggestion of an enterprise structure change.

5. CONLUSIONS

Better planning, in the manner supported by proposed approach, can improve companies' competitiveness through satisfying budgetary constraints and improving utilization of resources from a cash-flow perspective.

A computer implementation of the proposed methodology should provide a new generation DSS supporting one in cases of online resource allocation and tasks scheduling as well as production orders batching and routing. Such a tool should be especially helpful in cases when actually processed products portfolio do not spend all company's capability reserves, i.e. there is a room for additional work order considerations.

The proposed approach can be widely implemented in a number of production contexts.

REFERENCES

[1] BADELL M., ROMERO J., HUERTAS R., PUIGJANER L., *Planning, scheduling and budgeting value-added chains*, Computers and Chemical Engineering, 2004, 28, 45–61.

[2] BANASZAK Z., BOCEWICZ G., *Abductive reasoning driven approach to project-like production flow prototyping*, Foundations of Management, 2009, Vol. 1, No. 1, 43–64.

[3] BANASZAK Z., *CP-based decision support for project driven manufacturing*, [in:] *Perspectives in Modern Project Scheduling*, International Series in Operations Research and Management Science, Springer, 2006, Vol. 92, U.S., 409–437.

[4] BANASZAK Z.A., TANG X.Q., WANG S.C., ZAREMBA M.B., *Logistics models in Flexible manufacturing*, Computers in Industry, 2000, 43, 237–248.

[5] BANASZAK Z.A., ZAREMBA M.B., *Project-driven planning and scheduling support for virtual manufacturing*, Journal of Intelligent Manufacturing, 2006, Vol. 17, 641–651.

[6] CHANG S.L., WANG R.C., WANG S.Y., *Applying fuzzy linguistic quantifier to select supply chain partners at different phases of product life cycle*, International Journal of Production Economics, 2006, 100, 348–359.

[7] DANG Q.V., NIELSEN I., BØGH S., BOCEWICZ G., *Modelling and Scheduling Autonomous Mobile Robot for a Real-World Industrial Application*. IFAC Conference on Manufacturing, Modelling, Management and Control, 2013, 2098–2103, DOI 10.3182/20130619-3-RU-3018.00261.

[8] DANG Q.V., NIELSEN I., STEGER-JENSEN K., *Scheduling a single mobile robot incorporated into production environment*, [in:] P. Golinska (ed.), *EcoProduction and Logistics – emerging trends and business practices*, Part 2, Springer-Verlag, Berlin–Heidelberg, 2013, ISBN 978-3-642-23552-8, ISBN 978-3-642-23553-5 (eBook), 185–201.

[9] DANG Q.-V., NIELSEN I.E., BOCEWICZ G., *A genetic algorithm-based heuristic for part-feeding mobile robot scheduling problem* Advances in Intelligent and Soft Computing, AISC, 2012, 157, 85–92.

[10] DANG Q.-V., NIELSEN I., STEGER-JENSEN K., MADSEN O., *Scheduling a single mobile robot for part-feeding tasks of production lines*, Journal of Intelligent Manufacturing, 2013, 1–17 (in Presss).

[11] TOMCZUK-PIRÓG I., MAURY M., BANASZAK Z., *CLP-based approach to decision support for SCM systems*, Systems Science, 2007, 33 (1), 39–52.

CEP, business process,
business rules, adaptive rules,
HFD, market data

Eva ZÁMEČNÍKOVÁ, Jitka KRESLÍKOVÁ*

DESIGN OF ADAPTIVE BUSINESS RULES MODEL FOR HIGH FREQUENCY DATA PROCESSING

In this paper we would like to discuss high frequency data processing and the use of complex event platform in combination with business rules approach. For such a high volume of data it is suitable to use complex event platform (CEP), because CEP allows for big data processing in real time. We would like to focus on improvement of decision making process under the condition of dynamical adaptation of the process on the fly. We will use pattern recognition for detecting and predicting the trends in data by mining this information from historical data. After the distinguishing patterns we will build the set of business rules according to which the process runs and we will control the process flow by defining the restrictions.

We would like to use this model for building trading systems. Algorithmic trading applies complex event processing by calculating complex algorithms that indicate when to sell or buy based on real-time processing. Market data can be viewed as events. This data needs to be analyzed in real time in order to identify the trends in data and to react to these trends automatically. Traditional approach for detecting anomalies on stock market has been statistical analysis, but a CEP-based approach is able to react faster than the traditional approach.

1. INTRODUCTION

Nowadays there exist many models (or rather, whole platforms) for big data processing. By the notion of big data we generally understand high volume of source data from multiple sources like live streams, databases etc. For processing of such a high volume of data we found as the most suitable the Complex Event Platform (CEP). Lately CEP (also abbreviated as complex event processing) is in the spotlight mainly because of its capability to process data in real time. In the

* Faculty of Information Technology, Brno University of Technology,
e-mail: {izamecni|kreslika}@fit.vutbr.cz

next chapters there will be given the definition and basic knowledge about this platform and how it processes the data. More detailed insight will be devoted to decision making process and authors will suggest the use of adaptive business rules to control the flow of process events. First chapter shortly introduces complex event processing to reader of this paper. In the following chapter, a more detailed information about CEP and its solution is provided and the rules for real-time processing are explained. Third chapter discusses high frequency data domain and its applications in financial markets. The fourth chapter covers the theme of business rules and the use of these rules for controlling of data processing and the decision making part. The conclusion and the ideas for future research will be provided at the end.

2. COMPLEX EVENT PROCESSING

Complex event processing (CEP) is an emerging technology that generates actionable knowledge from distributed message-based systems, data streams and historical data in real time or near real time. There are some CEP engines but only few are capable of integrating data from multiple sources and working with high event volumes. Environmental measurements processing and making wide-spread use of it is a big data problem. The emphasis on real-time data processing is becoming an essential requirement.

According to [1] we can abstract many different levels of CEP. For example if we consider trading and financial markets, at the lowest level there will be a stock trader responsible for executing trades. A trade might consist of several executed bids, offers, payments and other financial transfers. Complex event, indicating how much profit or loss the trader generated during some time interval has predictive power which can be used in following decisions [1, 3].

Complex event processing offers its users a way to automate the detection of anomalies or other detectable and exploitable phenomena. It is way too tedious for the auditor to correlate all the trades made by all the traders to detect all the various errors they might have done. On the other hand, a CEP system, if properly configured, is able to react faster than a human. At this point, we would like to use a power of adaptive rules which can speed up the process by an automatic set up of process on the fly as a response to a specific change in context. Traditional approach for detecting anomalies on stock market has been statistical analysis after the trades were completed. For fast reactions for frauds, real-time monitoring is a necessity and can be used as a complementary method. The approach based on streaming data proved to be several orders of magnitude faster than a traditional approach based on relational database and single issue queries [5].

2.1. EVENTS

There are two parallel definitions for the concept of an „event“. First, it can mean anything that happens, or is happening, e.g. a financial trade is made. Second, it may be seen as the object acting as the manifestation of something that happens, a purchase order is sent [1]. Because of the nature of the CEP, we will use event as a representation-based definition. In complex event processing, multiple rules are applied to the events that flow through the system. These rules are applied in Event Processing Agents (EPA), which are the fundamental building blocks of CEP. EPAs monitor the patterns in event flows and react according to defined function.They take events as input and produce new events as output according to the given set of rules. Luckham [1] classfies agents as input filters, maps and constraints. Filters are event patterns that remove irrelevant events from the streams. Only relevant events are passed further to maps and constraints. Maps are used to create higher level complex events by aggregating multiple lower level events. These aggregations specify event hierarchies. Constraints can detect the presence or absence of an event or a complex event in a stream. They create notification events, when the constraint is violated (broken). Transformation is an abstract supertype of translate, aggregate, split and compose and never used by such alone. Translation means directly mapping one event to another event. This part was based on [5].

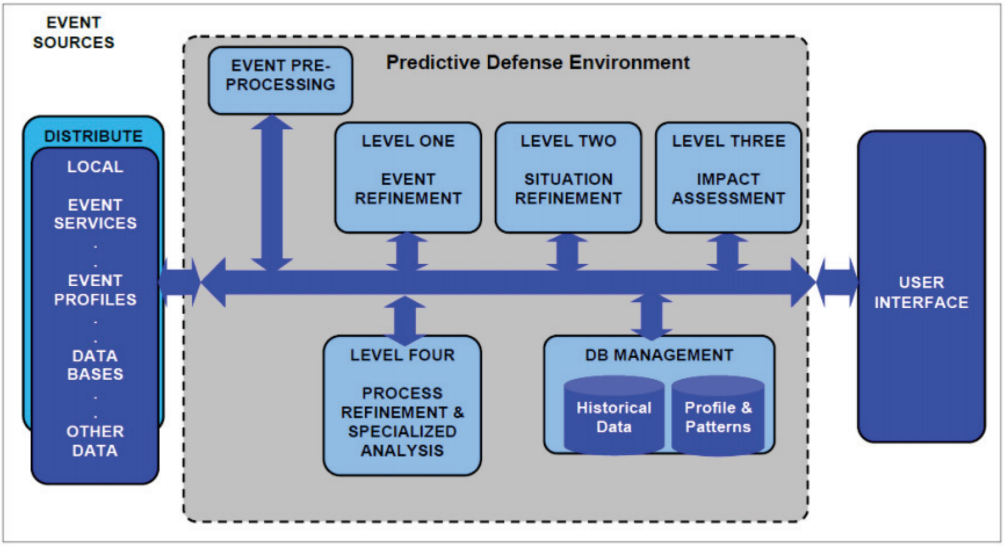


Fig. 1 Complex event processing reference architecture - adapted from [2]

On figure 1 a schema of a complex event architecture is presented. Processing of events is divided into several levels which conform to desired level of inference. At the lowest level, the event preprocessing runs – during this phase we clean the input data stream to produce comprehensible data. On the next level, the events that were detected in input data are refined and subsequently initial decisions and correlations are done. The main challenge is to find relevant data. Then, situation refinement and impact assessment follows. At the level of impact assessment, we may predict the intentions of subject or to estimate potential opportunities or threats. At the end, the process refinement is done. Information was taken from [7].

All the results of event processing and operational visualization at all levels are summed up in a human readable format via user interface.

Most of current CEP platforms solutions fall into one of these two categories:

1. Aggregation-oriented CEP, or
2. Detection-oriented CEP.

The first approach uses real time processing of event data which enters the system. As an example we might take an algorithm, performing some calculations within a moving window of a given size. On the other hand there is a detection-oriented solution which focuses on examination of data and detection of patterns or recurring behavior. Many applications use combination of both approaches.

2.2. RULES FOR REAL TIME PROCESSING

In [8] are presented rules for real time processing. As we want to design model for events processing which can react in real time we will design the model with respect to these requirements.

The rules and their brief introduction:

1. **Keep the data moving** – to process messages "in-stream", without any requirement to store them, to perform any operation or sequence of operations in order to achieve low latency. An additional latency problem arises for systems that are passive, that means the system requires applications to continuously poll for conditions of interest.
2. **Query using SQL on streams (*StreamSQL, CQL*)** – a traditional SQL system knows when the computing is finished when it gets to the end of a table, but the streaming data never ends so the stream processing engine must be instructed when to finish such an operation and output an answer. The window concept serves this purpose by defining the "scope" of a multimessage operator such as an aggregate or a join. Depending on the choice of window size and slide parameters, windows can be constructed as isolated or overlapping.
3. **Handle stream imperfections (delayed, missing and out-of-order data)** – we need to have built-in mechanisms to provide adaptability against stream "imperfections", including missing and out-of-order data, which are commonly present in real-world data streams.
4. **Generate predictable outcomes** – a stream processing system must process time-series messages in a predictable manner to ensure that the results of

processing are deterministic and repeatable. The ability to produce predictable results is also important from the perspective of fault tolerance and recovery.

5. **Integrate stored and streaming data** – requires to have the capability to efficiently store, access, and modify state information, and combine it with live streaming data. For seamless integration (without modifying the application code), the system should use a uniform language when dealing with either type of data.
6. **Guarantee data safety and availability** – to preserve the integrity of mission-critical information and avoid disruptions in real-time processing, a stream processing system must use a high-availability solution. We must ensure that the applications are running and available, and the integrity of the data is maintained at all times, despite of the failures.
7. **Partition and scale applications automatically** – to have the capability to distribute processing across multiple processors and machines to achieve incremental scalability. Stream processing systems should also support multi-threaded operation to take advantage of modern multi-processor (or multicore) computer architectures. Ideally, the distribution should be automatic and transparent.
8. **Process and respond instantaneously** – a stream processing system must have a highly-optimized, minimal-overhead execution engine to deliver real-time response for high-volume applications.

3. HIGH FREQUENCY DATA

One of the ideal source of high frequency data are financial markets. For processing of these data we need to put together statistic, mathematic, economic and also informatic methods and algorithms. Statistical methods can predict time series well but the results are not so stable when there is noise in the time series – such as inaccurate or incomplete data. Market data is highly variable and each time interval (known as tick) generates new logical unit of data. Main focus of today's research is not only the development of high-quality descriptive systems but also the ability to produce predictions of future movements of data. Information for this section were mainly taken and updated from [2].

In terms of adaptive rules generation, the real-time event processing is a key part we would like to focus on. Fast and automated data analysis won't yield any advantage if every next step in processing requires human approval. But transforming a business system to react in real-time requires not only new technologies, but a new way of thinking as well.

Most apparent are applications to high frequency financial data, which are characterized by a set of contemporaneously correlated trade marks, many of them are discrete in nature at high or ultra high frequency. In empirical studies on financial market microstructure, characteristics of the multivariate time varying conditional densities (moments, ranges, quantiles, etc.) are crucial. For instance, with our model we are able to derive multivariate conditional volatility or liquidity measures.

4. ADAPTIVE BUSINESS RULES

In CEP, the processing takes place according to user-defined rules, which specify the relations between the observed events and the phenomena to be detected. We would like to focus on event processing from the decision point of view. After the data has been processed by a CEP engine we can distinguish recurring behavior in data. We can describe these phenomena by patterns. Designed model for decision making during processing of data takes into account these patterns and so the system may react to this behavior and may apply on data the most suitable rule with which the process will continue.

For example when the trader has to decide when to buy or sell, we might apply the following rule:

- If `price_of_security` reaches `threshold_value` → open buy or sell position

Next step of processing will be chosen according to information obtained from historical data and information given by user. User may update the behavior according to preconditions, which will be in system described by business rules. Designed model of system will allow the user to specify his or her own input rules.

Next steps in decision making might also be correlated by learning of model on meaningful sample of data. But this is not the aim of this thesis.

4.1. DECISION MAKING

For decision making of some more complex situation requiring calculation we might use an engine which can communicate with our CEP solution. We might use existing tool which can generate action as an output based on a given set of data (eg. FICO Blaze Advisor) or we can implement our own solution. We still get a result which is set up on the fly without need of redeployment of the running process. Communication with external solution might be provided via web services – this design is used in Service Oriented Architecture (SOA) approach. Service oriented architecture (SOA) models were created to facilitate the design of enterprise software.

SOA addresses the following concerns:

1) many systems need to be integrated to a single interoperable entity and serve as a service for other systems

2) the existing components don't have to be implemented strictly in the same language in order for them to communicate to each other
3) businesses implement new products rapidly. Another source of integration requirements are mergers, which bring new, incompatible systems to the ecosystem. Designed model must scale to support high volumes of events [5].

4.2. CEP PATTERNS

As we have historical data we might simulate the designed solution with dynamically set up rules on these data. After the run we will compare the experimental results with real data processing. Pattern detection works always in some context. The context defines the relevant events impacting the pattern matching. It can be temporally or spatially bounded. Context can also be based on semantics of mutually referenced objects or entities. This context is called a window. According to the [5] the patterns are divided into two categories – basic patterns and dimensional patterns. They consist of

- logical operations (conjunction, disjunction, negation),
- threshold patterns – triggers when a number of events of some type has been processed,
- subset selection patterns – can select significant values in a set,
- modal patterns – check if some assertion is true.

For example, there are patterns to detect if one instance of each type of the participant set (or none of them) has been seen. These patterns without clearly defined rules can be very ambiguous. Pattern policies allow us to express evaluation, cardinality, repeated type, consumption and order policies. Evaluation policy defines whether we want to evaluate the pattern every time a new event is observed. Cardinality policy determines how many times one event can be part of a pattern matched [3,5].

Traditional approach for detecting anomalies in the stock market has been statistical analysis, after all the trades have been completed. CEP solution has the benefit that analyses according patterns can be run when needed. CEP is suitable for fast reactions for frauds, real-time monitoring is a necessity and can be as a complementary method. For example we may detect a fraud for payment by credit cards according to a spatial restriction – if we encounter two payments from different places with very long distance in short period of time, we can be almost sure about the fraud.

CEP systems are often developed bottom-up by first identifying the event information available. However, in [10] a top-down approach is described. First of all, the key performance indicators and other abstract measures are defined and then we hierarchically proceed down to find the correct low level events in a changing environment to calculate them. CEP distinguishes several scalability attributes:

- Events volume,
- Event processing agents,
- Producers and consumers,

- Window size,
- Computational complexity,
- Environment,
- Constants.

For stream analytics it is a key capability that complex event processing systems are able to scale out in order to process all incoming events in a timely fashion as required by the application domain.

4.3. USE OF BUSINESS RULES PATTERNS

By defining the business rules patterns, we will control the events flow. According to the context, we dynamically update and apply the restrictions on business rules. These restrictions might be defined as in the following example.

For simplicity let's consider we have 4 event types occurring during data processing – *A*, *B*, *C*, *D*. Then the set of restrictions could be defined:

1. *A* – event *A* must occur during data processing;
2. *NOT A* – event *A* must not occur;
3. *AB* – events *A* and *B* must occur during data processing;
4. *A* + *B* – one of the event *A* or *B* must occur;
5. *A*→*B* – event *B* will occur only after the processing of event *A*;
6. *AB*→*C* – event of type *C* will occur only if events *A* and *B* were processed before.

Also, CEP distinguishes a number of event processing patterns, for instance:

- adoption patterns,
- business patterns,
- integration patterns,
- workflow patterns, etc.

More detailed information can be found in [9].

The use of CEP patterns might be helpful if we want to generate the rules automatically during the process run.

4.4. USE OF WORKFLOW PATTERNS FOR COMPLEX EVENT PROCESSING

The decision making in nodes might be controlled similarly like we do when using the workflow patterns when controlling business process. We will focus on use of following basic patterns (also known as sequence patterns):

- AND-join pattern,
- AND-split pattern,
- XOR-join pattern: Simple Merge,
- XOR-split pattern: Exclusive Choice.

In process-aware information systems various perspectives can be distinguished. The control-flow perspective captures aspects related to control-flow dependencies between various tasks (e.g. parallelism, choice, synchronization etc). Originally, twenty patterns were proposed for this perspective, but this number has grown to over forty patterns. The data perspective deals with the passing of information, scoping of variables, etc, while the resource perspective deals with resource to task allocation, delegation and others. Finally the patterns for the exception handling perspective deal with the various causes of exceptions and the various actions that need to be taken as a result of exceptions occurring. This information has been adapted from [11].

5. CONCLUSION AND FUTURE WORK

In this article we introduced the idea of processing of big data by using the complex event platform with adaptive business rules for decision making. We mentioned the advantages of its usage and the overview of methods and theoretical basis.

The next step of our research is to formally design and consequently implement model for decision making of processing events by using CEP with adaptive model of business rules. After that we will perform test measurments and compare experimental results with real result. Simulation will be conducted on real historical data on relevant sample of data implementation of system according to adaptive business rules. We will experiment with some scalable attributes such as the size of window, the volume of event and others and compare partial results.

ACKNOWLEDGEMENT

This work was partially supported by the BUT FIT grant FIT-S-14-2299, “Research and application of advanced methods in ICT”.

REFERENCES

[1] LUCKHAM D., *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems*, Addison Wesley Professional, May 2002, ISBN: 0201727897.

[2] STEINBERG A., BOWMAN C., *Handbook of Multisensor Data Fusion*, JDL Laboratories, CRC Press, 2001.

[3] GENCAY R., DACOROGNA M., MULLER U.A., PICTET O., OLSEN, R., *An Introduction to High-Frequency Finance*, Academic Press, London 2001, ISBN 978-0-12-279671-5.

[4] ETZION O., NIBLETT P., *Event Processing in Action*, 1st ed., Manning Publications Co., Greenwich, CT, USA, 2010.

[5] AALTO A., *Scalability of Complex Event Processing as a part of a distributed Enterprise Service Bus*, master thesis, Aalto University, Esbo 2012.

[6] What is Complex Event Processing?, 2007, [online, visited June 2014] <http://www.thecepblog.com/what-is-complex-event-processing/>

[7] ZÁMEČNÍKOVÁ E., KRESLÍKOVÁ J., *Time Series Prediction Based on Event Driven Business Process Management*, International Journal of Computational Engineering Research (IJCER), Ghaziabad, Uttar Pradesh: International Journal of Computational Engineering research, 2014, Vol. 4, Iss. 4, 1–6, ISSN 2250-3005.

[8] STONEBRAKER M., ÇETINTEMEL U., ZDONIK S., Newsletter, ACM SIGMOD Record, Vol. 34, Iss. 4, December 2005, 42–47, ACM New York, NY, USA, [online, visited June 2014] <http://cs.brown.edu/~ugur/8rulesSigRec.pdf>

[9] PASCHKE A., *Design Patterns for Complex Event Processing*. 2nd International Conference on Distributed Event-Based Systems (DEBS ’08), Rome, Italy, 2008 [online, visited June 2014] <http://arxiv.org/ftp/arxiv/papers/0806/0806.1100.pdf>

[10] KELLNER I., FIEGE L., *Viewpoints in complex event processing: industrial experience report*. Proceedings of the Third ACM International Conference on Distributed Event-Based Systems, DEBS 2009, ACM, New York, NY, USA, 2009, 9:1–9:8.

[11] AALST W., *Workflow Patterns* [online, visited June 2014] <http://www. workflowpatterns.com/>

resource allocation,
time-cost trade off, fuzzy systems,
optimization

Donat ORSKI*, Piotr NURKOWSKI

ON LINEARIZATION OF THE BUDGET ALLOCATION PROBLEM IN PROJECT MANAGEMENT UNDER TWO-LEVEL UNCERTAINTY

The project management problem is considered in which for the given project *deadline* and *budget,* under two-level uncertainty concerning duration of operations, the goal is to determine budget allocation maximizing certainty of the successful project completion. It is assumed that upper boundary time-resource function only is given for each operation, with a fuzzy parameter inside. Such two-level uncertainty is recommended for use in real-life situations in which just rough estimations of operation durations are available. For typical upper boundary functions the optimization problem is nonlinear and non-convex, which disables application of classical optimization algorithms. This chapter presents a method of transforming the problem to the form allowing application of linear programming. Two ways of linearization are introduced and evaluated. The experimental results support the usefulness of the presented approach.

1. INTRODUCTION

The chapter concerns a class of project management problems in which order of project activities (operations) is depicted by AoA (*activity on arc*) graph (see Fig. 1), and operation durations depend on the amount of resources allocated to them. We assume that just one type of a resource is used. It is continuous, nonrenewable and may be interpreted as project funding, i.e. the project budget. In the literature (e.g., [3], [25]) problems concerning distribution of a limited amount of a resource so as to achieve the required project completion time are known as *time-resource trade off* or *time-cost trade off* problems. The time-resource or time-cost function assumed for

* Wrocław University of Technology, Institute of Informatics, Wybrzeże Wyspianskiego 27, 50-370 Wroclaw, Poland.

every operation describes the relationship between the amount of a resource allocated to the operation and its duration. These problems were introduced shortly after CPM and PERT [18], addressed as PERT/cost problems and solved for time-cost functions linearized in time intervals between given *crash* and *normal* times (Fig. 2) [12], [17] or for convex time-cost functions [1]. With the appearance of fuzzy sets theory and its applications not only CPM and PERT were given the second life [8], [9], but also PERT/cost (e.g. [19]). The rationale behind treating either operation durations, or some parameters in time-resource trade off problems, as fuzzy numbers is to eliminate unrealistic assumptions of CPM, PERT and PERT/cost. Full knowledge on deterministic relationships or knowledge of prior probability distributions and stochastic independence of operations are usually indicated as such. The use of fuzzy variables which need to be characterized by experts, e.g. project managers, team leaders, seems especially motivated in case of innovative, non template-based projects carried out in new environments.

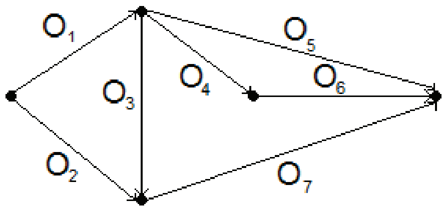


Fig. 1. Example structure of project operations O_1 to O_7

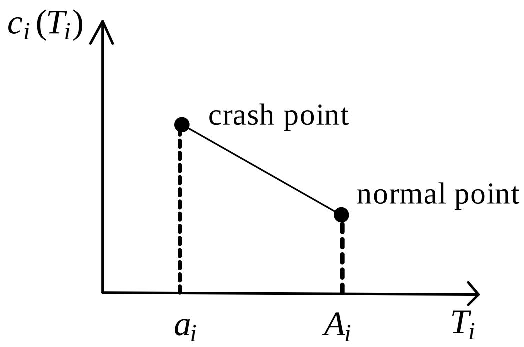


Fig. 2. Linear time-cost function defined for operation O_i between the *crash* (a_i) and *normal* (A_i) times

However, most research done within the framework of fuzzy variables application to time-cost trade off problem ([10], [11], [14]–[16], [19], [20], [24]) looks as template-based. The authors make use of simplified linear time-cost functions defined for all operations in the intervals between their so called *normal* and *crash* points only (Fig. 2). In subsequently published papers the authors assume fuzziness of more and more parameters and variables, e.g. times, costs, deadlines, without clear practical motivation. This approach appears incremental, on one hand, and does not affect linearity of the model, on the other hand.

Some other research, originating in knowledge-based systems ([21], [22], [23]), is motivated by the idea that a mathematical model of reality should be adequate to precision and completeness of available knowledge. Lack of precise knowledge on deterministic time-resource relationships for particular operations motivates taking advantage of typical functions defining upper boundaries only. Each upper boundary function contains a parameter describing the *size* of a task to be completed during the operation. As the size is relative to competence and efficiency of a team chosen for carrying it out, this team's leader (expert) is asked to characterize it approximately. The authors employ the formalism of uncertain variables ([4]–[7]), i.e., fuzzy variables recommended for dealing with parametric uncertainties in mathematical models.

This chapter is related to the second research framework. The concept of imprecise knowledge of both time-resource relationships and their parameters is called here the two-level uncertainty. Despite precisely defined project *deadline* and *budget*, project success remains undetermined, and only the *certainty* of success may be calculated. The optimization goal is to find a budget allocation maximizing the certainty of project success, which is understood as the project being completed (i) on or before the deadline and (ii) within the given budget. Unfortunately, no well known optimization methods may be applied to solve the problem, because its formulation contains non-linear and non-convex elements. The purpose of this chapter is to show how linearization based on Taylor expansion may be used to transform the budget allocation problem under consideration to a form appropriate for linear programming.

The chapter is organized as follows. Section 2 introduces two-level uncertainty and presents the formulation of the budget allocation problem. Section 3 shows transformation of the problem into an epigraph form and describes linearization of constraints. Section 4 explains how the optimized certainty index may be transformed to a form eligible for linearization and describes its linearization. Section 5 contains results of numerical experiments performed to evaluate quality of linear approximation.

2. BUDGET ALLOCATION PROBLEM FORMULATION

2.1. TWO-LEVEL UNCERTAINTY

In the presented knowledge-based approach, each operation is described by the inequality

$$T_i \le \varphi_i(u_i, x_i), \quad i = 1, 2, ..., k \tag{1}$$

where T_i is the execution time of the *i*-th operation, u_i is the amount of a resource allocated to the *i*-th operation, an unknown parameter $x_i \in R^1$ is a value of an uncertain variable $\bar{x}_i$ described by a triangular certainty distribution $h_i(x_i)$ (Fig. 3) given by an expert, and $(\bar{x}_1, ..., \bar{x}_k) = \bar{x}$ are independent. Typically simple upper boundary time-resource functions, $\varphi_i(u_i, x_i) = \frac{x_i}{u_i}$, are assumed. As the total amount of a distributed resource, i.e. the budget, is limited to *U*, every allocation $(u_1, ..., u_k) = u$ should satisfy the constraints:

$$u_i > 0 \quad \text{for each } i \qquad \text{and} \qquad u_1 + u_2 + ... + u_k = U \,. \tag{2}$$

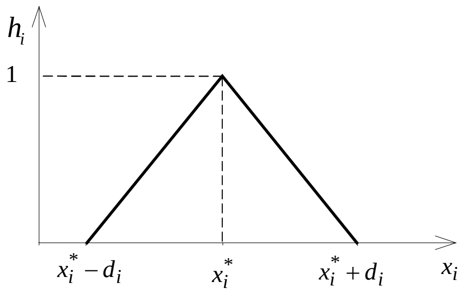


Fig. 3. Triangular certainty distribution for the uncertain variable $\bar{x}_i$

2.2. BUDGET ALLOCATION UNDER UNCERTAINTY

For the given budget *U* and the acceptable project duration $a > 0$, one may determine the certainty index [7], [21]

$$v[T(u;\bar{x}) \,\tilde{\leq}\, a] \overset{\Delta}{=} v(u;a,U) \tag{3}$$

for the property "project duration *T* is approximately less or equal to *a* for the allocation *u* and the uncertain parameter $\bar{x}$", where $T(u; x)$ denotes the upper boundary function for project duration. According to [21] $v(u; a, U)$ may be determined on the basis of operations boundary functions $\varphi_i(u_i, x_i)$ and calculations defined for certainty indexes [7]:

$$v(u;\alpha,U) = v\{[\varphi_1(u_1,\bar{x}_1) \,\tilde{\leq}\, a_1)] \wedge [\varphi_2(u_2,\bar{x}_2) \,\tilde{\leq}\, a_2)] \wedge ... \wedge [\varphi_k(u_k,\bar{x}_k) \,\tilde{\leq}\, a_k)]\}$$

$$= \min_i v_i(u_i; a_i) \tag{4}$$

where $a_1, a_2, ..., a_k$ denote acceptable durations of project operations. These durations are not given, so must be included in the problem formulation as unknowns which are subjects to linear constraints related to paths in a graphical model of a project structure. For example, in case of project depicted in Fig. 1 the following constraints must be included: $a_1 + a_5 \le a$, $a_1 + a_3 + a_7 \le a$, $a_1 + a_4 + a_6 \le a$, $a_2 + a_7 \le a$. With triangular certainty distributions h_i one obtains [7], [21]

$$v_i(u_i; a_i) = \begin{cases} 1 & \text{for } x_i^* \le a_i u_i \\ \frac{1}{d_i} a_i u_i - \frac{x_i^*}{d_i} + 1 & \text{for } x_i^* - d_i \le a_i u_i \le x_i^*, \\ 0 & \text{otherwise} \end{cases} \qquad i \in \overline{1,k}. \tag{5}$$

Then, the budget allocation problem under two-level uncertainty may be formulated as the following optimization problem.

Given: project structure, U, a, x_i^*, d_i, $i \in \overline{1,k}$.

Find: u and $a_1, a_2, ..., a_k$ maximizing

$$\min_i v_i(u_i; a_i), \tag{6}$$

subject to constraints (2) and constraints concerning $a_1, a_2, ..., a_k$.

All constraints are linear, however, the optimized function (6) is not. It not only contains *min* operator, but also nonlinear and non-convex terms "$a_i u_i$" in (5).

In subsequent sections the two ways of problem transformation will be presented, developed for removal of the *min* operator and for linearization of nonlinear terms.

3. LINEARIZATION OF CONSTRAINTS

The *min* operator may be removed by using the so called epigraph form of the optimization problem [2]. This approach was used also in case of the rate allocation under uncertainty [13].

3.1. EPIGRAPH FORM OF THE BUDGET ALLOCATION PROBLEM

The idea of this approach consists in introducing a new variable representing the result of *min* operation and in replacing *min* operator that compares all *k* certainty indexes by *k* inequality constraints, each comparing only one certainty index with the new variable. This leads to the following problem formulation.

Given: project structure, U, a, x_i^*, d_i, $i \in \overline{1, k}$.
Find: u, $a_1, a_2, ..., a_k$ and $\bar{v}$ maximizing

$$\bar{v} ,$$

subject to the constraints introduced previously in Sec. 2.1, and to the new ones:

$$0 \le \bar{v} \le 1 ,$$
$$v_1(u_1; a_1) \ge \bar{v} , \quad v_2(u_2; a_2) \ge \bar{v} , \quad ..., \quad v_k(u_k; a_k) \ge \bar{v} . \tag{7}$$

3.2. LINEARIZATION

The formulation in Sec. 3.1 is still nonlinear because of k nonlinear terms "$a_i u_i$". They appear in separate constraints in functions $v_i(u_i; a_i)$. Here we suggest to approximate these functions using linear components of Taylor expansion. Let (a_{0i}, u_{0i}) and $f_i(u_i, a_i)$ denote the expansion point and the linear approximation of $v_i(u_i; a_i)$, respectively.

Application of Taylor expansion on $v_i(u_i; a_i)$ yields

$$v_i(u_1; a_1) = \frac{1}{d_i} a_{0i} u_{0i} - \frac{x_i^* - d_i}{d_i} + \frac{u_{0i}}{d_i}(a_i - a_{0i}) + \frac{a_{0i}}{d_i}(u_i - u_{0i}) + \frac{1}{d_i}(a_i - a_{0i})(u_i - u_{0i})$$

$$= \frac{1}{d_i} a_{0i} u_{0i} - \frac{x_i^* - d_i}{d_i} + \frac{u_{0i}}{d_i}(a_i - a_{0i}) + \frac{a_{0i}}{d_i}(u_i - u_{0i})$$

$$+ \frac{a_1 u_i - a_i u_{0i} - a_{0i} u_i + a_{0i} u_{0i}}{d_i} , \qquad i \in \overline{1, k} . \tag{8}$$

Elimination of nonlinear components produces the following linear approximation

$$f_i(u_i, a_i) = \frac{1}{d_i} a_{0i} u_{0i} - \frac{x_i^* - d_i}{d_i} + \frac{u_{0i}}{d_i}(a_i - a_{0i}) + \frac{a_{0i}}{d_i}(u_i - u_{0i}) , \qquad i \in \overline{1, k} . \tag{9}$$

4. LINEARIZATION OF THE CERTAINTY INDEX

In this section, to deal with *min* operator in (6), we will employ the idea other than used in Sec. 3. It is based on the fact that in a non-trivial situation, i.e., when $0 < v_i(u_i; a_i) < 1$, certainty indexes $v_i(u_i; a_i)$ are strictly increasing functions of u_i. It may be noticed that the optimal budget allocation should satisfy the constraints (2) and the set of equations

$$v_1(u_1;a_1) = v_2(u_2;a_2) = ... = v_k(u_k;a_k)\,. \tag{10}$$

4.1. PARTLY ANALYTICAL OPTIMIZATION OF THE CERTAINTY INDEX

The set of k equations (10) and (2), linear with respect to $u_1, u_2, ..., u_k$, may be solved analytically. Though, the obtained u and v still functionally depend on $a_1, a_2, ..., a_k$:

$$u_i(a_1, a_2, \ldots, a_k) = \frac{d_i}{\sum_{j=1}^{k} d_j}\left(U - \sum_{j=1}^{k}\frac{x_j^*}{a_j}\right) + \frac{x_i^*}{a_i}\,, \qquad i \in \overline{1,k}\,, \tag{11}$$

$$v(a_1, ..., a_k) = 1\frac{\sum_{i=1}^{k}\frac{x_i^*}{a_i} - U}{\sum_{i=1}^{k}\frac{d_i}{a_i}}. \tag{12}$$

The analytical calculations reduced the budget allocation problem formulated in Sec. 2.2 to the optimization problem with unknown durations $a_1, a_2, ..., a_k$, nonlinear function (12) to be maximized, and linear constraints imposed on $a_1, a_2, ..., a_k$, the same as in Sec. 2.2.

4.2. LINEARIZATION

Unlike Sec. 3, now only one function needs to be linearly approximated, nevertheless, it contains k variables and has nonzero higher order derivatives. We shall approximate (12) by taking only linear components of Taylor expansion. Let $(a_{01}, a_{02}, ..., a_{0k})$ and $f(a_1, a_2, ..., a_k)$ denote the expansion point and the linear approximation of $v(a_1, ..., a_k)$, respectively.

Application of Taylor expansion on (12) yields

$$f(a_1, ..., a_k) = \frac{U - \sum_{i=1}^{k}\frac{x_i^* - d_i}{a_{0i}}}{\sum_{i=1}^{k}\frac{d_i}{a_{0i}}} + \sum_{i=1}^{k}\frac{\frac{x_i^*}{a_{0i}^2}\sum_{j=1}^{k}\frac{d_j}{a_{0j}} - \frac{-d_i}{a_{0i}^2}\left(U - \sum_{j=1}^{k}\frac{x_j^*}{a_{0j}}\right)}{\left(\sum_{j=1}^{k}\frac{d_j}{a_{0j}}\right)^2}(a_i - a_{0i}). \tag{13}$$

5. QUALITY OF APPROXIMATION

Computational experiments have been performed to evaluate the quality of approximation. Obviously, the approximation should be more accurate in closer neighbourhood of the expansion point. However, one should not expect that for budget allocations and/or operation durations distant from the expansion point satisfactory accuracy may be achieved. We assume that the final solution algorithm shall work in an iterative way, searching for the optimal allocation in a close neighbourhood of the expansion point only, and then taking this local solution as the expansion point for the next iteration. Thus the main purpose of the experiments was to roughly estimate the neighbourhood radius r around the expansion point, which could be recommended for application in this iterative algorithm.

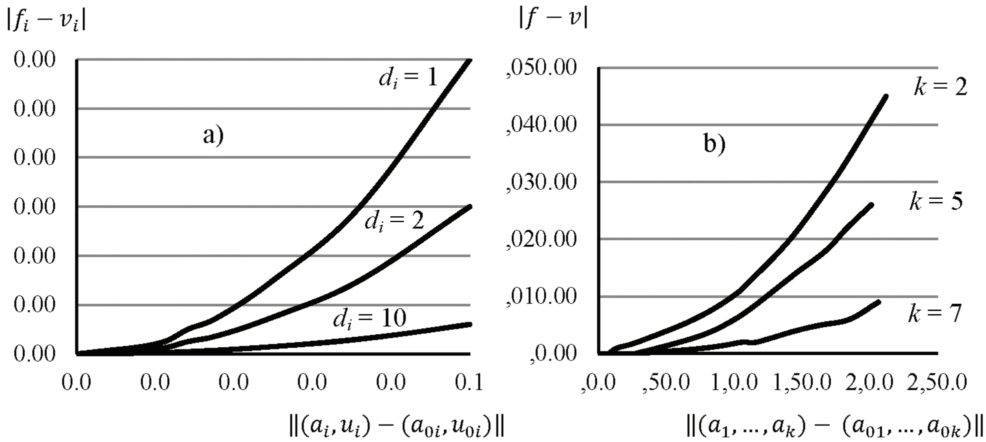


Fig. 4. Approximation error vs. distance from the expansion point determined for:
a) different degrees of expert's uncertainty about x_i^* and linearization as in Sec. 3,
b) different numbers of operations and linearization as in Sec. 4

The experimental results presented in Fig. 4a were obtained for $x_i^* = 15.5$ and $(a_{0i}, u_{0i}) = (3, 5)$, while those presented in Fig. 4b for $x_i^* = 10$, $d_i = 5$, $a_{0i} = 5$, $i \in \overline{1, k}$, and $U = k$. They show that inaccuracy less than 0.05 may be achieved on approximation of certainty indexes v_i and v within the neighbourhoods of $r_a \approx 0.3$ and $r_b \approx 2$, respectively. It may be also observed that weaker expert's confidence (greater d_i) results in better quality of approximation, in the first case, and so does greater number of operations, in the second case.

6. FINAL REMARKS

In this chapter we presented two linearization methods developed for the problem of budget allocation under two-level uncertainty. The second approach could be recommended for complex projects, though, both methods need further investigation in linear programming procedures used iteratively in a global optimization algorithm.

Because time, resource and certainty are exchangeable one for another, the more general trade off problem may be considered under uncertainty. It could be called time-resource-certainty trade off problem, and, in fact, its different formulations have been already presented e.g. in [21]. The problem linearization suggested in this chapter and the solution method may be adopted also to solving deadline or budget problems for the assumed satisfactory level of certainty (i.e., the certainty threshold).

The allocation problem formulation considered in this chapter was based on a *basic definition* of the uncertain variable. The *complex definition* of the uncertain variable (i.e., *C*-uncertain variable) [6], [7], [23] is described as the one which allows better exploitation of expert's knowledge. It would be reasonable to apply the two methods described here to linearize the problem formulation developed for the complex definition of the uncertain variable.

The presented approach may be considered for possible application to the rate allocation problem defined for computer networks with uncertain link capacities and utility function parameters [13]. Necessary enhancements should include introduction of multiple types of continuous, nonrenewable resources corresponding to capacities of network links.

REFERENCES

[1] BERMAN E.B., *Resource allocation in a PERT network under continuous activity time-cost functions*, Management Sciences, 1964, Vol. 10, No. 4, 734–745.
[2] BOYD S., VANDENBERGHE L., *Convex Optimization*, Cambridge University Press, Cambridge 2004.
[3] BRUCKER P., DREXL A., MOHRING R., NEUMANN K., PESCH E., *Resource-constrained project scheduling: Notation, classification, models, and methods*, European Journal of Operational Research, 1999, No. 112, 3–41.
[4] BUBNICKI Z., *Uncertain variables and their applications for a class of uncertain systems*, International Journal of Systems Science, 2001, Vol. 32, No. 5, 651–659.
[5] BUBNICKI Z., *Uncertain variables and their application to decision making*, IEEE Trans. on SMC, Part A: Systems and Humans, 2001, Vol. 31, No. 6, 587–596.
[6] BUBNICKI Z., *Uncertain Logics, Variables and Systems*, Springer-Verlag, Berlin, London, New York 2002.
[7] BUBNICKI Z., *Analysis and Decision Making in Uncertain Systems*, Springer-Verlag, Berlin, London, New York 2004.
[8] CHANAS S., KAMBUROWSKI J., *The use of fuzzy variables in PERT*, Fuzzy Sets and Systems, 1981, Vol. 5, 1–19.

[9] CHANAS S., DUBOIS D., ZIELINSKI P., *On the sure criticality of tasks in activity networks with imprecise durations*, IEEE Transactions on Systems, Man, And Cybernetics – Part B: Cybernetics, 2002, Vol. 32, No. 4, 393–407.
[10] CHEN S.-P., TSAI M.-J., *Time–cost trade-off analysis of project networks in fuzzy environments*, European Journal of Operational Research, 2011, No. 212, 386–397.
[11] EVANGELINE JEBASEELI M., PAUL DHAYABARAN D., *A comparative study on fully fuzzy time cost trade off*, Annals of Pure and Applied Mathematics, 2012, Vol. 1, No. 2, 97–107.
[12] FULKERSON D.R., *A network flow computation for project cost curves*, Management Science, 1961, Vol. 7, No. 2, 167–178.
[13] GASIOR D., ORSKI D., *On the rate allocation problem under co-existence of uncertain utility function parameters and uncertain link capacities*, Journal of the Operational Research Society, 2014 (in press), DOI: 10.1057/jors.2013.113.
[14] GHANZAFARI M., SHAHANAGHI K., YOUSEFLI A., *An application of possibility goal programming to the time-cost trade off problem*, Journal of Uncertain Systems, Vol. 2, No. 1, 2008, 22–30.
[15] KAUR P., KUMAR A., *Modification in Chen and Tsai's method for solving time–cost trade-off problems of project networks in fuzzy environments*, Neural Computing and Applications, 2012, published online.
[16] KE H., MA W., GAO X., XU W., *New fuzzy models for time-cost trade-off problem*, Fuzzy Optimization and Decision Making, 2010, Vol. 9, 219–231.
[17] KELLEY J.E., Jr., *The critical path method: resources planning and scheduling*, [in:] Industrial Scheduling, Prentice-Hall, Englewood Cliffs, NJ, 1963.
[18] MALCOLM D.G., ROSENBLOOM J.H., CLARK C.E., FAZER W, *Application of a technique for R & D program evaluation (PERT)*, Journal of Operational Research, 1959, Vol. 7, No. 5, 646–669.
[19] MON D.L., CHENG C.H., LU H.C., *Application of fuzzy distributions on project management*, Fuzzy Sets and Systems, 1995, No. 73, 227–234.
[20] NIKOOMARAM H., HOSSEINZADEH LOTFI F., JASSBI J., SHAHRIARI M.R., *A new mathematical model for time cost trade-off problem with budget limitation based on time value of money*, Applied Mathematical Sciences, 2010, Vol. 4, No. 63, 3107–3119.
[21] ORSKI D., *Application of uncertain variables to knowledge-based resource distribution*, [in:] D. Krol, N.T. Nguyen (Eds.), *Intelligence integration in distributed knowledge management*, Hershey, Pa, Information Science Reference, 2008, 63–84.
[22] ORSKI D., HOJDA M., *Decision making under uncertainty for a class of production systems with serial and parallel operations*, Kybernetes, 2008, Vol. 37, No. 5, 677–692.
[23] ORSKI D., *Application of C-uncertain variables and systemic approach to resource distribution in a class of operation systems*, Kybernetes, 2009, Vol. 38, No. 7/8, 1265–1281.
[24] SATHISH S., GANESAN K., *Fully fuzzy time-cost trade-off in a project network - a new approach*, Mathematical Theory and Modeling, 2012, Vol. 2, No. 6, 53–65, published online.
[25] WEGLARZ J., JOZEFOWSKA J., MIKA M., WALIGORA G., *Project scheduling with finite or infinite number of activity processing modes – A survey*, European Journal of Operational Research, 2011, No. 208, 177–205.

legal form of economic activity,
company's market value,
Polish capital companies

Zofia WILIMOWSKA*,
Stanisław LEWIŃSKI VEL IWAŃSKI**
Monika KOTOWSKA-LEWIŃSKA***

EFFECTIVENESS OF MANAGEMENT OF ENTERPRISES NOT RATED ON STOCK EXCHANGE – CHOICE OF LEGAL FORM DUE TO THE MAXIMIZATION OF COMPANY'S VALUE

Effectiveness of company's management could be verified in various ways. In this context mainly profitability, market position and other factors are analyzed. In this article effectiveness of management will be analyzed from the company's market value maximization point of view. Company's market value is chosen as an suitable indicator, because of its strong explanation power. Firm's market value can show properly, whether both shareholders and lenders earn money on their capital investment. But this measure also represents all wealth of other company's stakeholders (workers, managers, contractors, governmental institutions and other) [18]. Main goal of this work is to verify, whether choice of enterprise's legal form has got an influence on effectiveness of its assets management. Answer on this question will be obtained through legal-economic analysis, performed on a sample of 16 non-stock firms, both Joint Stock Companies (J-S.C.) and Limited Liability Companies (LLC), which run its businesses constantly from year 2001.

* University of Applied Sciences, Armii Krajowej 7, 48-300 Nysa; Wrocław University of Technology, Institute of Industrial Engineering and Management, Smoluchowskiego 25, 50-372 Wrocław, Poland; e-mail: zofia.wilimowska@pwr.edu.pl

** Wrocław University of Technology, Institute of Industrial Engineering and Management, Smoluchowskiego 25, 50-372 Wrocław, Poland; e-mail: stanislaw.lewinski@pwr.edu.pl

*** The Adam Mickiewicz University in Poznań, The Faculty of Law and Administration, Niepodległości 53, 61-714 Poznań, Poland, e-mail: monika.kotowska@amu.edu.pl

1. INTRODUCTION

It is not possible to start and continue an economic activity, until we choose an appropriate legal form of our emerging enterprise. This choice is related to the preliminary analysis of the scope, nature and scale of our business. On the very beginning generally we must compare advantages and disadvantages of available legal forms of business conducting. In this context it would be well to have an appropriative analytic measures for such comparison. Apart from issue of descriptive characteristic of available legal forms of business conducting, we should also have numerical indicators, which could give us an answer to which extent which legal form of company is better for certain enterprise.

As a first step to prepare such numerical measurement could be study presented in this article. In this work we propose to measure changes in a market value growth, as a form of evaluation, which might be helpful for assessing which type of legal form is better for certain (rather large- and medium-scale) business. In our research we had taken into consideration only Joint-Stock Companies (J-S.C.) and Limited Liability Companies (LLC). Our study was performed on a sample of 16 medium and large size enterprises, constantly running its business from year 2001 till 2010.Economic Value Growth Added ($EVGA_t$) [13]measure had been taken to the study as an approximation of analyzed firms market value growth.

2. LEGAL FORM OF CONDUCTED BUSINESS AS A KEY DECISION OF EMERGING ENTERPRISES

Due to the organizational and legal form of business conducting we can identify enterprises running in a form of: civil law partnerships, personal companies, capital ones, cooperatives, associations or foundations.

In this article broader analysis will be performed for capital companies: Limited Liability Companies and Joint-Stock ones. The principles of functioning of capital companies were defined in rules of The Act of 15 September 2000 of the Code of Commercial Companies [1].

Historically the oldest (in international law) form of conducting business is a joint-stock company. Its beginnings reaches the fifteenth and sixteenth century. Its prototypes were English trade associations formed to service the newly discovered overseas countries, especially India and America. The very first regulations of joint stock company were published in the Prussian Landrecht of the seventeenth century and in the Napoleonic Commercial Code (1807) [8, p. 175].

In Poland, the stock law was regulated first by Regulation of President of the Republic of Poland of 22 March 1928 [11] and amended by Regulations of 3 December

1930 on the application of the law on joint stock companies (...) [10] and Regulation of 27 October 1933 introducing the Commercial Code [9] – basic legal document specifying functioning of joint stock companies and limited liability ones.

Limited liability company is historically (internationally) quite younger form of conducting business. Motherland of limited liability company is nineteenth century Germany (from the end of that century), where this form of economic activity found its normative form in the Act of 20 April 1892 on limited liability companies [2]. Early as in 1892, the majority of then existing export-import companies has been transformed into commercial companies, including a limited liability companies. Legal regulation of LLC after Germany was then introduced in Austrian law (1906), English (1907), French (1925) and in other law systems. In Poland, the activity of LLC was regulated for the first time in the Decree of 8 February 1919 on the limited liability companies [5]. Normative configuration of LLC was conditioned, like other companies, the needs of economic life [8, p. 143].

3. LAW CONDITIONS OF CONDUCTING BUSINESS IN A FORM OF JOINT-STOCK COMPANIES AND LIMITED LIABILITY COMPANIES

In this part it should be emphasized that, despite both joint stock companies and limited liability ones belong to one group of organizations, they have many features that distinguish them from each other, which in a different way form conditions of economic activity.

Summary comparison of the legal characteristic of limited liability companies and joint stock ones is presented in a Table 1.

Table 1. Comparison of features of LLC and J-S.C. companies.
Source: Own elaboration on basis of [8]

Limited Liability Companies	Joint-Stock Companies (J-S.C)
Preferable form of activity	
•Joint ventures with a small number of shareholders. •For partners who wish to keep the direct supervision. •For smaller size business ventures.	•Joint ventures with a significant number of shareholders. •For shareholders who wish to delegate the supervision to the supervisory board. •For grater size business ventures. •For companies planning initial public offerings.

The steps necessary to set up a company	
•Conclusion of the company's agreement in the form of a notarial deed or by electronic means. •Acquisition of contributions by the partners to cover the entire share capital. •Establishment of the Board. Establishment of the Supervisory Board or the audit committee is voluntary – but mandatory only if the following circumstances occur simultaneously: the share capital exceeds 500 thousand PLN, and the number of partners exceeds 25 people. •Registration in the Register of Entrepreneurs of the National Court Register.	•Conclusion of the company's agreement only in the form of a notarial deed. • Acquisition of contributions by the shareholders: shares subscribed for cash; required to pay at least in a quarter of their nominal value. •Establishment of the Board. Mandatory establishment of the supervisory board. •Registration in the Register of Entrepreneurs of the National Court Register.
The initial share capital	
•The minimum share capital amounts to 5,000.00PLN. •The nominal value of one share is a minimum of 50 PLN. •Share capital of the company is divided into shares of equal or unequal nominal value. •The share capital must be paid before registration in the National Court.	•The minimum share capital amounts to 100,000.00. •The nominal value of one share is a minimum of 1 penny. •Share capital of the company is divided into shares of equal nominal value. •The obligation to pay up at least a quarter of the initial capital of the company before registration. •Mandatory verification by a certified auditor of non-cash contributions to the company.
The increase of initial share capital	
•The initial capital may be increased by the partners pursuant to existing provisions of partnership agreement and by amending the articles of agreement. •The share capital may be increased by the new contributions by the shareholders or with financial resources of company. •New contributions don't need a form of notarial act.	•The share capital may be increased: o pursuant to a shareholders' resolution based on new contributions of shareholders or based on the financial resources of the company, o pursuant to a resolution of the Board, previously authorized by the shareholders. •Resolution to increase the share capital must be made in the form of a notarial act.
Shares	
•Immaterial nature – shareholders are written in the book of shares conducted by the Board. •The contract of sale and pledge of shares need to have written form with signatures notarial authenticated under pain of nullity. •The sale of shares is effective for the company since it was notified by one of the interested parties.	•Shares may be either tangible or intangible (pecuniary or non-pecuniary). •Shares may be: registered or bearer shares. •Bearer shares may not be issued prior to full payment; on proof of partial payment there are issued interim certificates. •The book of shares includes entries for the registered shares and interim certificates. •The shares are indivisible.

	•Disposal of registered or provisional certificate requires written statements and transfer of ownership of shares or certificates of relief. •Sales of bearer shares requires the transfer of possession of bearer shares. •Sales of bearer shares is rather simpler than shares in limited liability companies, because of constant rating of this type of shares in case of joint-stock companies listed on Stock Exchange (ex. Warsaw Stock Exchange [16] or New Connect [14]).
The liability of company	
•The company is responsible for its obligations with all its assets, without any restrictions.	
The liability of partners, the shareholders and board members for the obligations of company	
•Shareholders are not liable for obligations of the company. Board members shall be jointly and severally liable to the creditors of the company in respect of ineffective enforcement of the company's assets. Members of management board may avoid liability if they could prove, that in right time they reported for bankruptcy or initiated bankruptcy proceedings. •A claim to a board member for damages caused to a company shall expire after three years from the date on which the company became aware of the damage.	•Shareholders are not liable for obligations of the company. Board members are not liable for obligations of the company. •A claim to a board member for damages caused to a company shall expire after three years from the date on which the company became aware of the damage.
Regulations concerning the accounting of economic events and related to publication of financial reports and statements	
•Full accounting, in accordance with The Act on Accounting [3]. •Publication financial statements in National Court Register: mandatory – Balance Sheet, Loss and Profit Statement and voluntary-Cash Flow.	•Full accounting, in accordance with The Act on Accounting [3]. •Publication financial statements in National Court Register: mandatory: Balance Sheet, Loss and Profit Statement and Cash Flow. •Publication detailed reports according to Stock Exchange regulations.

According to the analysis presented in Table 1 it should be noted that establishment and operation of limited liability company is relatively much more simpler than the joint-stock company. For limited liability company there are envisaged fewer formal requirements related primarily to the activities carried out during the registration process. Also the limited liability company is obliged to conduct a smaller range of reporting and to have less formalized management structure (fewer company bodies). At the same time only 5.000,00 PLN is required as the initial capital of this company.

However, in the form of LLC we cannot run a certain types of business, such as a bank or investment fund. Also, in the case of this form of activity attracting new shareholders may be a bit more difficult, especially when in the case of a joint stock company we can reach a larger group of capital donors through Stock Exchange, by a public issue.

At the same time, it is important that, in the case of shareholders and the board of management of joint stock company there is wholly excluded their liability for the debts of the business entity (liability is limited to the amount of invested share capital). Where in the case of LLC the board members are jointly and severally liable to the company's creditors due to ineffective execution from the enterprise's assets.

Finally, it is also essential that in the limited liability company all share holders may keep the direct supervision over managing the enterprise. Where in the case of a joint stock company this situation is rather completely excluded.

To sum up this part, it should be noted that both the limited liability company andjoint-stock one have certain advantages and disadvantages. Based on only legal analysis, it is clearly hard to answer on the question, which of these forms of business activity will allow for the most efficient company's asset management. Therefore to answer on a question asked at the beginning of this article, it is necessary to conduct a detailed economic analysis.

4. ANALYSIS OF MANAGEMENT EFFECTIVENESS OF POLISH CAPITAL COMPANIES

The economic analysis (mentioned in 2nd chapter) was performed on a sample of 8 joint stock companies (J-S.C.) and 8 limited liability companies (LLC), which was running its businesses constantly in years: 2001 – 2010. The mentioned companies were deliberately chosen from a population of Polish non-stock companies. To analysis have been undertaken firms from 6 different sectors of Polish economy: 1) metal industry, 2) construction sector, 3) electric, electromechanical and energy industry, 4) wood and paper industry, 5) chemical and petrochemical industry and 6) trade branch. The main goal of this study was to verify which group of firms better manage its assets. In research as an indicator of effectiveness of assets management was undertaken the level of market value growth of analyzed companies.

In order to describe the growth of market value of analyzed firms a parametric measure *P*, basing on the Q Tobina ratio ([12] and [17]) was used, define on the basis of Economic Value Growth Added measure, with the following pattern[1]:

$$P = \frac{EVGA_E}{A} = \frac{NOPAT - IC * WACC}{A} \tag{1}$$

[1] Own elaboration on the basis of the literature of the subject: [6, p. 156], [7, p. 6] and [13].

where: $EVGA_E$ – Economic Value Growth Added (for owners), *NOPAT* – economic profit before taxation (with CIT), *IC* – invested capital in a company (Equity plus Debt), *WACC* – weighted average cost of capital, *A* – book value of assets.

Measure described with equation No. (1) can be used as an approximation of portion of a yearly growth of market value of certain company. This fact shows equation No. (2) [15]:

$$MVGA = \sum_{t=1}^{\infty} \frac{EVGA_t}{(1+WACC)^t} \tag{2}$$

where: *MVGA* – Market Value Growth Added, *t* – number of year, in which company runs its business, $EVGA_t$ – yearly $EVGA_E$.

Because of its partial "simplicity" and explaining power $EVGA_t$ measure has been chosen to final analysis. And to bring to a common level all values of $EVGA_t$'s calculated for companies from different branches, final $EVGA_t$ indicators was divided by enterprises yearly level of assets.

4.1. CHANGES IN MARKET VALUE GROWTH OF JOINT STOCK COMPANIES (J-S.C.) AND LIMITED LIABILITY COMPANIES IN LAST 10 YEARS

Here detailed results will be presented only for 4 companies from 2 different branches. First there will be described outcomes achieved for 2 different metal industry companies: CNPEP RADWAR Joint-Stock Company (J-S.C.) and FABRYKA BRONI LUCZNIK Limited Liability Company (LLC), which operates on metal industry market. The first one produces mainly radiolocation devices, artillery and air defense systems. The second one manufactures mainly hand guns and rifles. Their product's profiles are very similar, that's why their *EVGA*'s can be compared. Figure 1 presents results achieved for selected companies.

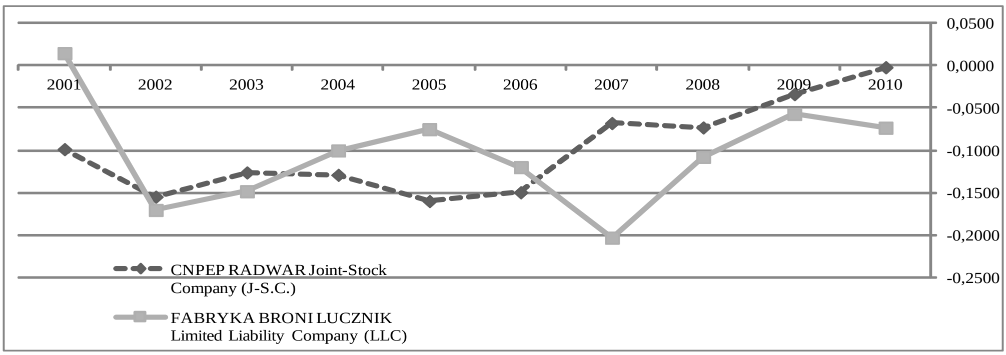


Fig. 1. An analysis of EVGA/A for metal industry companies

From figure 1 we can observe that from year 2001 till 2002 both companies' Economic Value Growth Added was decreasing. Next in years 2002 – 2004 theirs values was increasing. In 2004–2005 EVGA of FABRYKA BRONI LUCZNIK LLC was rising, but EVGA of CNEP RADWAR J-S.C. was declining. This situation changed into opposite in period 2005–2007. And finally from 2007 EVGA of both companies was increasing.

In this situation we noticed that general trend of growth of mentioned firms is almost the same. Only in some periods one company was "better" than the another, exactly before the global economic crisis in years 2004–2006 "better was behaving" *EVGA* of FABRYKA BRONI LUCZNIK. But from 2007 its general situation seemed to be worse than CNEP RADWAR. From presented results we can't say that one company was better managing its assets than the another.

Second sector for which achieved outcomes will be described is trade branch. Here EVGA was calculated for CITODAT J-S.C. and VIDICON LLC. Figure 2 presents the results.

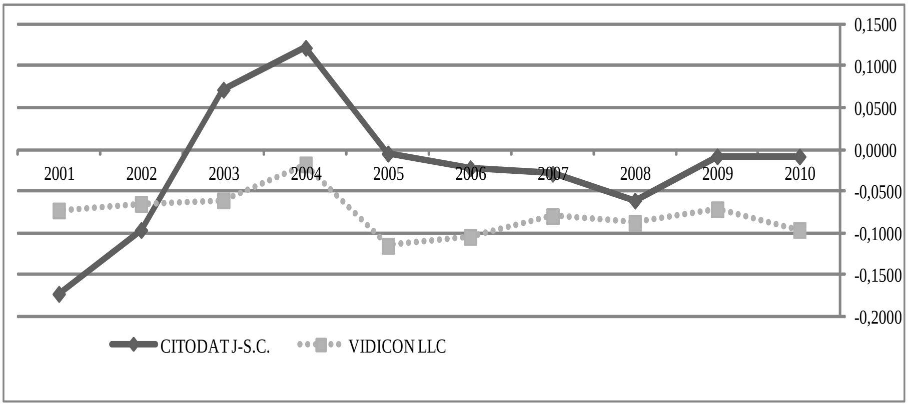


Fig. 2. An analysis of EVGA/A for trade branch companies

From figure 2 we notice, than in case of both companies changes in their EVGA are almost the same. In years: 2001 – 2004 values for CITODAT and VIDICON was rising, than in period: 2004 – 2008 they was decreasing and finally from 2008 they was increasing. Only different is that the CITODAT J-S.C. from year 2003 had got higher level of its parametric growth of economic value added, although it "started" from lower stage. It could prove, that joint-stock company might be more effective in context of its assets management.

In case of different 3 groups of companies similar behaviors were observed. The tendencies in changes of *EVGA* was almost the same in case of 12 next companies. Although in case of construction sector Limited Liability Company turned out to be more valuable than Joint-Stock company, but in case of wood and paper industry Joint-Stock Company proved to be slightly more effective.

4.2. COMPARISON OF COMPANY'S MARKET VALUE MANAGEMENT OF JOINT-STOCK COMPANIES AND LIMITED LIABILITY COMPANIES

The aim of last part of conducted study was to compare results achieved for both 8 J-S.C. and 8 LLC companies. During research there were calculated average values of Economic Value Growth Added in period 2001–2010, for every company from a sample of 16 enterprises. Summary outcomes presents figure 3.

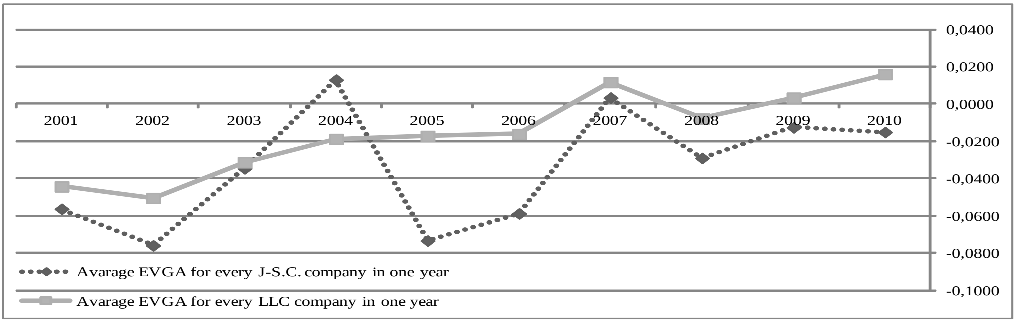


Fig. 3. Average EVGA/A for every J-S.C. and LLC company in certain year, in period 2001–2010

From figure 3 we can notice, that general direction of changes of average $EVGA_t$ was the same for both types of companies in analyzed period. From year 2002 till 2010 average approximation of firm's market value was constantly growing. Of course changes in $EVGA_t$ level was grater in case of J-S.C. companies and in 2004 this value was even higher than in case of LLC companies. But in the same time from year 2005 till 2010 average growth of J-S.C. firm's market value was lower than the LLC enterprises growth value.

5. CONCLUSIONS

Summing up we can say, that achieved results for a sample of 16 non-stock companies, don't determine entirely which type of companies better manage their assets, from value maximization point of view. The partial results, achieved in sectorial pairs (described in section 3.1) are showing, that in some branches J-S.C. companies better manage their market value, than the LLC ones. These outcomes are suggesting, that choice of legal form of business conducting can be influential in case of enterprise management. Although final analysis, performed on a sample of 16 enterprises are suggesting partly something opposite. Achieved average approximations of market value growth of chosen enterprises are higher for LLC ones since 2005 till 2010. However general tendency of changes of average $EVGA_t$ in every years: 2001–2010 was the same for both types of companies (not rated on stock exchange). These results are showing partially, that choice of legal form of business conducting doesn't have influence on effectiveness of enterprise's management.

In conclusion we can say, that obtained outcomes of analysis do not indicate an unambiguous answer, on a question posed at the beginning of this article. Although study was performed in detail for stable and economically strong companies (from different branches), the results don't suggest whether it is better to choose Joint-Stock Company or Limited Liability Company to conduct medium- and large-scale business.

Because all mentioned legal forms of enterprises have got their advantages and disadvantages, and for some companies one of them is better than the another, more detailed and wide range analysis should be performed. That's why further study need to be carried out in field of comparison of the market value growth between Joint-Stock Companies rated on a Warsaw Stock Exchange (GPW [16]) and Limited Liability Companies (not rated on a Stock Exchange). Achieved results, for a big sample of enterprises (about 200 firms) should give an answer, whether one type of analyzed forms of enterprises better manage their assets and to which extent it could be observed.

REFERENCES

[1] *Act of 15 September 2000. Commercial Companies Code* (unified text, J.L. 2013, item 1030, as amended).
[2] *Act of 20 April 1892 on the limited liability companies* (unified text, J.L. 1982, No. 82, item 602, as amended).
[3] *Act of 29 September 1994 on accounting* (unified text, J.L. 2013, item 330, as amended).
[4] COPELAND T., KOLLER T., MURRIN J., *Wycena: Mierzenie, kształtowanie i ocena wartości firm* (*Measurement, creation and firm's value evaluation*), Wig Press, Warsaw 1997.
[5] *Decree of 8 February 1919 on the limited liability companies* (unified text, J.L. 1919, No. 015, item. 201, as amended).
[6] SUSZYŃSKI C. (Ed.), *Przedsiębiorstwo, Wartość, Zarządzanie* (*Enterprise, Value, Management*), PWE, Warsaw 2007.
[7] GRUSZCZYŃSKI M., *Ekonometria nadzoru korporacyjnego* (*Econometrics of corporation supervision*), Works of Econometrics Institute, State Research, No. 03/S/0009/03, Szkoła Główna Handlowa, Warsaw 2003.
[8] KRUCZALAK K., *An outline of commercial law*, Lexis Nexis, Warsaw 2004.
[9] *Regulation of 27 October 1933. Regulations introducing Commercial Code* (unified text, J.L. 1933 No. 82, item 601, as amended).
[10] *Regulation of 3 December 1930 on the application of the law on joint stock companies to the authorities and to the liquidation of mutuals* (unified text, J.L. 1930, No. 66, item. 523, as amended).
[11] *Regulation of the President of the Republic of Poland of 22 March 1928* (unified text, J.L. 1928 No. 39, item 383, as amended).
[12] VOGT S., KEITH M.H., *On "q"*, Financial Review, 1996, No. 31, 265–286.
[13] WEAVER S.C., *Measuring Economic Value Added®: A Survey of the Practices of EVA® Proponents*, Journal of Applied Finance, 2001, 11, 1 (winter edition), 50–60.

[14] Web portal of *NewConnect* “alternative” Polish stock exchange: http://www.newconnect.pl/index.php?page=o_rynku- date of data upload 4 June 2014
[15] Web portal of Stern Stewart & Co: http://sternvaluemanagement.com/?content=proprietary&p=mva- date of data upload 4 June 2014.
[16] Web portal of Warsaw Stock Exchange: http://www.gpw.pl/o_spolce_en- date of data upload 4 June 2014.
[17] WILBUR L.G., BADRINATH S.G., *On the Measurement of Tobin's q*, Journal of Financial Economics, No. 44, 1997, 77–122.
[18] WRIGHT P., BLACK A., DAVIES J., *In search of shareholder value, second edition*, Price Water House Coopers, 2000.

electricity purchase strategy portfolio
optimization, risk hedging

Przemysław PUCHALSKI, Zofia WILIMOWSKA*

MODEL AND ITS COMPUTER SIMULATION FOR ELECTRIC ENERGY PURCHASE ALLOCATION AND RISK EVALUATION ON THE POLISH MARKET

Thanks to the changes in the Polish regulations large consumers can be participants in the energy market. Currently, according to the chosen strategy, it is possible to purchase energy within a tariff system from any trading company or directly from the market through energy exchange (in Poland it is TGE – Towarowa Giełda Energii). In order to fully take advantage of the present conditions, consumers who decide to actively participate in market have to seek purchase strategies which on the one hand reduce cost and at the same time limit risks. In this paper, attempt is made to verify if a purchase strategy which uses utility function to define the contribution of forward and spot contracts can effectively manage the purchase process on the Polish market. In the opinion of the authors, the simulated results presented confirm the usefulness of the proposed methodology.

1. INTRODUCTION

Over recent years, a lot of change has been done on the Polish energy market. In the past, there was a state monopoly in the area, whereas today a liberalization process follows in all the energy sector. Nobody could choose providers of energy, and nowadays, thanks to the implementation of TPA rules (True Part Access), which means that we can choose a seller with ease, consumers' decisions have influence on the energy cost. In the case of large consumers, it is even possible to purchase energy directly from the energy exchange, i.e. Towarowa Giełda Energii (thanks to intermediation of brokerage houses).

The simplest way of purchasing energy is to choose any trading company (tariff purchase) which has some advantages. The consumer knows the exact price of energy without having to consider the risk of bad estimation of energy load (when purchased

* University of Applied Sciences, Armii Krajowej 7, 48-300 Nysa; Wrocław University of Technology, Institute of Industrial Engineering and Management, Smoluchowskiego 25, 50-372 Wrocław, Poland; e-mail: zofia.wilimowska@pwr.edu.pl

energy is lower or higher than the actual load) and with no need to search for the cheapest energy at a given risk. In this case, all risks are transferred to the trading company but the fixed price for the consumer is high.

The consumer who rejects tariff has to take into account many variables and risks. There are many types of contracts on TGE and each of them features a different price, variance and execution time. Participation in the market implies the need to prepare a forecast load and makes it necessary to bear the costs of balancing (when purchased energy does not exactly cover the demand). In exchange for a way of purchasing that is more complicated and carries a certain risk, it is possible to obtain energy at a lower price. If a company decides to buy contracts on the exchange market, it faces a problem of management of electric energy purchase allocation. The strategy where most of demand is met by the spot market (RDN, Rynek Dnia Następnego on TGE) causes that the energy cost is very variable and difficult to predict but on the other hand, it is a solution where the consumer cannot forecast its load well in advance. When the load can be estimated much earlier, a futures contract also should be taken into account. On TGE, there are forward contracts which run at different time frames – from one week to one year. Forward contracts can be signed well ahead of time, e.g. monthly contracts are traded for six months before delivery.

Large consumers, who build a long-term purchase strategy, have to consider how much electricity should be bought on the spot market and how much in forward contracts. Different strategies can give various costs of energy, depending on market fluctuations. One of the approaches is the algorithm presented in [2]. In this article, we will verify if the proposed methodology could be applied on the Polish energy market, for Polish large consumers.

2. ALGORITHM AND SIMULATION ASSUMPTIONS

The purchase allocation problem can be regarded as a portfolio selection where the task is to attain maximum profit and minimum risk. The method proposed in [2] suggests analyzing a saving of portfolio built by forward contracts and spot contracts and comparing it with the cost of energy bought in tariff:

$$\pi = rq - [s(q - Q) + FQ], \tag{1}$$

where: π – saving of portfolio, r – electricity retail price (tariff) (PLN/MWh), q – total electricity demand (MWh), Q – the allocation amount on forward contract (MWh), s – spot price (PLN/MWh), F – forward price (PLN/MWh).

To find the optimal amount of forward contracts and spot contracts, we will maximize a utility function:

$$U(\pi, k) = E(\pi) - \frac{kV(\mu)}{\pi_0}, \tag{2}$$

where: k – the coefficient of risk aversion ($k \geq 0$), $E(\pi)$ – is the expected profit (saving) corresponding to a portfolio selection, $V(\pi)$ – is variance of the expected profit, π_0 – is energy purchase cost in tariff.

The utility function (2) can be represented by:

$$U(\pi, k) = -\frac{kV(s)}{(r-F)q}Q^2 + \left[\frac{2qkV(s)}{(r-F)q} + E(s) - F\right]Q + rq - qE(s) - \frac{kV(s)q}{r-F}. \quad (3)$$

In order to find the maximum utility, we equate the first derivative to zero:

$$\frac{dU}{dQ} = -\frac{2kV(s)}{(r-F)q}Q + \frac{2kV(s)}{r-F} + E(s) - F = 0. \quad (4)$$

After transformations, the optimal amount of forward contract Q^* is equal:

$$Q^* = q\left(1 - \frac{(F - E(s))(r-F)}{2kV(s)}\right). \quad (5)$$

After further analyzing:

$$Q^* = \begin{cases} q, & E(s) > F, \\ q\left(1 - \frac{(F-E(s))(r-F)}{2kV(s)}\right), & E(s) < F. \end{cases} \quad (6)$$

What is more, additional conditions can be added for equation (6):

- if $F > r$ then $Q^* = 0$, because the equation will lose its sense;
- $Q^* <= 90\%\ q$, this condition is not necessary, but it seems logical to have a marginal space for the spot market;
- If $k \geq (F - E(s))(r - F))/(2V(s))$ then $Q^* = 0$, because the equation will lose its sense;
- If $E(s) = F$ than Q^* can take any value between 0 and q, we determine value 90% q.

For the achieved portfolio, risk is evaluated. As a measure of contracts risk Value at Risk is computed. There are several methods of *VaR* estimation – in this article the delta-normal approximation is used [3]. The method assumes that the profits of the investments in a portfolio are normally distributed and consequently the portfolio profits are also normally distributed with variance equal to a weighted average of the covariance of the portfolio profits:

$$VaR_{95\%}(\pi^*) = E(\pi^*) - 1{,}65\sigma(\pi^*), \quad (7)$$

where $E(\pi^*)$ is the expectation of profit π^*, where π^* is the profit corresponding to the allocation of Q^* on forward contract, $\sigma(\pi^*)$ is standard deviation of this portfolio, 1,65 is a constant that corresponds to the specified one-tailed confidence interval for the normal distribution (here 95%). $VaR_{95\%}(\pi^*)$ means the value of profit (saving) which will be exceeded with a 95% probability.

There is a second measure of risk calculated. It is also Value at Risk, but it is computed for portfolio π':

$$\pi' = s(q - Q^*) + FQ^*, \tag{8}$$

VaR for π' is estimated also using the delta normal approximation with the same confidence level (95%). $VaR_{95\%}(\pi')$ mean an increase of costs, which will be exceeded with a 95% probability.

3. SIMULATION ASSUMPTIONS

The purpose of this article is to verify if the algorithm could be applied to the Polish energy market. The process of purchase allocation requires performance of the following independent tasks:

- carrying out a forecast of the electricity process spot price,
- carrying out a forecast of the forward contract price,
- carrying out a forecast of the electricity demand,
- determining the coefficient of risk aversion *k*,
- the energy cost in the tariff must be known.

Forecasting of the electricity price process, both in long and short term, is a vast subject, which lies beyond the scope of this article. There are many publications, which propose dedicated models, e.g. Seasonal ARIMA or Autoregressive GARCH time series, as well as Regime-Switching models [1], [4]. To avoid propagation of errors of forecast that may affect the assessment of the method's suitability, historical data of the electricity spot price will be used.

Simulations will be carried out using the BASE and PEAK monthly contracts (separate simulations), which are traded on the TGE. The PEAK contract guarantees a certain amount of energy (every hour the same) during peak hours from 7:00 to 22:00 on work days. The BASE provides specified amount in every hour of the day, 7 days a week. Monthly contracts are traded on the TGE for six months before delivery. In the simulations, the forward price will be established as a volume-weighted average transaction. A forecast of the forward price is not used for two reasons. First, to avoid the effect of the assessment on the suitability of the method. Secondly, it seems reasonable to assume that consumers who consider monthly contracts make their decisions with less advance time than six months. Both spot prices and forward contracts comes from the period 2011-01-01 to 2014-05-31.

The energy demand of large consumers is unknown to the authors of this article. For the purposes of article, it was assumed that:

- for the scenario with the BASE contract – the energy load is constant during every hour of the day, equaling 5 MWh,
- for the scenario with the PEAK contract – the energy load is constant during peak hours from 7:00 to 22:00 on work days, equaling 5 MWh.

Large consumers negotiate the energy price. Exact values are not published, therefore tariff prices are arbitrarily established. For the BASE scenario it is 230 PLN/MWh and for the PEAK is 250 PLN/MWh.

The objective of this article is to verify if the algorithm can be used to maximize the benefits while minimizing the risks. For this purpose, both scenarios will be conducted with different values of the risk aversion *k*. The parameter takes values from the set of {0.3, 0.5, 0.7, 1 ,1.5, 2, 3, 5, 7, 10}.

4. RESULTS

In accordance with the previous part of article, simulations were done for two scenarios (BASE and PEAK). Hours of demand correspond to appropriate futures contracts. Data comes from the TGE, from period of 2011-01-01 to 2014-05-31.

The first figure shows the average spot price, the price of forward contracts and the monthly variation of the spot price for the BASE scenario in time. Figure 2 shows the

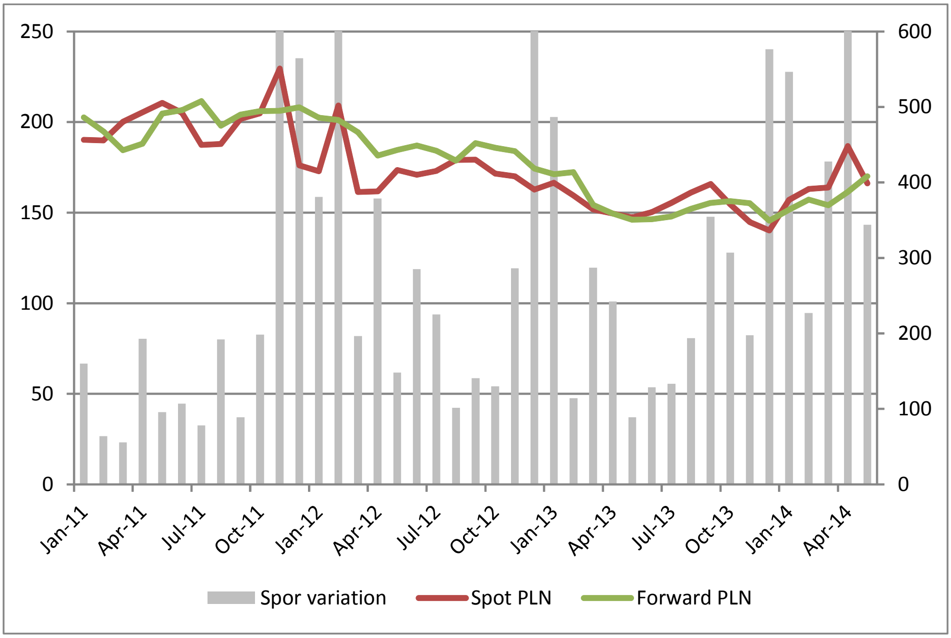


Fig. 1. The BASE scenario: the spot price, the spot variation and the forward contract price

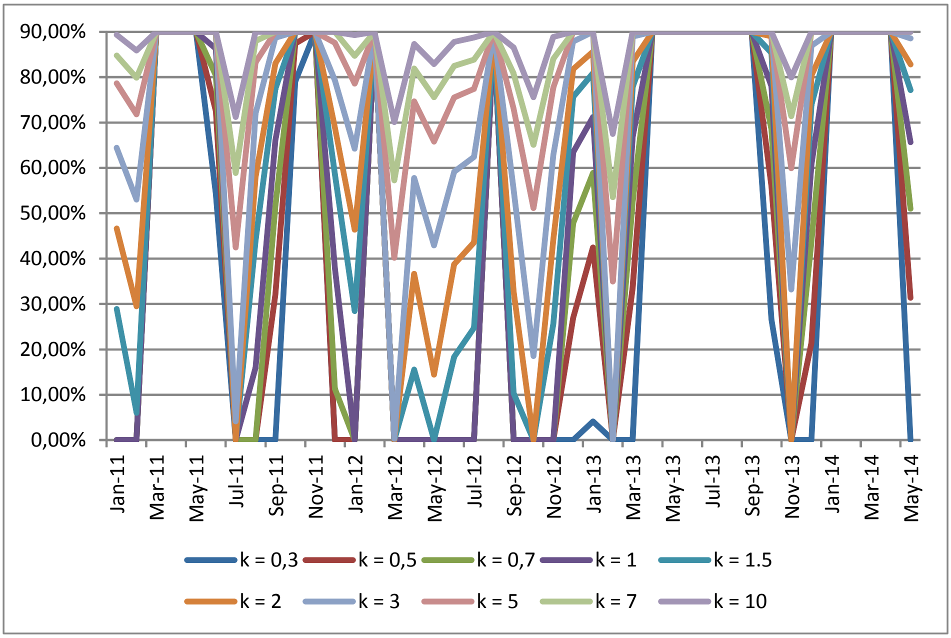


Fig. 2. The BASE scenario: the aversion of risk (*k*) and the share of forward contracts

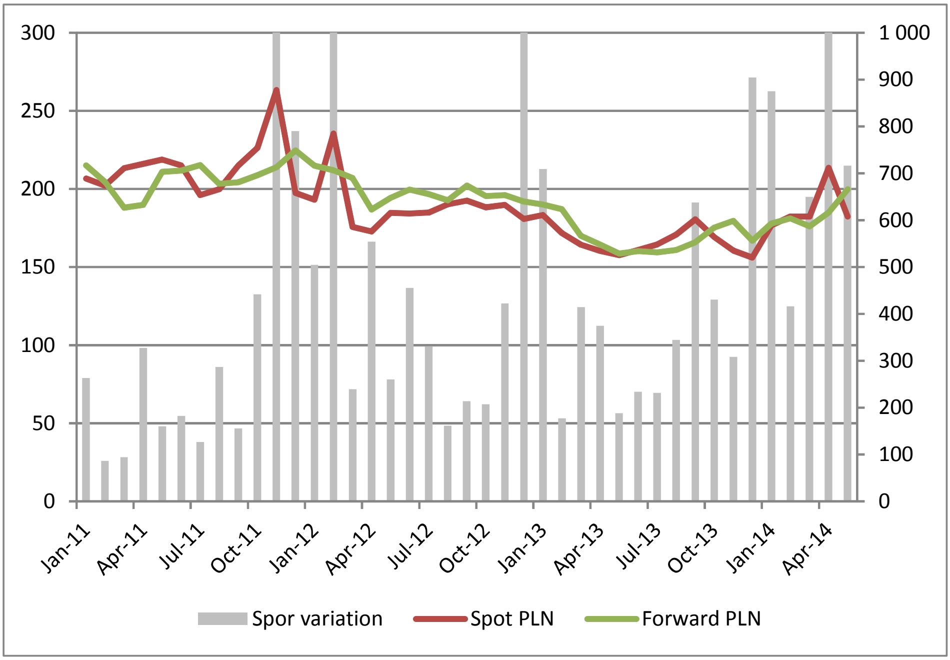


Fig. 3. The PEAK scenario: the spot price, the spot variation and the forward contract price

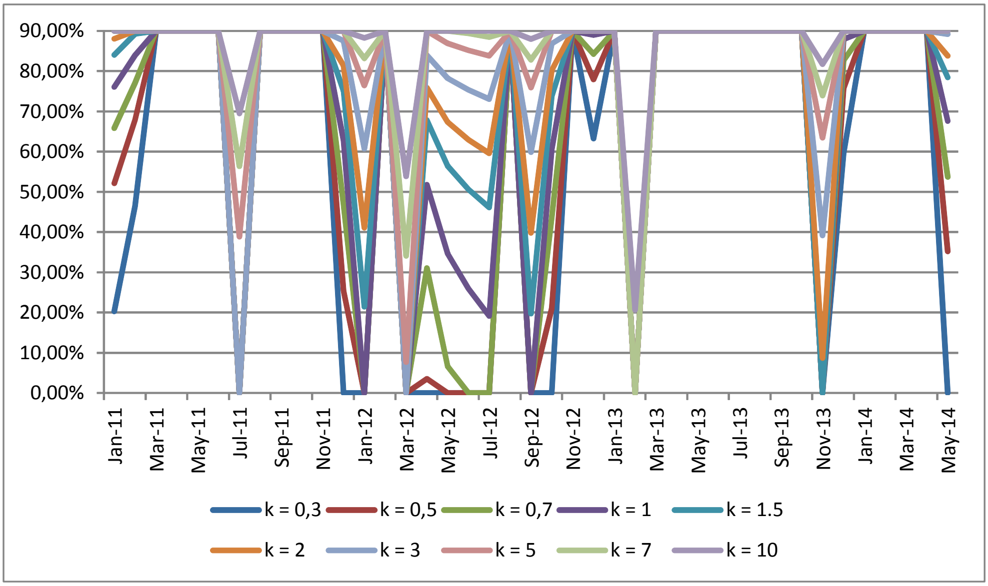


Fig. 4. The PEAK scenario: the aversion of risk (k) and the share of forward contracts

share of the forward contract for different *k* values. An interesting observation is that when variation of the spot price reaches very high values, the spot price is much higher than the forward (months 2011-11,2012-02, 2014-04). Of course according to the algorithm, allocation of forward contracts is 90%. Additionally, as expected over the months where the price of the forward contract is much higher than the spot price, the participation of the spot market is maintained at a high level even for bigger values of the risk aversion *k* (months 2011-12, 2012-12). The same conclusion can be drawn when the variance of the spot price is low (months 2012-03–2012-07 or 2012-10–2012-11). Figures 3–4 contain data for the PEAK scenario. Conclusion are the same as for the BASE scenario, please check the data from months e.g. 2011-12, 2012-12.

Next figures contain information about hypothetical savings and risk measures when the consumer resigns from the tariff energy. All values (portfolio saving, *VaR* for π^* and π') are presented in percentages because they are divided by the cost of the energy in tariff. Only the results for the BASE scenario are shown because conclusions for the PEAK simulations are the same. As expected, when the risk aversion k rises (and the the amount of forward contracts increase), profit from the portfolio (economy) and risk decreases. Furthermore, when data from relevant months is compared (e.g. 2011-08 and 2011-12 or 2012-12 and 2013-01), we notice cases where the expected profits are the different but $VaR_{95\%}(\pi')$ are the same. These months show us the influence of the variation on risk measures.

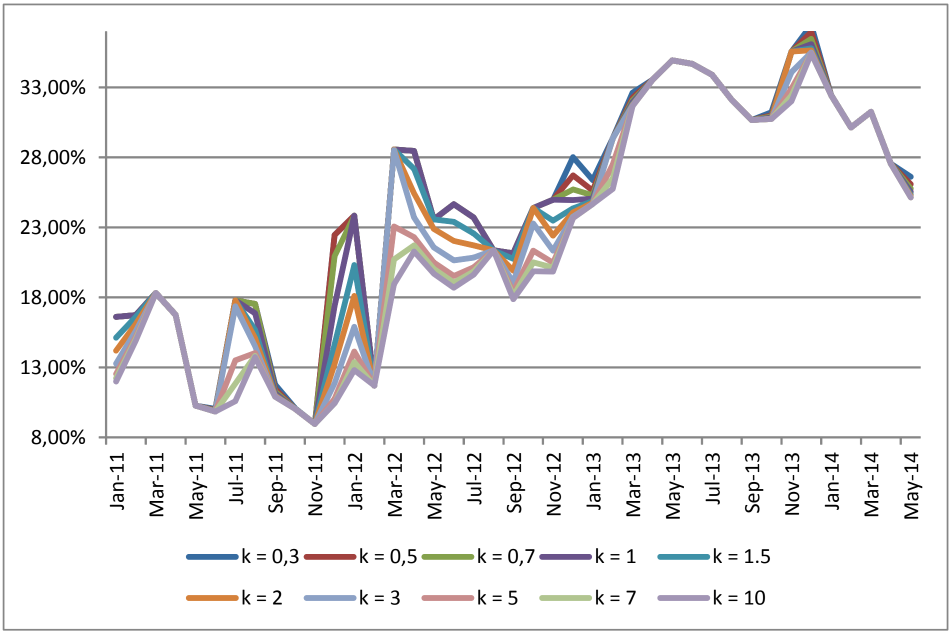


Fig. 5. The BASE scenario: the savings for the portfolio (%)

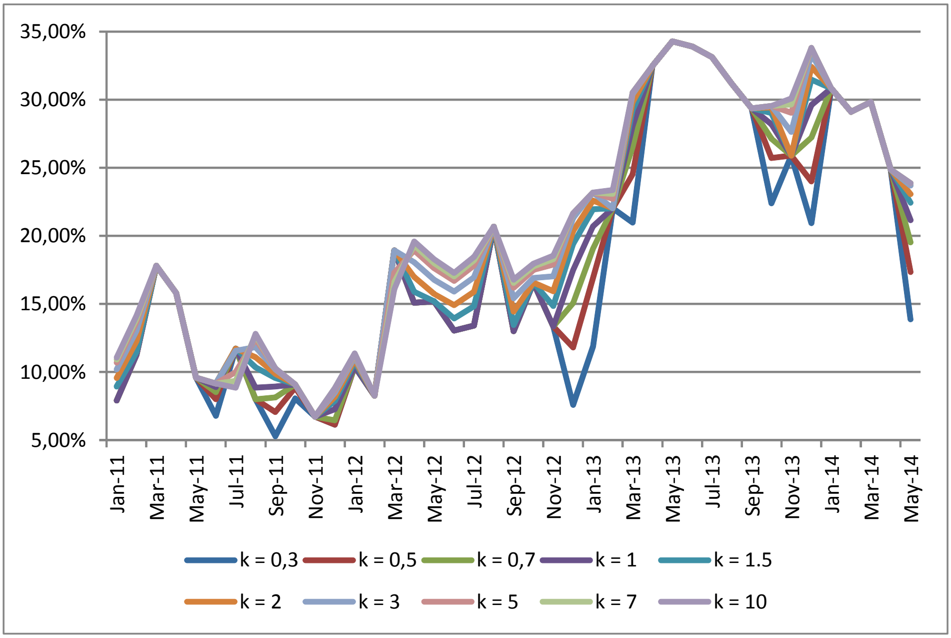


Fig. 6. The BASE scenario: the Value at Risk (π) and the risk aversion (k)

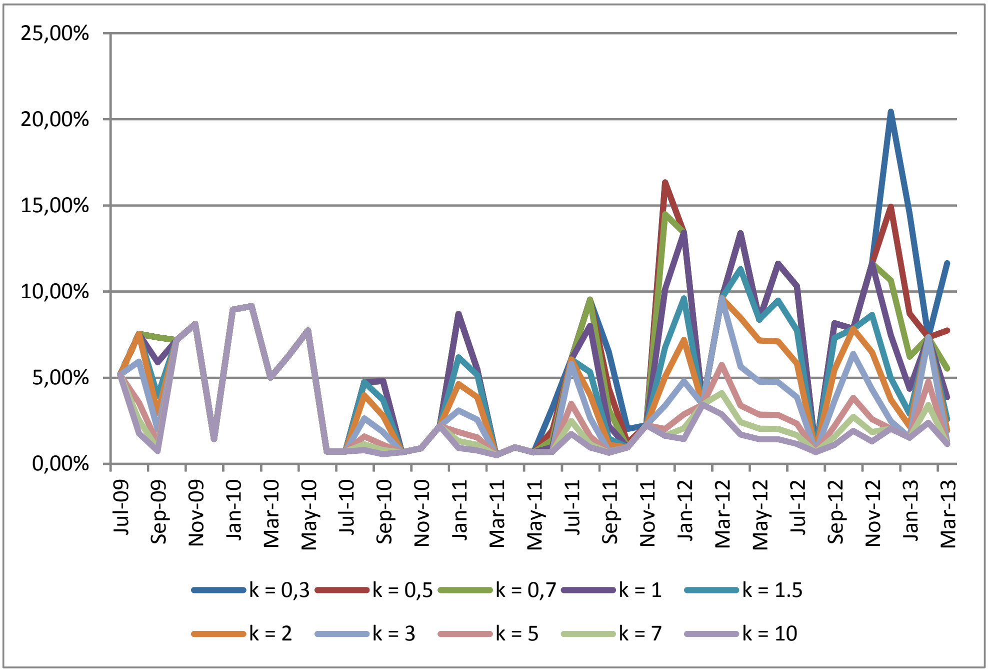


Fig. 7. The BASE scenario: the Value at Risk (π') and the risk aversion (k)

5. SUMMARY

It must be emphasized that the simulations presented carry many simplifications. In reality, the impact on the outcomes would have the quality of the spot price and the forward contract forecasts. Additionally, we cannot forget about the transactions cost, as well as the influence of the disparity between the forecasted demand and the actual demand. In spite of all the simplifications, the results shown in this article confirm that the proposed algorithm of purchase can be useful for the Polish energy market.

When a large consumer knows the coefficient of the risk aversion k, the algorithm determines the number of the forward contracts, in accordance with the variation of the market.

Undoubtedly, the proposed methodology has some disadvantages. The first of them is that Value at Risk is different every month, according to market fluctuations. It seems that a more interesting approach one where *VaR* does not exceed a specified value. The second desired alternative could be a method where the maximum cost is given and the algorithm minimizes the risk.

REFERENCES

[1] BROCKWELL P.J., RICHARD A.D., *Introduction to Time Series and Forecasting*, Springer--Verlag, 1996.
[2] MING Z., GENGYIN L., YIXIN N., *Long-Term Electricity Purchase Allocation Strategy and Risk Evaluation in Power Markets*, [in:] Power Systems Conference and Exposition, PSCE ’06 IEEE PES, Oct. 29, 2006–Nov. 1, 2006.
[3] WERON A., WERON R., *Giełda Energii. Strategie zarządzania ryzykiem*, CIRE, Wrocław 2000.
[4] WERON R., *Modeling and Forecasting Electricity Loads and Prices. A statistical Approach*, Wiley, 2006.

Economic Order Quantity,
mathematical programming, optimization

Jacek BOROŃ*
Paweł BARTYLA**

OPTIMIZATION OF WAREHOUSE SUPPLY BY THE METHODS OF MATHEMATICAL PROGRAMMING

Building materials warehouses and construction sites are places where contemporary managers have to face various decision-making problems related to production, management policy, or the right course of action. It is often the case that, due to many variables, complex selection criteria, and the constraints on the solution space, a seemingly uncomplicated problem becomes difficult to solve, and finding even a satisfactory solution may take a lot of time or turn out almost impossible with just "pen and paper". The article presents, in a friendly to building materials warehouse managers, the practical application of methods of mathematical programming as tools for computer-aided purchases organization. The problem is illustrated example of optimizing purchases for small warehouse.

1.1. INTRODUCTION

Using a spreadsheet-type of software, you can build a transparent mathematical model of a particular decision-making problem which reflects the problem's true nature in a faithful or sufficiently accurate manner. It is usually possible to describe the real relationships in the form of mathematical functions, and the same goes for the criterion itself, which is an indispensable measure for finding a solution or a course of action within the set of feasible solutions. Solving an optimization problem, i.e. "combing" the set of feasible decisions and choosing the best one, is possible thanks to the Solver® add-in [1]. A properly conducted optimization process allows us to make the right decision which was not "visible" before but is better than the previous

* Department of Civil Engineering, Faculty of Building Engineering, Wrocław University of Technology, Wrocław, Poland, e-mail: jacek.boron@pwr.edu.pl
** Pekabeks PREF sp. z o.o., Poznań, Poland.

one (the one based on assumptions or intuition). It has the potential to bring substantial cost savings in the area of production or resource management, and the bigger the scale of the problem, the bigger the savings [2], [3].

The aim of this article is to present how a spreadsheet can be applied to building and solving specific and, more importantly, real-life decision-making problem related to the management of a single warehouse company.

2. THE PROBLEM OF ORDERING POLICY IN A BUILDING MATERIALS WAREHOUSE

Let's assume that a certain building materials warehouse operates in the N city in Lower Silesia (Poland). The warehouse consists of a hall for storing materials that require weather protection and of a partially covered storage yard for storing weatherproof materials of more sizeable overall dimensions. Every month, the warehouse manager has to face the same dilemma: "What will be the demand for building materials next month and what quantities of these materials should I order to keep the warehouse upkeep costs to a minimum?". In literature, a problem of this type is referred to as the *Economic Order Quantity* (*EOQ*). The easiest way to illustrate the problem is through the chart below:

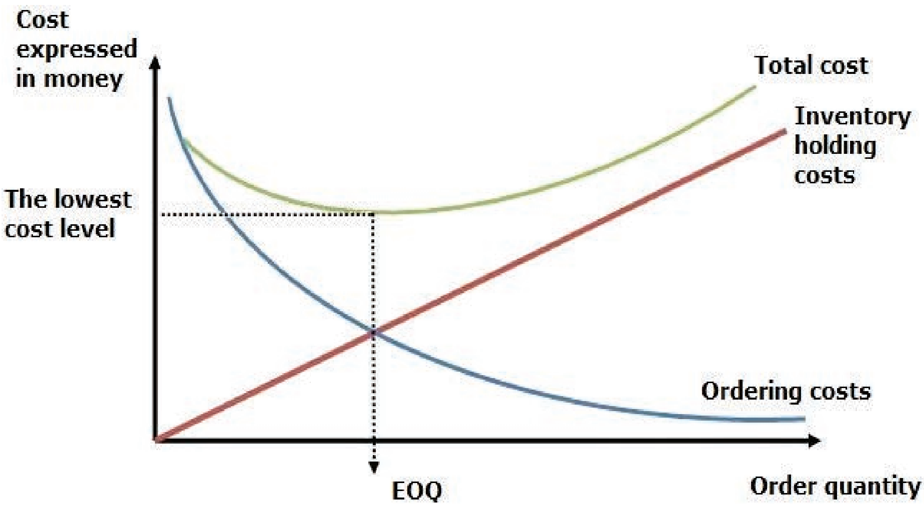


Fig. 1. The chart illustrating the concept of Economic Order Quantity (EOQ)

In a nutshell, the costs of maintaining a building materials warehouse primarily include the following:

– ordering costs, e.g. fuel costs (theoretically, the more you order at once, the smaller the cost per one unit of material);
– inventory holding costs in the warehouse, e.g. tax per 1 square meter of the storage area or wages for employees handling the materials (i.e. employees who unload, move, organize the inventory and make sure it is kept clean). There is a formula which can be used to calculate the optimal quantity of the product to order, knowing the one-time cost per order, the demand, the unit purchase price, and the holding cost of the product (measured as a percentage of the unit purchase price). However, the above formula does not include limited storage space and, therefore, does not take into account the mutual relations between other materials. The table below (Table 1) shows a sample inventory of products stored on a storage yard and their descriptions. For this small warehouse, we will attempt to build a mathematical model which will serve to determine the quantities of materials to order each month so that the ordering and holding costs (and, therefore, the warehouse upkeep costs) are at their minimum. One limitation which makes the use of the EOQ formula impossible is the limited storage area of only 300 sq. m net. This is the yard's storage area less traffic routes and any other type of non-storage area (i.e. where products cannot be stored).

In addition, the following data for our model warehouse has been assumed:
– tax per 1 sq. m of surface area = PLN 0.74;
– one month's salary for a warehouse employee = PLN 2,500.00;
– average monthly cost of fuel for a forklift = PLN 700.00;
– transportation cost for ordered materials = PLN 1.00 per 1 kilometer.

2.1. DEVELOPING THE MATHEMATICAL MODEL

Based on the above parameters, we can now proceed to build the mathematical model of EOQ optimization. The decision criterion is the minimum possible monthly cost of warehouse operation; therefore, what needs to be calculated are the expenses related to each product (see Fig. 2). The demand for each product is provided based on the statistics from previous years and the warehouse manager's predictions and assumptions.

The cost of the order:
– transportation cost x distance from manufacturer x 2 x 1,3 vehicle depreciation.

The cost of storage:
– surface storage unit x [tax+(fuel for forklift + salary for a warehouse employee)/ storage area].

Table 1. The inventory of warehouse products stored on a partially covered storage yard

The range of construction materials stored in the storage yard									
Construction material	Distance from manufacturer [km]	Dimensional characteristics w × d × h	Type of unit	Unit sizes	Manner of storage	Surface occupied [m^2]	Surface area per unit [m^2]	Price (per piece) [PLN]	Quantity per pallet (or in the stack)
Airbricks MAX	120	188 × 288 × 220	pallet	1.2 × 0.8 m	up to 3 pallets vertically	1.76	0.587	4.38	96
Airbricks SILKA E24	230	240 × 333 × 199	pallet	1.2 × 0.8 m	up to 3 pallets vertically	1.76	0.587	6.07	45
Airbricks YTONG	230	Energo 300 × 599 × 199	pallet	1.2 × 0.8 m	up to 2 pallets vertically	1.76	0.880	15.04	40
Full bricks ordinary	50	Class 25 250 × 120 × 65	pallet	1.2 × 1.2 m	up to 2 pallets vertically	2.56	1.200	1.48	408
Ceiling beams POROTHERM	120	16 × 6000 × 17.5	piece	6.0 × 0.16 m	only one stock 5x5	4.80	0.192	247.48	[–]
Flooring blocks POROTHERM	120	250 × 525 × 230	pallet	1.2 × 1.2 m	only 1 pallet vertically	2.56	2.560	9.83	50
Cement (in sacks)	300	Portland cem. 25 kg R 32.5	sack	[–]	only 1 pallet vertically	2.56	0.073	10.25	35
Ceramic roofing tiles Röben	400	[–]	pallet	1.2 × 0.8 m	up to 2 pallets vertically	1.76	0.88	4.00	240
Ceramic roofing tiles Braas	60	[–]	pallet	1.2 × 0.8 m	up to 2 pallets vertically	1.76	0.88	3.60	300
Formwork pine boards (20 mm)	40	200 × 6000 × 20	piece	[–]	up to 2 stocks vertically (5 × 12)	6.00	0.05	11.60	60
Roof battens	40	60 × 40 × 6000	piece	[–]	up to 2 stocks vertically (20 × 10)	7.20	0.018	9.60	200

Characteristics of building materials stored in the storage yard						
No	Construction material	Surface area per unit	Unit	Demand	Ordering cost	Inventory holding unit cost
[-]		[m²]	[-]	piece/month	[PLN]	[PLN/piece]
1	Airbricks MAX	0,587	pallet	150	312	6,69
2	Airbricks SILKA E24	0,587	pallet	30	598	6,69
3	Airbricks Ytong	0,880	pallet	60	598	10,04
4	Full bricks ordinary	1,200	pallet	20	130	13,69
5	Ceiling beams POROTHERM	0,192	piece	60	312	2,19
6	Flooring blocks POROTHERM	2,560	pallet	28	312	29,20
7	Cement (in sacks)	0,073	sack	400	780	0,83
8	Ceramic roofing tiles Röben	0,880	pallet	140	1 040	10,04
9	Ceramic roofing tiles Braas	0,880	pallet	80	156	10,04
10	Formwork pine boards (20 mm)	0,050	sack	450	104	0,57
11	Roof battens	0,018	piece	4 500	104	0,21

Fig. 2. The list of stored products with their key characteristics

The demand for each product is provided based on the statistics from previous years and the warehouse manager's predictions and assumptions. The next step is to determine the range of cells to be changed by Solver® (marked in grey) and linking them with costs and the surface area used by each product (see Fig. 3).

Let's define the range of cells L19:L29 as "ilosci" (quantity) and enter the initial formula = L19 x J19 into cell M19. However, as the warehouse products are sold gradually over time, the formula = L19 x J19/2 is much more correct. The same course of action needs to be pursued with the remaining cells in column M. In cell N19, you enter the formula = (H19/L19) x I19 for ordering cost and copy it into the remaining cells of this column. The "total cost" column is the sum of costs from columns M and N. Cell N33/34 should sum up the range of cells O19:O29 and be defined as "koszt_laczny".

The next step is developing an algorithm which calculates the surface area used by each product stored in the warehouse. In cell P19, you enter the formula =(L19 x F19)/2. The same goes for other cells of column P. Like in the case of holding costs, it is assumed that if there is 100% of the inventory on the storage yard at the beginning of the period (i.e. between one order and another), there is 0% at the end. Therefore, on average, there is 50% of the products on the yard throughout the entire period.

Naturally, it is possible to make a different assumption which will be more suitable for a particular warehouse. After summing up the surface areas taken up by different materials, the formula = SUM(P19:P29) should be entered into cell P33 and the cell should be defined as “zuzyta_powierzchnia” (surface used).

⌂	L	M	N	O	P	Q
12–14	SOLVER - nonlinear GRG model (starting point)					
15–16	Quantity	Inventory holding cost	Ordering cost	Total cost	Surface occupied	Ordering submitted every ...
17–18	[piece]	[PLN]	[PLN]	[PLN]	[m^2]	[month]
19	150	501,89	312,00	813,89	44,000	1,00
20	30	100,38	598,00	698,38	8,800	1,00
21	60	301,14	598,00	899,14	26,400	1,00
22	20	136,88	130,00	266,88	12,000	1,00
23	60	65,70	312,00	377,70	5,760	1,00
24	28	408,81	312,00	720,81	35,840	1,00
25	400	166,86	780,00	946,86	14,629	1,00
26	140	702,65	1 040,00	1742,65	61,600	1,00
27	80	401,51	156,00	557,51	35,200	1,00
28	450	128,33	104,00	232,33	11,250	1,00
29	4500	461,97	104,00	565,97	40,50	1,00
30						
31						
32						
33–34		Monthly cost of storage [PLN]		7 822,13	296,0	Warehouse surface [m^2]

Fig. 3. The mathematical model of optimization (hidden)
for the non-convex nonlinear problem

2.2. SOLVING THE ORDERING POLICY PROBLEM

Another step (see Fig. 4) is entering the data into the dialogue box of the optimization software. For the warehouse inventory model, the GRG Nonlinear model should be chosen. Since it is a non-convex model, the Use Multistart check box should be selected in GRG options (which automatically limits the population size to 1 000 and the (iteration) convergence to 0.00001). Other Solver® options should be left as default (earlier, such parameters were tested through trial and error).

The starting values in green cells (Fig. 3, range L19:L29) can be any non-negative integers (in practice, they are usually greater than or equal to 1). To avoid the risk of re-scaling, it is recommended not to use starting values which are bigger than the possible future solution. In this specific model, the starting values are based on the monthly demand (Fig. 2, the grey column). The final step is running the optimization software by clicking *Solve*. The results obtained are presented in Fig. 5.

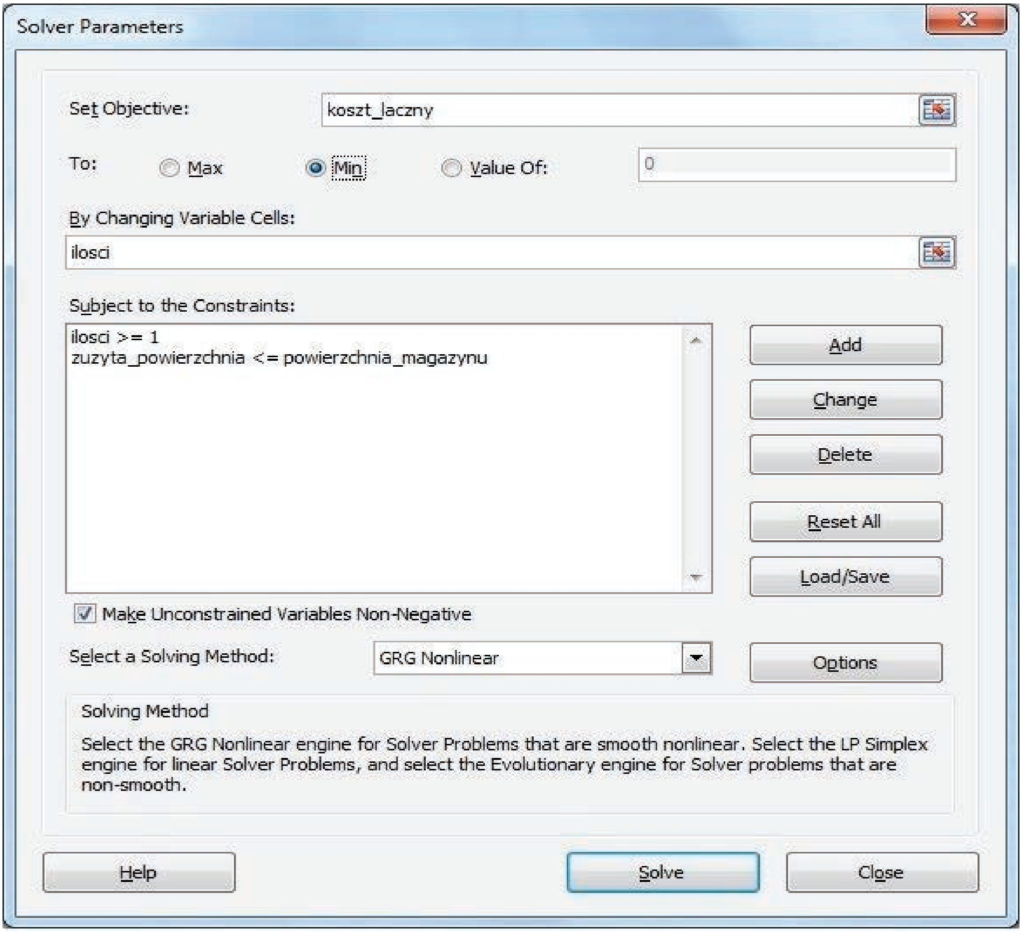


Fig. 4. The Solver® add-in dialogue box
– for the building materials warehouse ordering policy

SOLVER - nonlinear GRG model (solution)

Quantity [piece]	Inventory holding cost [PLN]	Ordering cost [PLN]	Total cost [PLN]	Surface occupied [m²]	Ordering submitted every ... [month]
116	386,97	404,66	791,63	33,924	0,77
72	239,58	250,54	490,13	21,004	2,39
83	414,97	433,95	848,93	36,380	1,38
19	130,45	136,41	266,86	11,436	0,95
128	140,01	146,41	286,42	12,274	2,13
24	349,24	365,22	714,46	30,617	0,85
846	352,79	368,93	721,72	30,928	2,11
167	835,94	874,18	1710,11	73,285	1,19
49	244,74	255,93	500,67	21,456	0,61
396	112,97	118,14	231,11	9,904	0,88
2088	214,34	224,15	438,49	18,79	0,46
	Monthly cost of storage [PLN]		7 000,52	300,0	Warehouse surface [m²]

Fig. 5. Solution of the problem related to the ordering policy in a building materials warehouse

3. CONCLUSIONS ON THE PROBLEM OF ORDERING POLICY

To conclude the analysis of this problem, it is worth mentioning that it is possible to obtain a different end result after changing the GRG parameters. This is due to the specific nature of non-convex nonlinear models and the applied Multistart algorithm. The problem has been solved dozens of times, with different starting points and different parameters. The lowest value of the costs obtained was **PLN 7,000.52**.

Therefore, it can be assumed with high probability that the above solution is the unquestionable globally optimal solution for the above-defined problem conditions. Having compared the obtained solution with the **Economic Order Quantity** shown in Figure 6, it is clear that the EOQ result is slightly better in terms of reducing the total cost, but it does not satisfy the most important and the only constraint - the constraint related to the available storage area.

⌂	L	M	N	O	P	Q
12–14	**Economic Order Quantity (EOQ)**					
15–16	Economic quantities of building materials	Inventory holding cost	Ordering cost	Total cost	Surface occupied	Ordering submitted every …
17–18	[piece]	[PLN]	[PLN]	[PLN]	[m^2]	[month]
19	118	395,72	395,72	791,43	34,692	0,79
20	73	245,00	245,00	490,01	21,479	2,44
21	85	424,36	424,36	848,72	37,203	1,41
22	19	133,40	133,40	266,79	11,695	0,97
23	131	143,18	143,18	286,35	12,552	2,18
24	24	357,14	357,14	714,28	31,310	0,87
25	865	360,77	360,77	721,54	31,628	2,16
26	170	854,84	854,84	1709,69	74,942	1,22
27	50	250,27	250,27	500,54	21,941	0,62
28	405	115,52	115,52	231,05	10,128	0,90
29	2135	219,19	219,19	438,38	19,216	0,47
30						
31						
32						
33–34		Monthly cost of storage [PLN]		6 998,77	306,8	Warehouse surface [m^2]

Fig. 6. Solution of the problem related to Economic Order Quantity (EOQ)

4. SUMMATION AND GENERAL CONCLUSIONS

Based on the above-described example, it can be concluded that the optimization used to solve managerial issues can bring tangible benefits such as increased productivity and savings in terms of costs, time, and material. Problems which are difficult or impossible to solve with the proverbial pen and paper can be solved in a short period of time using the MS EXCEL® spreadsheet and the Solver® add-in. This is possible thanks to efficient optimization algorithms and the high computing power of today's computers, which is more than sufficient in the case of most problems. Although the Solver® add-in only offers the possibility of single-criterion optimization, it is possible to find a compromise between two criteria by developing your model and conducting the iterative calculations in an appropriate way (vector optimization). Many problems and relationships can be easily described with numbers and mathematical functions, and intuitive work in a spreadsheet makes it easier to design a model. It is sometimes the case that the obtained solution, although theoretically optimal model-wise, does not seem to be rational and fit for real life. A solution which is worse (but closer to the optimum) may turn out to be much better, because it can be implemented in reality. Optimization by way of mathematical programming still requires some rational thinking, and although the real-life relationships can be described mathematically, it is necessary to distinguish those occurring in a spreadsheet from those occurring in reality [4].

REFERENCES

[1] Frontline Solvers http://www.solver.com/
[2] BAKER K.R., *Optimization Modeling with Spreadsheets*, 2nd ed., Wiley, 2011.
[3] BOROŃ J., *Optimizers in solving managerial and structural problems*, [in:] Information Systems Architecture and Technology: *Model Based Decisions*, J. Świątek et al. (Eds.), Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2008, 15–25.
[4] POWELL S.G., BAKER K.R., *Management Science: The Art of Modeling with Spreadsheets*, 4th ed., Wiley, 2013.

PART 3

IMAGE PROCESSING AND PATTERN RECOGNITION

Hilbert optics,
objects identification

Adam SUDOL*, Sławomir STEMPLEWSKI**, Viktor VLASENKO*

METHODS OF DIGITAL HILBERT OPTICS IN MODELLING OF DYNAMIC SCENE ANALYSIS PROCESS: AMPLITUDE-PHASE APPROACH TO THE PROCESSING AND IDENTIFICATION OBJECTS' PICTURES

Article presents the results of researches into possibilities of digital Hilbert optics (DHO) applications to amplitude-phase images of dynamic scenes objects. The effectiveness is evaluated with DHO-images correlations and its' run-rotation movement dependences. As the main method for identification the maxima-correlation analysis into amplitude-phase images of anisotropic Hilbert transformed complex shape objects is used. The identified object image is correlated with sets of run-rotated templates and maximum coefficient of correlation (MCC) decision is approved. The comparative analysis of common used MCC into identity transformed images and suggested method results shows that CHO-based methods of dynamic objects identification have more advantages and could be recommended for application into new generation of automatic systems for visualization and interpretation of dynamic scenes, and run-rotated objects identification.

1. INTRODUCTION

Identification of objects acting as dynamic scene (spatial-time) elements is very important question to the artificial intelligence and it is usually based on filtering and evaluation of parameters of objects' shape and trajectory. Increasing the sensitivity of the systems results in an improvement of discriminatory identify abilities and is practically justified. Methods of digital optics are developed in digital signal and image processing for the last 40–50 years [1–3 et al.], the methods based of Fourier transforms are investigated and developed most [4–7]. Application of digital Hilbert optics

---

* University of Opole, Department of Technology, Dmowskiego str. 7–9, 45-365Opole, e-mail: vlasenko@uni.opole.pl

** Technical University of Opole, Department of Electrotechnics, Automatics and Informatics, Prószkowska 76, 45-758 Opole.

(DHO) methods – digital Hilbert transforms for processing and analysis of multidimensional is not widely used, and its development is on experimental stage [2, 6–8 et al.].

Object recognition and dynamic scenes analysis based on digital Hilbert optics (DHO) requires investigations on discriminative ability of Hilbert analysis algorithms and evaluations of sensitivity to changes in shape and motion parameters of objects in scenes. Research in this area is carried out by the authors of this article for over ten years [10–12 et al.]. The basic methods of DHO are two-dimensional (2D-) isotropic (HTI) and anisotropic (HTA) Hilbert transformations. The methods of amplitude-phase analysis of images of objects, textures and complex scenes in the whole as creating so called “phase” images with their correlational analysis is the significant contribution to above mentioned set of method, notably efficient in very small changes in a series of analyzed pictures. Object identification is based on preliminary analysis and image normalization, processing of their Fourier transformations, computing the Hilbert and Foucault transform and finally on analysis of amplitude and phase fields. Evaluation of effectiveness of this approach is based on analysis of correlation matrices of investigated images.

The primary goal of this paper is to present and discuss the results of research of the properties and efficiency of amplitude-phase algorithms for analysis and identification the images of complex shape objects (CSO), as the elements of dynamic scene pictures distorted by three dimensional angular transformations. The new proposed approach to the problem is the application of modified algorithm for computation of amplitude-phase fields of investigated objects, based on S.L. Hahn’s method [8], and presentation of results of conducted experiments involving above mentioned algorithm.

This paper opens a series of studies showing the different aspects above mentioned methodology for amplitude-phase analysis and object identification. Their topics are presentation of structure-functional DHO block schemas, presentation of sample images of transform, investigated objects and results of analysis of correlational matrices of object in phase space of HTA and HTI transforms.

2. BASIC METHODS OF PHASE-AMPLITUDE IMAGES’ ANALYSIS IN HTA DOMAIN

Figure 1 contains a diagram of information technology concept (ITs) based on digital Hilbert optics (DHO) method of amplitude-phase analysis and proposed for application in analysis and identification of complex shape objects (CSO). The method implements summing of normalized Hilbert anisotropic transform (HTA) and Hilbert isotropic transform (HTI) and aggregation of phase – summary and differential – fields. The proposed method is the modified method of S.L. Hahn [9].

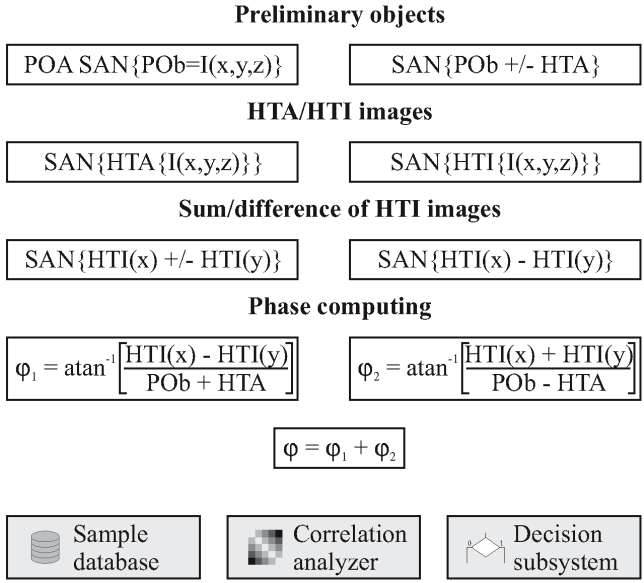


Fig. 1. General conception of Hahn's method modification:
POA – preliminary object acquisition, POb – preliminary object,
SAN – symmetrical amplitude normalization, HTI – Hilbert transform – isotropic,
HTA – Hilbert transform – anisotropic

Within the scope of this paper it is proposed to use a Hahn's modified method by calculating the sums of total and differential phase fields – based on HTA – for calculating the phase (result) images.

HTA moves a large amount of energy of preliminary image to so called "characteristic point of shape" (CPoS) thus strengthening the system's identification ability and increasing the sensitivity to changes of shape and motion parameters.

Correlation analysis of images' Hilbert transforms enables the classification of object by their type. However filtering and spatial tracking of object with known and constant shape and trajectory became more accurate utilizing the correlational analysis in the phase space. Thus depending on identification context it is necessary to choose proper type of transformation and analysis. Current research aims to determine and describe the set of transforms, their order and rules to follow to define most efficient identification subsystem in the arbitrary set conditions. The most distinct classification bases on the way the object can be seen by the recognition subsystem. At this time two ways are investigated: 2D with rotary distortion (like text) and 3D with fully 3D distortion capabilities (like planes).

a)

b)

c)

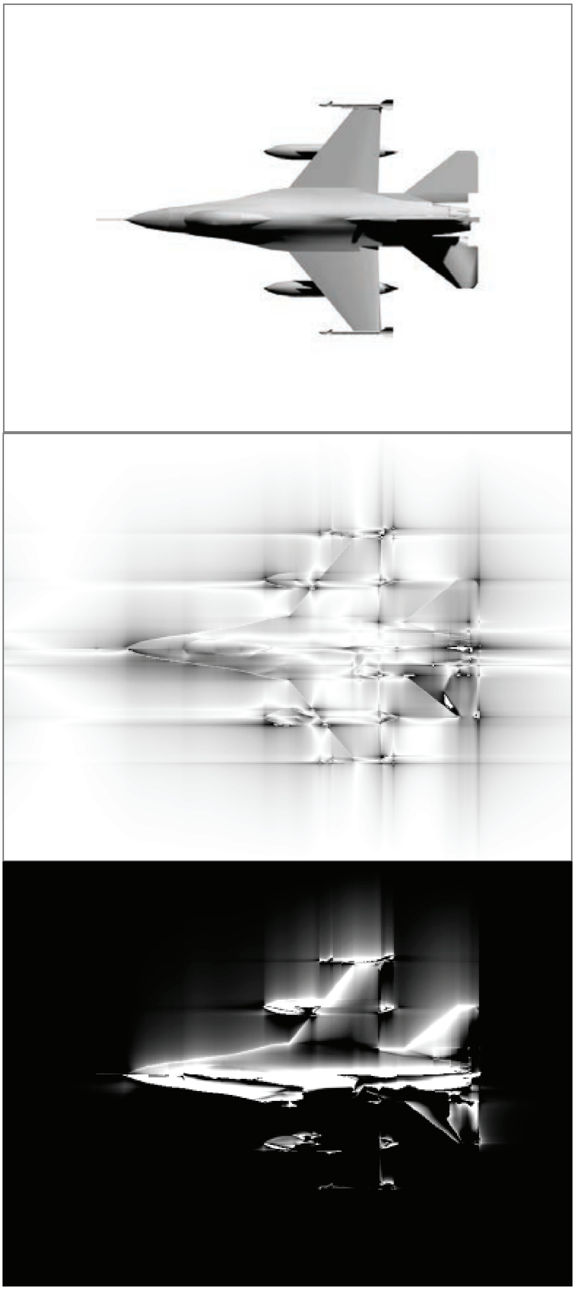
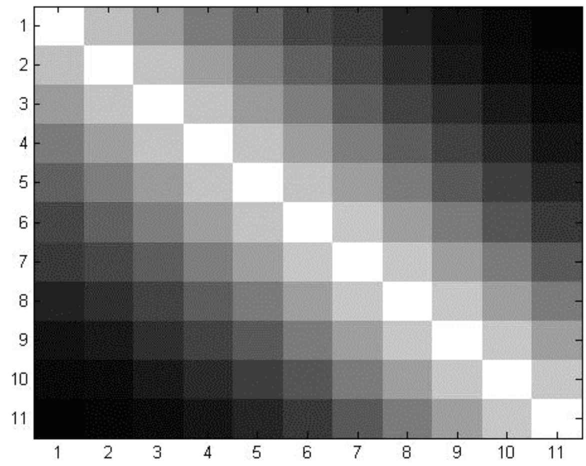


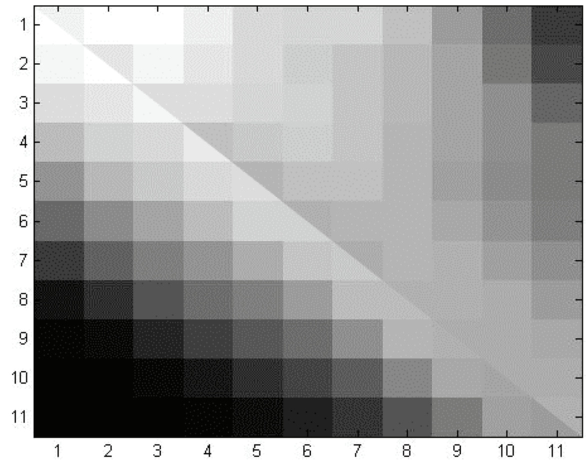
Fig. 2. Examples of phase images illustrate the basic steps of IT:
a) Flat image of 3D-object, b) 2D-HTA (symmetrical normalization of the transform),
c) the phase image

3. PRESENTATION OF THE RESULT OF UNDERLYING PROCEDURES FOR OBJECT IDENTIFICATION

The analysis of histograms indicates the possibility of separate classes (types) of objects processed with the methods discussed in this paper – inter-class identifications at the decision level threshold $B_{CC} \geq 0.75$ has been made.



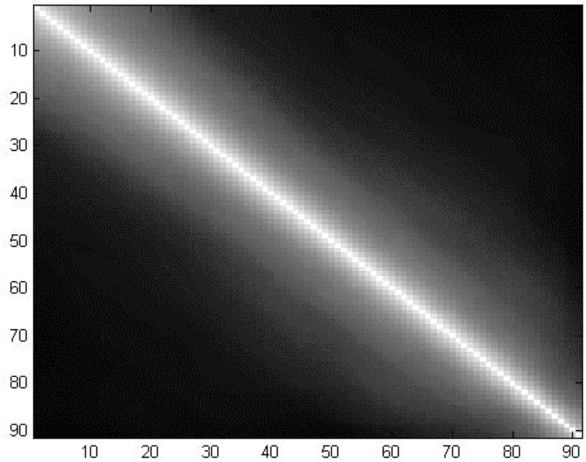
a_1) In-class correlation matrix – object 2325



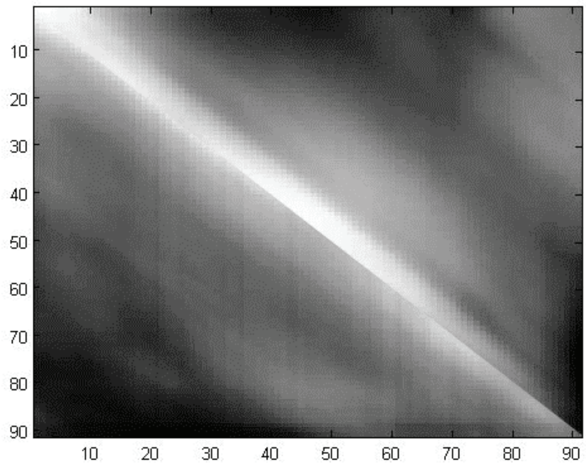
a_2) Inter-class correlation matrix – 2325 with 2364 – left half, 2325 with 2368 – right half

Fig. 3a. Examples of image correlation matrix of complex shape object on same and different types in different rotational positions in the range from 0 to 10° with a step of 1.0°

Application of phase methods increases the discriminative ability for identification on rotational movements of CSO and minimizes the influence of object shape differences on its effectiveness of assessment of motion vectors, which is evident by the histogram concentration close to each other in the low-mid range values of correlation coefficients ($B_{CC} < 0.5$) with significant discriminations of correct orientation of tested objects by the angular location patterns ($B_{CC} \geq 0.9$).



b$_1$) In-class correlation matrix – object 2325



b$_2$) Inter-class correlation matrix – 2325 with 2364 – left half,
2325 with 2368 – right half

Fig. 3b. Examples of image correlation matrix of complex shape object on same and different types in different rotational positions in the range from 0 to 90° with a step of 5.0°

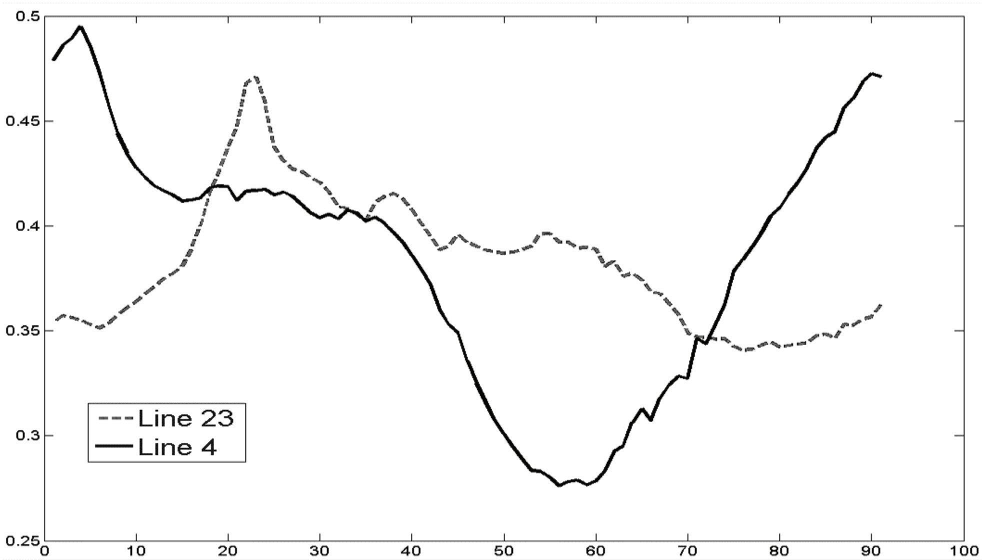
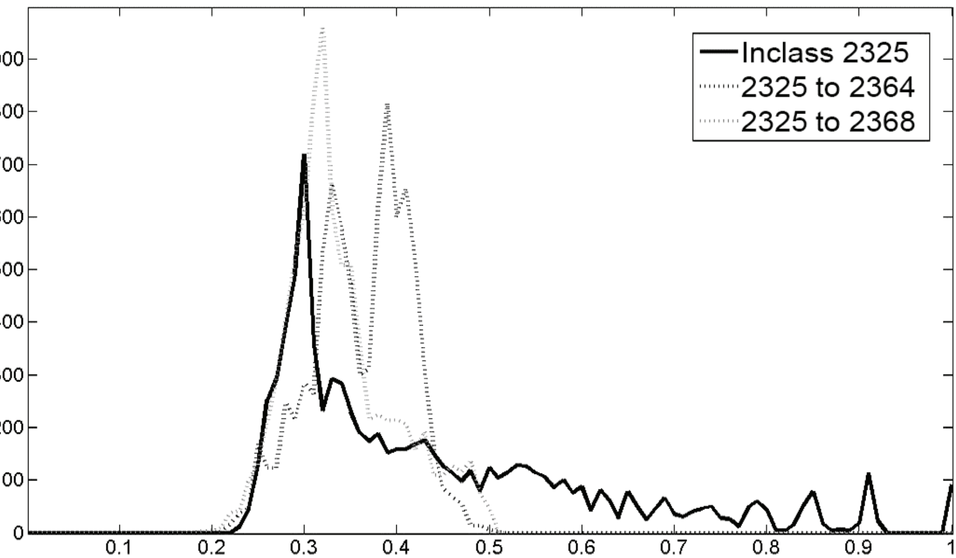
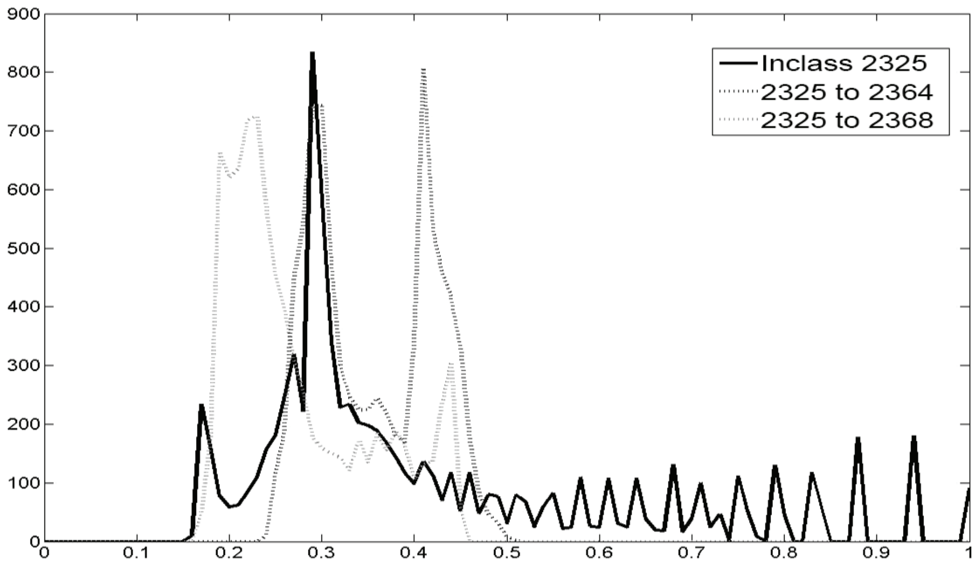


Fig. 4. Cross-lines of correlation matrixes (3b, b2)

The results of analysis and identification based on, presents fig. 5 as the histograms of correlation coefficients of objects in domains: a – preliminary object (non-processed), b – anisotropic transformation with symmetrical normalization (PH-HTA).



a)



b)

Fig. 5. Histograms of correlation coefficients objects' images:
a) non-processed, b) transformed in Hilbert amplitude-phase domain

4. ANALYSIS OF RESULTS AND CONCLUSIONS

Application of underlying IT procedures built on digital Hilbert optics in combination with nonlinear normalization methods, preparation and amplitude-phase analysis provides significant benefits in the process of identifying structural elements of dynamic scenes. Analysis of the results allows to draw the following *conclusions*:

– DHO methods exhibit the characteristic point, edges and the internal structure of CSO, enhancing the visualization of shape anomaly and weak objects;
– Analysis of amplitude-phase 2D- (3D-) images based on DHO procedures, especially on modified S. L. Hahn's method, allows to raise the resolution of identification (classification) systems based on the shape and motion parameters of CSO;
– Structural analysis allows for identification of Hilbert transforms of objects based on dependencies of shape parameters (structures) and motion thus minimizing the volume of etalon database.

5. PERSPECTIVES FOR FURTHER RESEARCH

The most interesting directions of research DHO-systems, in the opinion of the authors are:

– Analysis of effectiveness of algorithms for phase-difference methods of identification;
– Simplification of complex shape object (CSO) models in the domains of DHO by their signature- and histogram-goniometric signatures and assessment of the effectiveness of these techniques;
– The development of structural models based on characteristic points model CPM methods and images of phase fields in color pictures;
– Research the dependencies of higher order moments of Hilbert transforms and amplitude-phase pictures from objects' motion parameters.

REFERENCES

[1] ROZENFELD AZRIEL, *Picture Processing by Computer*, Academic Press, New York, London 1969.
[2] OPPENHEIM A.V., SCHAFER R.W., *Digital Signal Processing*. Prentice-Hall, Englewood-Cliffs, N.J., 1975.
[3] PRATT W.K., *Digital Image Processing*. Wiley – Interscience Publication, John Wiley and Suns, New York–Chichester–Brisbane–Toronto 1978.
[4] STARK H. (ed.), *Applications of Optical Fourier Transforms*, Academic Press, New York 1982.
[5] VAN DER LUGT A. (ed.), *Optical Signal Processing*, Wiley Series in Pure and Applied Optics, John Wiley, New York, NY, USA, 1992.
[6] PRATT W.K., *Digital Image Processing*, PIKS Inside, Third Edition, John Wiley & Sons, Inc., N.Y., 2001.
[7] ZIELIŃSKI T.P., *Cyfrowe przetwarzanie sygnałów. Od teorii do zastosowań*. WKŁ, Warszawa 2007.
[8] HAHN S.L., *Hilbert Transforms in Signal Processing*, Artech House, Boston–London 1996.
[9] VLASENKO V., SUDOŁ A., *DHO-Methodology for Complex Shape Objects and Textures at Dynamic Scenes Identification: Structure Design, Modeling and Verification*, Systems Science, 2009, Vol. 35, No. 3.
[10] VLASENKO V., VLASENKO N., SEMENOV D., SUDOL A., *Information technologies for dynamic objects and textures identification based on methods of digital Hilbert optics*, [in:] *Information Systems Architecture and Technology. Decision Making Models*., J. Świątek, L. Borzemski, A. Grzech, Z. Wilimowska (eds.), Wrocław 2007, 153–160.
[11] LAVRINENKO S.V., PIERAŃSKI W., VLASENKO N.K., VLASENKO V.A., *Zastosowanie metod cyfrowej optyki Hilberta w analizie oraz identyfikacji obiektów i tekstur w złożonych scenach dynamicznych*. IV Krajowa Konferencja „Metody i systemy komputerowe w badaniach naukowych i projektowaniu inżynierskim", Kraków 2003, s. 785–787.

[12] VLASENKO V.A., LAVRINENKO S.V., *Identification of Known-Form Objects Based on Anisotropic Two-Dimension Hilbert Filtering*. Proceedings of USRIRT, 2001, No. 1 (25), 67–74 (Journal of Ukrainian Scientific-Research Institute of Radio and Television).

data mining, medicine supporting, predictive systems,
classification, clustering, association rule mining

Adam BRZESKI*
Paweł ROŚCISZEWSKI*
Jan CYCHNERSKI*

DATA MINING APPLICATIONS AND METHODS IN MEDICINE

In this paper we describe the research area of data mining and its applications in medicine. The origins of data mining and its crucial features are shortly presented. We discuss the specificity of medicine as an application area for computer systems. Characteristic features of the medical data are investigated. Common problems in the area are also presented as well as the strengths and capabilities of the data mining methods. Finally, we describe and compare a set of modern applications of data mining designed for clinical use.

1. INTRODUCTION

Medicine has always been an exceptional field of application for computer science. Participating in patients treatment and life saving processes means responsibility for human health. Therefore, systems dedicated for supporting healthcare must meet numerous requirements and standards in order to assure their reliability and high level of confidence. At the same time, in recent years the contribution of computer science for healthcare has become a necessity. The capabilities of computers, information systems and electronic devices opens vast possibilities for saving people's health. Therefore, many computer systems assist medical personnel nowadays. As a results, constantly increasing amounts of medical data are collected and stored in the purpose of improving medical treatment. But large databases of medical data are also a great opportunity for medical researchers for discovering unknown knowledge and finding new relations between observed variables.

---

* Department of Computer Architecture, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, 11/12 Gabriela Narutowicza Street, 80-233 Gdańsk, Poland.

Along with the rapid growth of amounts of data to be analysed, a need for providing new tools has emerged in computer science. Traditionally, the problems of data analysis were solved by mathematical approach involving verifying hypothesis using statistical analysis. However, this approach is not efficient for large numbers of variables to analyse. Also, the data began to appear in more complex forms, like long time-series or multidimensional arrays. Consequently, it motivated the automation of the data analysis, which lead to establishing a new field of science referred to as data mining.

According to Guidici [12], data mining is the process of selection, exploration and modelling large quantities of data to discover regularities or relations that are at first unknown with the aim of obtaining clear and useful results for the owner of the database. In practise, data mining is a fusion of multiple concepts from mathematics, and statistics with computer science, utilizing various techniques related to large databases, visualisation of data, machine learning, classification and clustering. The term of data mining is closely related to another popular discipline – Knowledge Discovery in Databases (KDD). Data mining is often presented as one of steps of KDD processes, but the two terms also tend to be used exactly in the same context.

2. DATA MINING IN MEDICINE

Data mining in medicine is also specific. The nature of medicine implies not only increased requirements on the developed systems, but also brings additional difficulties of various kinds. Significant part of them are related to the nature of medical data. On the other hand, achievements of great value and importance are within the capabilities of data mining. It also has a reasonable set of features justifying its presence in the world of medicine.

Data mining applications are typically divided into two groups: predictive and descriptive data mining. Predictive data mining assumes selection of a special variable referred to as outcome variable. The remaining variables are called predictors. The aim is to build a model to predict the value of outcome variable basing on values of the predictors, which is typically a task of classification. Predictive data mining is the most common application of data mining in medicine. Descriptive data mining, on the other hand, aims to discover unknown relations between a set of variables and enable extracting clear rules about the regularities. While in prediction tasks the rules may remain hidden, descriptive data mining aims to discover certain knowledge about the analysed data and to provide clear explanations. Therefore it is a hard task to complete and successful applications are quite rare. Descriptive data mining is mostly related to the task of clustering and association.

Problematic aspects of data mining in medicine, firstly, relate to the sensitivity of data. Medical data need to be highly protected for numerous privacy-related reasons. Storing the data requires therefore an increased effort. On the other side, it results in limiting the access to the data for the researchers. Collecting medical data is also a highly expensive process, which hardens gathering of large amounts of data for particular research. For that reason medical data sets often contain small number of samples, which makes the analysis of the data more difficult. Furthermore, medical data is prone to many sources of noise and measurements errors, both in case of clinical observations and laboratory measurements. Also, clear and reasonable explanations of the system's decisions are expected in order to increase the credibility of the system. Data samples may include significant number of missing values, which need to be addressed by analysis techniques. Furthermore, data related to observed variables tend to be heterogeneous. Data may include high-level, descriptive clinical parameters as well as low level laboratory values or even genetic sequences. For some applications analysis of both types of data is required. Another problem is the dimensionality of data. Except single valued variables, data may come in a form of sequences of data, time-series, or even two and three dimensional arrays or images. Furthermore, as pointed out by Cios [6], medical data is hard to characterize and describe within computational models, and also the diversity in nomenclature and terminology makes establishing canonical form of the medical concepts nearly impossible. Finally, critical, human-health context of medicine requires particular attention to the reliability of data mining methods. Precise measurements of the quality of the algorithms need to be supplied. Methods should also provide the level confidence in terms of probability.

Mentioned problems make a successful data mining application a challenging task. However, as pointed out by Belazzi and Zupan [4], the data mining approach has essential capabilities entitling its application in medicine. Those include:

- systematic and integrated process – data mining is a pretty well established and standardized process with precisely defined steps. Even though the widely accepted CRISP-DM standard is not well suited for medical applications [20], alternative standards were proposed in the literature [5, 7, 10],
- explanations – many of data mining techniques are capable of providing explanations of the generated results,
- utility of domain knowledge – some of the data mining methods enable inclusion of the initial domain knowledge into the model, which is a substantial feature. Unfortunately, no standard approach was yet established in this area.

2.1. NOTABLE WORKS IN THE AREA

Despite the relatively short history of data mining in medicine, it is a dynamically developing field of science. Among numerous applications of the data mining techniques, several notable summary papers were also published. Below we shortly describe selected review papers.

Cios and Moore [6] presented a decent analysis of data mining in medicine, indicating its uniqueness among other fields of applications resulting from the specific nature of medicine. Authors characterize the specificity of medical data in the context of its gathering and analysis. Also, statistical difficulties are presented with examples of erroneous statistical inferences from medical data.

Belazzi and Zupan [4] provided a comprehensive review of the state of the art of predictive data mining in medicine. They describe most commonly used techniques that include decision trees, logistic regression, artificial neural networks, support vector machines, naive Bayes classifiers, Bayesian networks and *k*-nearest neighbours algorithm. Authors present the history of data mining applications in medicine with mentioning examples of published works. They present a set guidelines and tasks related to data mining on medical data, drawing attention to important issues that need to be considered, which are: defining the problem and setting the goals, data preparation, modelling and evaluation, construction of the target predictive model, deployment and dissemination.

Yoo et al. [23] presented data mining techniques in 3 categories: classification, clustering and association and provided respective guidelines regarding their practical use. Authors described many of techniques in more detail, namely: naive Bayes classifier, neural network, decision tree, support vector machine, classification based on association, ensemble, AdaBoost, hierarchical clustering algorithms, partitional clustering algorithms and Apriori algorithm. Furthermore, 16 scientific works presenting applications of data mining in medicine are referenced with outlining the particular techniques used by the authors. Also three examples of data mining application in medical industry were presented.

In this paper we attempt to describe and compare more cases of applications of data mining in medicine. We also focus on the most recent works which were not covered in the review papers presented above. We focus on clinical data rather than low level genetic or protein data. However, we cover both predictive and descriptive data mining applications and methods.

3. APPLICATIONS AND METHODS

For the presented review we selected a set of 10 papers presenting applications of data mining in medical domain. A detailed comparison of the methods was presented in Tables 1 and 2. Below we present a short description of selected works.

Firstly, Stel et al. [21] managed to establish medical clues for the risk of recurrent falling in older persons. Authors utilized tree-structured survival analysis

(TSSA) [11]. The TSSA model is mostly used in medical research and is an alternative approach to popular Cox hazards model [8]. It is based on binary classification tree. The tree is constructed by inserting chosen predictors as consecutive tree nodes. Each node represents a division of the patients population into two groups. The selection of predictors is based on the discrimination power of the predictors, which is measured as a difference in prognosis according to log rank. The nodes are inserted into the tree until a stop condition is met, e.g. when further partition is statistically not significant or when the terminal node covers a small group of cases, that is 5–10. Authors of the work, however, constructed a full tree and then applied a pruning algorithm [17] in order to remove uninformative nodes. Also, authors used Kaplan–Meier survival curves before the log rank test. The procedure enabled identification of 11 groups of patients characterised by a particular set of up to 6 predictors that had high correlation to the outcome variable (the risk of becoming a recurrent faller). TSSA was also used by Lopes et al. [15] in order to predict mortality in patients with hospital acquired infection, which lead to indicating significant factors correlated with the outcome variable. A similar decision-tree based approach was used by Liu et al [14]. They used a classification and regression tree (CART) to find predictors for successful return to spontaneous breathing after mechanical ventilation.

Yeh et al. [22] presented a notable approach of considering time series (laboratory data collected over time) together with domain knowledge acquired from the physicians using temporal abstraction [19]. Authors also present their procedure for handling missing and outlier values.

Amini et al. [1] presented stroke prediction method using a C4.5 decision tree and *k*-nearest neighbours. The research investigated 50 risk factors of stroke evaluated for 807 stroke and normal patients. The prediction based on the risk factors achieved efficiency of 90%. Authors however did not indicate which predictors were of highest significance.

Andrews et al. [3] used logistic regression and decision trees to predict the outcome of head-injured patients. After establishing the prediction model, the authors managed to identify the most meaningful clinical attributes, that strongly supported the predictive accuracy (PA). The authors precisely measured the influence of selected attributes, proving their correlation with the outcome variable.

Finally, Eastwood et al. [9] analysed data of older patients suffering from hip fracture and investigated factors determining treatment outcome in 6 months. By using cluster analysis authors divided the patients into clusters and found correlations between them and the outcome variable. Finally, the authors constructed a simple decision tree enabling classification of the patients by the most significant predictors and assignment to a proper cluster.

Table 1. General information about the analyzed data in the selected works

Ref	Data – rows	Data – columns (attributes, predictors)	Outcome variable
[9]	571 hospitalized adults aged 50 and older with hip fracture	29 numerical characteristics including information on age, gender, ethnicity, prefacture residence, presence of paid help in the home, cognitive impairment	outcome in 6 months (functioning, nursing home residence, death)
[3]	124 head injured patients	Clinical and laboratory data (patient age, admission Glasgow Coma Scale score, Injury Severity Score (ISS), pupillary response on admission, and insults durations, low cerebral perfusion pressure (CPP))	death
[21]	1365 community-dwelling older persons (Z65) from the Longitudinal Aging Study Amsterdam (LASA)	34 predictors of falls; numerical values established by surveys, divided into 8 categories: sociodemographic characteristics, chronic diseases and medication use, physical impairments and general health, body composition, activity and mobility, psychosocial functioning, lifestyle, and other potentially fall-related predictors	recurrent falling
[15]	754 patients with hospital-acquired infections	33 variables (number of admissions, age, gender, nutritional status on admission , underlying diseases invasive procedures, number of antibiotics used, time from admission to onset of the first NI)	death
[2]	909 records	13 attributes (personal data, clinical view, laboratory data)	heart disease
[14]	113 elderly patients in the medical intensive care unit on ventilation for >48 hours	16 demographic data values and 20 values from respiratory data	ability to spontaneous breathing
[13]	125 cases (55 IHD, 70 normal)	1,152 descriptors extracted from 36 time series from a Magnetocardiogram (MCG)	Ischemic Heart Disease
[22]	8223 samples from chronic hemodialysis (HD) patients with end stage renal disease	Basic attributes (gender, age, hospitalization, admission date) and time series of 26 laboratory	need for hospitalization after HD treatment
[18]	200 patients (124 brain tumor, 76 normal)	13 laboratory parameters with qualitative values (high, normal or low)	brain tumor
[1]	807 healthy and sick hospital patients	50 stroke risk factors for stroke (such as history of cardiovascular disease, diabetes, hyperlipidemia, smoking and alcohol consumption)	stroke

Table 2. Data mining methods and result achieved in the selected works

Ref	Methods	Results
[9]	– cluster analysis – decision tree	90% classification accuracy in assigning proper cluster
[3]	– logistic regression – decision trees	hypotension and low cerebral perfusion pressure (CPP) were found to have 9.2% improvement in Predictive Accuracy for predicting death. Low CPP, patient age, hypocarbia, and pupillary response had 5.1% improvement in PA for predicting good outcome.

[21]	– tree-structured survival analysis (TSSA) [11] with tree pruning [17] – Kaplan–Meier	11 end groups of patients with given values of 6 selected predictors associated correlated with the outcome variable
[15]	– TSSA – Kaplan–Meier	– invasive procedures and use of two or more antibiotics predictors indicated as significant – two groups of patients with high mortality rates were identified
[2]	– genetic search for attributed reduction – *k*-fold cross validation – decision tree-structured – naive bayes – classification by clustering	6 significant attributes were identified; 99% classification accuracy using the decision tree
[14]	– decision tree (CART analysis) – Gini index – 5-fold cross-validation	– 5 end groups with given values of selected predictors associated, correlated with the outcome variable – 3 significant variables were identified
[13]	– neural network – Bayesian neural network – probabilistic neural network – support vector machine – 10-fold cross-validation	Classification sensitivity of 96.65% using neural network
[22]	– temporal abstraction (TA) [19] – MSApriori algorithm – C4.5 decision tree	Two alternative, meaningful sets of rules for establishing probable outcome; the most significant predictor was identified
[18]	– normalized regression – jackknifing, cross-validation – R^2 analysis – *Y*-randomization	8 rules with established support, confidence and correlation rates; identification of 4 significant predictors
[1]	– *K*-nearest neighbours – C4.5 decision tree	Sensitivity: 97.21%, Specificity: 85,6%, Precision: 97,36%

4. DISCUSSION

As we attempted to show in previous sections, the field of data mining in medicine is a popular direction of research. More and more data analysis techniques are being applied for exploration of medical data. At the same time increasingly large datasets are being collected for analysis, posing a valuable source for discovering medical knowledge. Reliable datasets are also often published online (for instance: Lung Cancer Dataset, Breast Cancer Wisconsin Dataset), which is extremely valuable and helpful for further research in the area. These aspects can be considered a proper direction of the research development in the area. However, still significant problems exist, which were also mentioned in section 2. In the first place it seems that the significance and the contribution of data mining for medicine is still rather low. We suspect that the reason lies in widely understood lack of trust and confidence in the computer analysis, which is most often carried out beyond human control and understanding. The complexity of the computations pose a risk of containing hidden errors. At the same time the data analysis is hard to test and verify, which is therefore rarely accomplished in a satisfying degree. Another serious problem is related to the correctness of application of the statistical methods. Further errors can be made in this kind of analysis, which can be hard to detect, and are also present in scientific papers [16]. This also causes a certain degree of skepticism about the results reported by the scientists. Therefore, we consider focusing on increasing credibility and clarity of the methods as the most important direction for further development in the area.

5. CONCLUSIONS

Data mining is a relatively new field of science, which emerged to address the problem of analysis of large amounts of data. Medicine is a specific field of application for data mining methods. Even though, numerous problems need to be overcome when considering medical data, modern data analysis techniques have capabilities and potential for successful clinical applications. Many attempts on constructing medicine supporting data analysis systems were presented in literature. Most of them concern classification of data samples in order to predict the value of the outcome variable. However, successful works on descriptive, explanatory systems were also accomplished, which resulted in discovering relations and rules hidden in the data. The contribution of data mining for medicine is, however, still rather small. Credibility of the systems is one of the most important aspects to improve for their wider application in clinical use.

ACKNOWLEDGEMENTS

This research was partially funded by grants from National Centre for Research and Development (PBS2/A3/17/2013, *Internet platform for data integration and collaboration of medical research teams for the stroke treatment centers*).

REFERENCES

[1] AMINI L., AZARPAZHOUH R., FARZADFAR M.T., MOUSAVI S.A., JAZAIERI F., KHORVASH F., NOROUZI R., TOGHIANFAR N., *Prediction and control of stroke by data mining*, International Journal of Preventive Medicine, 2013, Vol. 4, 245–249.

[2] ANBARASI M., ANUPRIYA E., IYENGAR N.C.S.N., *Enhanced prediction of heart disease with feature subset selection using genetic algorithm*, International Journal of Engineering Science and Technology, 2010, Vol. 2, No. 10, 5370–5376.

[3] ANDREWS P., SLEEMAN D., STATHAM P., MCQUATT A., CORRUBLE V., JONES P., HOWELLS T., MACMILLAN C., *Predicting recovery in patients suffering from traumatic brain injury by using admission variables and physiological data: a comparison between decision tree analysis and logistic regression*, Journal of Neurosurgery, 2002, Vol. 97 No. 2, 326–236.
[4] BELLAZZI R., ZUPAN B., *Predictive data mining in clinical medicine: Current issues and guidelines*, International Journal of Medical Informatics, 2008, Vol. 77, No. 2, 81–97.
[5] CIOS K.J., MOORE G.W., *Medical data mining and knowledge discovery: an overview*, Medical data mining and knowledge discovery, Springer, Heidelberg 2000, 1–16.
[6] CIOS K.J., MOORE G.W., *Uniqueness of medical data mining*, Artificial Intelligence in Medicine, 2002, Vol. 26, No. 1–2, 1–24.
[7] CIOS K.J., TERESINSKA A., KONIECZNA S., POTOCKA J., SHARMA S., *Diagnosing myocardial perfusion SPECT bull's-eye maps – a knowledge discovery approach*, IEEE Eng. Med. Biol., 2000, Vol. 19, No. 4, 17–25.
[8] COX D., *Regression Models and Life-Tables*, Breakthroughs in Statistics, 1992, 527–541.
[9] EASTWOOD E.A., MAGAZINER J., WANG J., SILBERZWEIG S.B., HANNAN E.L., STRAUSS E., SIU A.L., *Patients with hip fracture: subgroups and their outcomes*, Journal of the American Geriatrics Society, 2002, Vol. 50, No. 7, 1240–1249.
[10] FAYYAD U.M., PIATETSKY-SHAPIRO G., SMYTH P., *Knowledge discovery and data mining: towards a unifying framework*, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996, Vol. 96, 82–88.
[11] GORDON L., OLSHEN R., *Tree-structured survival analysis*, Cancer Treatment Reports, 1985, Vol. 69, No. 10, 1065–1069.
[12] GIUDICI P., *Applied data mining: statistical methods for business and industry*, John Wiley & Sons, 2003, p. 2.
[13] KANGWANARIYAKUL Y., NAENNA T., NANTASENAMAT C., TANTIMONGCOLWAT T., *Data mining of magnetocardiograms for prediction of ischemic heart disease*, EXCLI Journal, 2010, Vol. 9, 82–95.
[14] LIU Y., WEI L.Q., LI G.Q., LV F.Y., WANG H., ZHANG Y.H., CAO W.L., *A decision-tree model for predicting extubation outcome in elderly patients after a successful spontaneous breathing trial*, Anesthesia & Analgesia, 2010, Vol. 111, No. 5, 1211–1218.
[15] LOPES J., GOULART E., SIQUEIRA A., FONSECA I., DE BRITO M., STARLING E., *Nosocomial infections in brazilian pediatric patients: using a decision tree to identify high mortality groups*, Brazilian Journal of Infectious Diseases, 2009, Vol. 13, No. 2, 111–117.
[16] REINHART A., *Statistics done wrong*, Web. 4 August 2014.
[17] SEGAL M., *Regression trees for censored data*, Biometrics 44, 1988, 35-47.
[18] SENGUPTA D., SOOD M., VIJAYVARGIA P., HOTA S., NAIK P.K., *Association rule mining based study for identification of clinical parameters akin to occurrence of brain tumor*, Bioinformation, 2013, Vol. 9, No. 11, 555–559.
[19] SHAHAR Y., *A framework for knowledge-based temporal abstraction*, Artificial intelligence, 1997, Vol. 90, No. 1, 79–133.
[20] SHILLABEER A., RODDICK J.F., *Establishing a lineage for medical knowledge discovery*, Proceedings of the sixth Australasian conference on data mining and analytics, 2007, Vol. 70, 29–37.
[21] STEL V., PLUIJM S., DEEG D., SMIT J., BOUTER L., LIPS P., *A classification tree for predicting recurrent falling in community-dwelling older persons*, Journal of the American Geriatrics Society, 2003, Vol. 51, No. 10, 1356–1364.

[22] YEH J.Y., WU T.H., TSAO C.W., *Using data mining techniques to predict hospitalization of hemodialysis patients*, Decision Support Systems, 2011, Vol. 50, No. 2, 439–448.

[23] YOO I., ALAFAIREET P., MARINOV M., PENA-HERNANDEZ K., GOPIDI R., CHANG J., HUA L., *Data mining in healthcare and biomedicine: a survey of the literature*, Journal of Medical Systems, 2011, Vol. 36, No. 4, 2431–2448.

endoscopy, disease detection, image and video analysis, gastropathy, portal hypertensive, edge detection, Local Binary Pattern, local maxima clustering

Jan CYCHNERSKI*, Piotr DOROŻYŃSKI*,
Tomasz DZIUBICH*

AN ALGORITHM FOR PORTAL HYPERTENSIVE GASTROPATHY RECOGNITION ON THE ENDOSCOPIC RECORDINGS

Symptoms recognition of portal hypertensive gastropathy (PHG) can be done by analysing endoscopic recordings, but manual analysis done by physician may take a long time. This increases probability of missing some symptoms and automated methods may be applied to prevent that. In this paper a novel hybrid algorithm for recognition of early stage of portal hypertensive gastropathy is proposed. First image preprocessing is described. Then disease symptoms characteristics are presented and hybrid algorithm scheme combining edge detection, Local Binary Patterns and local maxima clustering is shown. Finally the detailed description of these methods are provided. The parameters of the algorithm are also described with ranges used in tests and their best values (obtained empirically) are presented. The proposed algorithm is tested and compared to a few other algorithms showing it's comparable in terms of effectiveness in general case and a bit better than other ones in recognition of early stage of PHG.

1. INTRODUCTION

Actual state of medicine, diagnostics and medical imaging lead to abrupt growth of diagnostic data available to physician. Amount of this data is often so big that in many cases physician are not able to analyze and interpret it acceptably accurately and quickly. In that case it is necessary to assist them in this very important task in diagnosis process.

Wireless Capsule Endoscopy (WCE) is an example of such need. In digestive tract diagnostics this non-invasive method of imaging becomes more and more popular nowadays. WCE examination begins with swallowing an electronic capsule. While the capsule moves naturally through whole digestive tract, it makes a dozen thousand

* Department of Computer Architecture, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland, e-mail: jan.cychnerski@eti.pg.gda.pl, piotr.doro.dorozynski@gmail.com, dziubich@eti.pg.gda.pl

digital photos/video frames of inside of selected part of patient's digestive tract. In the future, number of these images may grow up to a million. Review of such recording may take a very big amount of time – for a typical 8-hour long recording a physician needs typically 1–2 hours of work to analyze it carefully.

Computer diagnosis assistance in WCE is therefore essential for medical doctors, which resulted in many algorithms created for this task. Most of these algorithms are based on some advanced mathematical operations (e.g. wavelet transforms [1], Local Binary Patterns [2] and many more). Because of their complexity, it is usually extremely difficult or even impossible to explain their actions physically. This fact makes the algorithms unpredictable in many applications, which strongly reduces their potential use in real medicine, where proves of algorithm's correctness are needed.

Therefore, in this article an algorithm fulfilling this requirement (need of explanation why such methods used by the algorithm have been used) is proposed. It is especially designed for detecting one of the popular gastroenterologic diseases: portal hypertensive gastropathy, PHG, as authors couldn't find in literature any algorithm dedicated for this disease. Construction of the algorithm is completely justified by physical and visual features of the disease, which makes it distinguishable from most other algorithms designed for universal disease recognition.

2. ALGORITHM

2.1. IMAGE PREPROCESSING

Due to the fact, that acquired medical data comes from several sources (including various endoscopic devices), image preprocessing is necessary in order to standardise and improve the quality of the images. Depending on the endoscope, the true picture may occupy from 50% [3] to 100% of the image, moreover, the image may contain a variety of control information (date, time, name of the patient, technical information, etc.). All these elements interfere with recognition algorithms – so it was necessary to remove them. For this purpose each image was preprocessed as in [4], shown in Fig. 1.

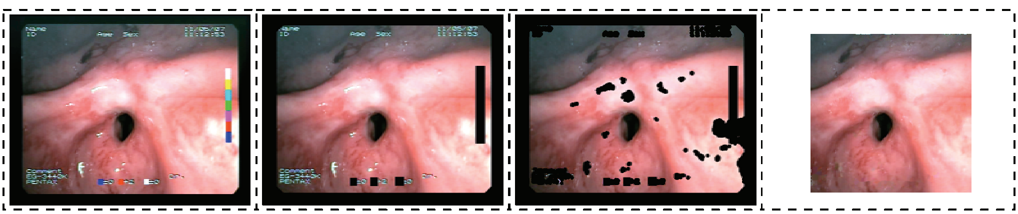


Fig. 1. Image preprocessing procedure

2.2. THE CORE ALGORITHM

Portal hypertensive gastropathy is caused by difficulties in blood outflow from stomach veins and increased artery pressure in patients with portal hypertension [5], which appears as characteristic mosaic often called “snake-skin” as shown in Fig. 2.

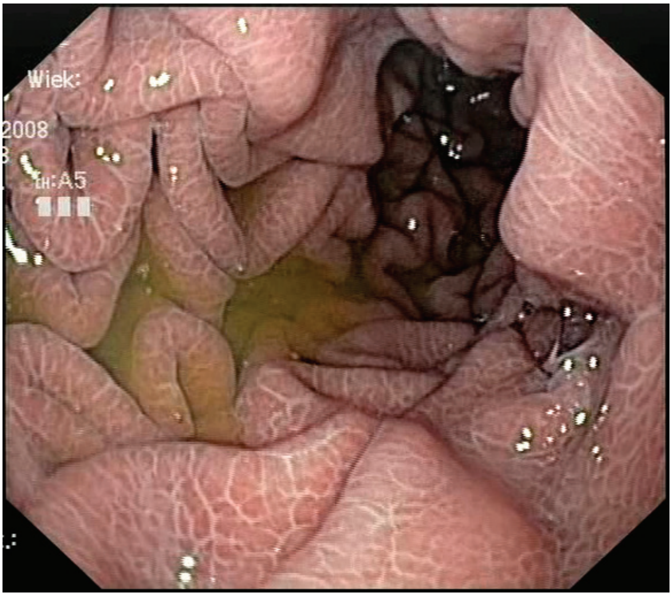


Fig. 2. Example of “snake-skin” texture in early stage of PHG

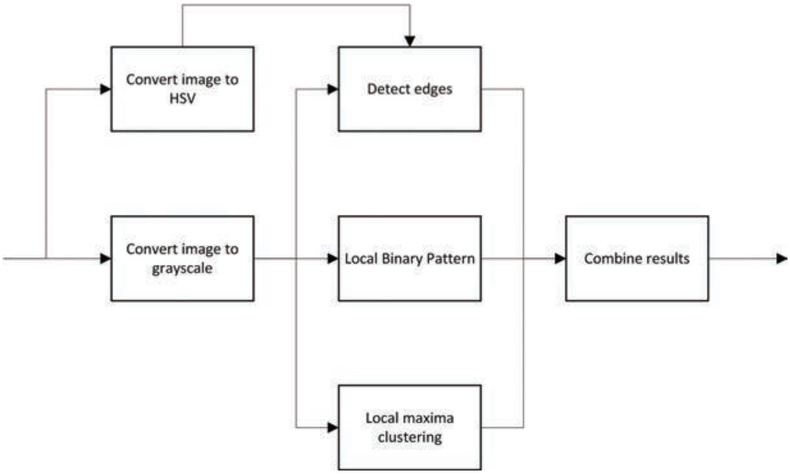


Fig. 3. Core algorithm scheme

In later stages of disease appears redness and finally bleeding. The proposed algorithm is designed for recognition of first stage of PHG ("snake-skin") and is hybrid combining 3 methods: edge detection, Local Binary Pattern (LBP) and local maxima clustering. The algorithm scheme is shown in Fig. 3. For general texture recognition the $LBP_{P,R}^{riu2}$ operator [6] is used on grayscale image. Other transformations are described in follow up sections. Described algorithm creates feature vector which is then used as input to two classifiers: artificial neural network (ANN) and support vector machine (SVM).

2.3 EDGE DETECTION

Since "snake-skin" texture has a lot of edges created by visible veins and arteries, edge detection was applied to grayscale image and resulting number of pixels corresponding to edges divided by total number of pixels in image was added to feature vector. To eliminate detected edges caused by natural tissue corrugation edge detection is done twice with different parameters: once to find only highly visible edges caused by natural tissue corrugation and once to find all edges. Algorithm described in [7] (Canny edge detector) was used for edge detection. The edges caused by natural tissue corrugation are then masked to get only edges caused by PHG. Finally to remove detected edges caused by noises like light reflections only pixels that in HSV space have S value above threshold and V value in specified range are kept. The whole process is shown in Fig. 4.

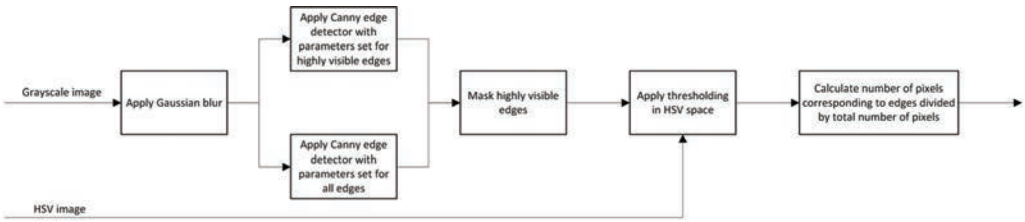


Fig. 4. Edge detection scheme

2.4. LOCAL MAXIMA CLUSTERING

Highly visible veins and arteries in early stages of PHG characterizes in higher intensity values than neighbouring pixels and thus finding local maxima is proposed. In classical local maxima search the pixels are only compared with neighbours and if they're value is higher or equal than all neighbours they are considered as maximum. For endoscopy images this gives a lot of maxima and thus some noise tolerance is needed. To achieve it, the continuous neighbourhood in which all pixel intensities are higher or equal to g_c – *noiseTolerance* is created for each pixel, where g_c is intensity value of a pixel for which the neighbourhood is created. The pixel is considered as maximum only if none of the pixels in this neighbourhood has higher intensity value than this pixel. Finally to eliminate maxima caused by noises in image (e.g. light reflections), the maxima are clustered by dividing the image into fixed size blocks and number of maxima in each block is considered. All blocks in which there is at least one maximum and number of maxima is smaller than predefined threshold are counted and number of this blocks divided by total number of blocks is added to feature vector.

2.5. ALGORITHM PARAMETERS

The proposed algorithm has few parameters on which results may vary. Their description, range used in experiments and the value which gives the best result (obtained empirically) is presented in table 1.

Table 1. Algorithm parameters

Name	Description	Range	Best value
lowThreshold1	Lower threshold in hysteresis thresholding in Canny edge detector	0–100	19
highThresohold1	Higher threshold in hysteresis thresholding in Canny edge detector	100–200	152
lowThreshold2	Lower threshold in hysteresis thresholding in Canny edge detector	0–100	34
highThresohold2	Higher threshold in hysteresis thresholding in Canny edge detector	100–200	167
kernelSize	Kernel size in Gaussian blur	1,3,5	3
Delta	Kernel's standard deviation	0–100	61
SThreshold	Threshold for S channel	0–150	21
VThresholdLow	Lower bound of V range	0–150	61
VThresholdHigh	Higher bound of V range	150–255	232
P	Number of points in LBP	8–28	16
R	Radius of points in LBP	1–3	3
invert	If true invert intensity value before finding maxima. Changes algorithm from finding maxima to finding minima	True, False	True
noiseTolerance	Noise tolerance for finding maxima	5–30	8
blockSize	Size of block in percent of image size	2,4,5,10	10
blockThreshold	Threshold for block expressed in fraction of number of pixels in block	0.005–0.02	0.0174

3. EXPERIMENT

3.1. TEST DATA

As the input data for training, testing and evaluation of the algorithms, the database of endoscopic examinations was used. This database contained several dozen real-life endoscopic recordings. Each recording was completely, frame by frame marked by experts in terms of their content – every frame received a set of labels defining its content. Labels included among others:

- blurry frame,
- sharp, clear frame,
  - healthy tissue,
  - abnormal tissue,
    - gastropathy 1 (general PHG),
    - gastropathy 2 (first stage of PHG),
    - cancer.

Each recording had a patient assigned. Videos of one patient are always considered jointly in the training and testing procedure, as so examples derived from one patient are not separated into different cross-validation sets. Contents of the database are described in table 2.

Table 2. Input data sets

Test case	Total samples	Positive samples	Positive exams	Negative samples	Negative exams
Blur	9405	4541	158	4864	164
Cancer	2758	1477	49	1281	47
Gastropathy 1	2777	1296	44	759	52
Gastropathy 2	1662	181	6	759	52

3.2. PARAMETER ASSIGNMENT

The algorithm has a set of input parameters, therefore, before the final assessment of efficacy, their specific optimal values must be obtained. For this reason, the behaviour of the algorithm was tested for different values of the input parameters. To do this, many sets of parameters and their values were selected, and then for each parameter set, a full 4-fold cross-validation training-testing procedure was carried out.

The whole data was assigned to the cross-validation subsets with algorithm described in [8], in a way that balanced number of disease examples over all subsets, without splitting one patient's images into different subsets.

Due to the fact that this process is extremely time-consuming and computationally demanding, from each video up to 15 frames were selected for each class (blurred, sharp, abnormal, etc.).

To improve the evaluation process, additional test was conducted on the smoothness of the resulting classification function of the entire videos in accordance with [9]. This process involved performing full classification of entire videos and evaluation of the resulting function changing speed. Due to the fact that most neighbouring frames in the typical video are very similar, the resulting classification function also should return similar results. If the resulting function behaves chaotically, it probably means the overtraining the classifier.

The classification efficiency score is defined by the formulas:

$$P_\alpha(a,b) = \min(a,b) + \alpha|a-b| \tag{1}$$

$$F_\beta(a,b) = (1+\beta^2)\frac{ab}{a+\beta^2 b} \tag{2}$$

$$S_{\mathrm{CV}} = P_{0.1}\,(\text{sensitivity, specificity}) \tag{3}$$

$$S_{\mathrm{PA}} = F_{0.2}\,(S_{\mathrm{smoothness}}, S_{\mathrm{CV-4}}) \tag{4}$$

where:

$\alpha \in [0, 1]$ is a ratio of preferring max(*a*, *b*) over min(*a*, *b*),
$\beta \in [0, \infty)$ is a ratio of preferring *a* over *b*,
$S_{\mathrm{CV}\text{-}k}$ is the total score for *k*-fold cross-validation,
$S_{\mathrm{smoothness}}$ is a smoothness test score like in [9],
S_{PA} is the final Parameter Assignment score.

3.3. TRAINING AND TESTING

After determining the specific values of the parameters of the algorithm, more precise tests were performed. For this purpose, only classifier parameters were tuned, at constant values of the core algorithm's parameters.

The test was performed by 16-fold cross-validation. Due to the fact that the videos had very diverse length (from a hundred to a few thousand frames) – some of the recordings would be largely over-represented in the learning process. For this sake, a maximum of 30 frames of each class from every video were selected.

Like the parameter assignment process, this evaluation process also contained output function smoothness testing.

The exact final classification efficiency score is defined by the formula:

$$S_{\mathrm{TT}} = F_{0.1}\,(S_{\mathrm{smoothness}}, S_{\mathrm{CV-16}}) \tag{5}$$

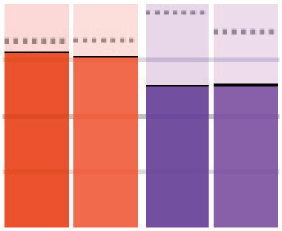
In order to evaluate the proposed algorithm, it was compared to a number of other gastrointestinal diseases recognition algorithms known from the literature:

A. Poh-Chee-Khun Color [10],
B. Poh-Chee-Khun Texture [10],
C. Kodogiannis [11],
D. BaopuLi [12],
E. Proposed algorithm.

4. RESULTS

4.1. PARAMETER ASSIGNMENT

The results of parameter assignment experiment for the best parameter set are shown in Fig. 5. The height of the bars represent achieved score, solid horizontal lines represent sensitivity and specificity, and dotted line represents smoothness score. The result isn't as good as expected (for gastropathy), but the experiment was done on all samples labelled as gastropathy, even with symptoms from later stages of PHG. On the other hand the algorithm achieved surprisingly good result in recognizing blurred images.



Score	78	76	64	64
Sensitivity	78	76	63	63
Specificity	78	76	63	63
Smoothness	82	83	95	86
Classifier	ANN	SVM	ANN	SVM
Test case	Blur		Gastropathy	
Signature	E			

Fig. 5. Parameter assignment results

4.2. TRAINING AND TESTING

The results of training and testing for gastropathy are shown in Fig. 6. The result of proposed algorithm compared to other tested algorithms for recognition of general PHG are similar, especially for ANN classifier. The situation is a bit different for recognition of first stage of PHG in which the proposed algorithm is about 10 percentage points better compared to other algorithms.

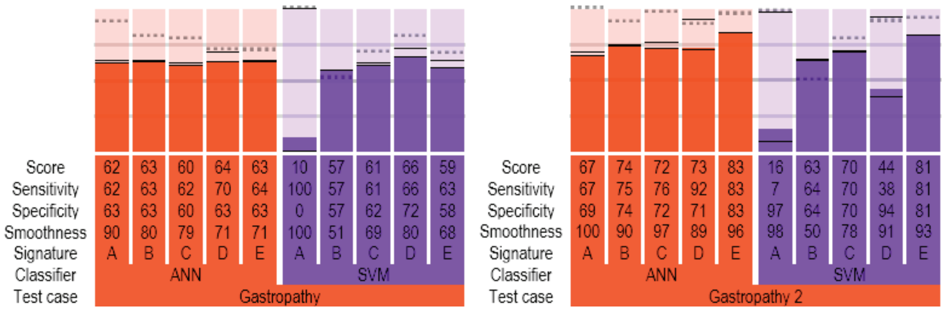


Fig. 6. Results of test *gastropathy 1* (on the left) and *gastropathy 2* (on the right)

Good results achieved for blurred images in parameter assignment experiment encouraged to test proposed algorithm also in other fields of image classification thus training and testing was done for blurred images and cancer recognition and the results are shown in Fig. 7. In both cases the proposed algorithm achieved comparable results to other tested algorithms (in fact being one of the best) and achieved final score at around 75%.

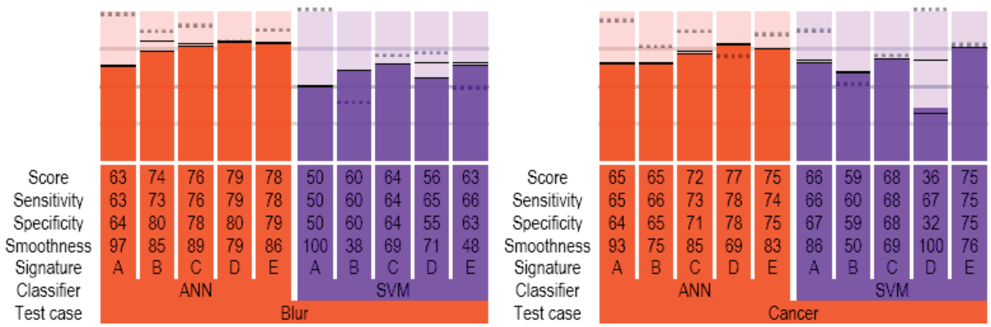


Fig. 7. Results of test *blur* (on the left) and *cancer* (on the right)

5. CONCLUSIONS

In this paper a novel hybrid algorithm for portal hypertensive gastropathy recognition has been proposed. The algorithm combines edge detection with elimination of edges caused by natural tissue corrugation and colour analysis, Local Binary Patterns and local maxima clustering with noise tolerance and elimination of too dense clusters. The algorithm effectiveness is comparable to other tested algorithms on the same input samples in general case and slightly better (about 10 percentage points) for early stage of PHG.

The proposed algorithm doesn't address later stages of PHG i.e. redness and bleeding. For the latter one blood detection algorithms like in [13] or [14] can be used. For the former colour analysis may be used, but no algorithm has been developed yet.

ACKNOWLEDGEMENTS

This research was funded by grants from National Centre for Research and Development (PBS2/A3/17/2013, *Internet platform for data integration and collaboration of medical research teams for the stroke treatment centers*).

REFERENCES

[1] BARBOSA, D.J.C., RAMOS, J., LIMA C.S. *Detection of small bowel tumors in capsule endoscopy frames using texture analysis based on the discrete wavelet transform*, Engineering in Medicine and Biology Society, 2008, EMBS 2008. 30th Annual International Conference of the IEEE, 2008, 3012–3015.

[2] LI B., MENG M. Q.-H., *Texture analysis for ulcer detection in capsule endoscopy images*, Image and Vision Computing, 2009, Vol. 27, No. 9, 1336–1342.

[3] CYCHNERSKI J., BLOKUS A., BRZESKI A., *Endoscopic Videos Deinterlacing and On-Screen Text and Light Flashes Removal and Its Influence on Image Analysis Algorithms' Efficiency*, International Journal of Image Processing and Visual Communication, 2013, Vol. 2, No. 1.

[4] CYCHNERSKI J., BLOKUS A., BRZESKI A., *Anonimizacja Obrazów z Nagrań Endoskopowych w Systemie Wspomagania Diagnostyki Chorób Przewodu Pokarmowego*, Zeszyty Naukowe Wydziału Elektroniki, Telekomunikacji i Informatyki Politechniki Gdańskiej, 2012, No. 10.

[5] HARTLEB M., *Zespół dysfunkcji krążeniowej towarzyszący marskości wątroby*, Przegląd Epidemiologiczny, 2005, Vol. 59, No. 2.

[6] OJALA T., PIETIKAINEN M., MAENPAA T., *Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, Vol. 24, No. 7.

[7] CANNY J., *A Computational Approach to Edge Detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, Vol. PAMI-8, No. 6.

[8] CYCHNERSKI J., *Anytime polynomial heuristic algorithm for partitioning groups of data with preserving class proportions for cross-validation*, Information Systems Architecture and Technology, 2014, No. 35.

[9] CYCHNERSKI J., BRZESKI A., BLOKUS A., *Method of training the endoscopic video analysis algorithms to maximize both accuracy and stability*, ICT Young 2013, No. 10.

[10] POH CHEE KHUN, ZHANG ZHUO, LIANG ZI YANG, LI LIYUAN, LIU JIANG, *Feature Selection and Classification for Wireless Capsule Endoscopic Frames*, Biomedical and Pharmaceutical Engineering, ICBPE '09, International Conference, 2009.

[11] KODOGIANNIS V.S., BOULOUGOURA M., *An Adaptive Neurofuzzy Approach for the diagnosis in Wireless Capsule Endoscopy Imaging*, International Journal of Information Technology, 2007, Vol. 13, No. 1.

[12] LI B., MENG M. Q.-H., *Tumor Recognition in Wireless Capsule Endoscopy Images Using Textural Features and SVM*, IEEE Transactions On Information Technology in Biomedicine, 2012, Vol. 16, No. 3.

[13] YUN SUB JUNG, YOUNG HO KIM, DONG HA LEE, JONG HYO KIM, *Active Blood Detection in a High Resolution Capsule Endoscopy using Color Spectrum Transformation*, International Conference on BioMedical Engineering and Informatics, 2008. BMEI 2008, Vol. 1.

[14] JIE LI, JINWEN MA, TILLO T., BAILING ZHANG, ENG GEE LIM, *A training based Support Vector Machine technique for blood detection in wireless capsule endoscopy images*, IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2012.

cross-validation, artificial intelligence training,
output averaging, endoscopy, cancer & blur recognition

Jan CYCHNERSKI*, Adam BRZESKI*, Paweł ROŚCISZEWSKI*

OBTAINING A WELL-TRAINED ARTIFICIAL INTELLIGENCE ALGORITHM FROM CROSS-VALIDATION IN ENDOSCOPY

The article shortly discusses endoscopic video analysis problems and artificial intelligence algorithms supporting it. The most common method of efficiency testing of these algorithms is to perform intensive cross-validation. This allows for accurately evaluate their performance of generalization. One of the main problems of this procedure is that there is no simple and universal way of obtaining a specific instance of a well-trained algorithm which has efficiency comparable to efficiency suggested by cross-validation. In this paper, a method resolving this problem (at some circumstances) is proposed and examined in the task of recognizing cancer, healthy tissue, blurred frames and sharp frames on endoscopic videos by two exemplary artificial intelligence algorithms designed for this task, using neural networks and support vector machines. The results show that proposed method allows to obtain algorithms trained a little better results than the average algorithm after cross-validation, without requiring any additional testing nor training.

1. INTRODUCTION

Algorithms for automated computer aided medical examinations recently are gaining increasing popularity. This is due, among other things, an increasing amount of data provided to physicians with different kinds of examinations. Such a situation occurs, inter alia, in the case of capsule endoscopy, where a single examination consists of 2–8 hours of video recording of the entire gastrointestinal tract. It is therefore important for physicians to speed up the video analysis process by some computer-aided assistance [1].

For this purpose, a number of disease recognition algorithms on endoscopic videos were created. Common feature of most of them is the use of artificial intelligence

* Department of Computer Architecture, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland, e-mail: jan.cychnerski@eti.pg.gda.pl, brzeski@eti.pg.gda.pl, pawel.rosciszewski@pg.gda.pl

methods – such as artificial neural networks (ANN) or support vector machines (SVM). Such algorithms require precise learning and testing. In medical systems it is also crucial to know predicted efficiency (e.g. sensitivity and specificity) of their components. For these purposes almost always cross-validation is used, which consists of multiple learning and testing the main algorithm with chosen subsets of the whole input data [1], [2].

On the other hand, the purpose of the learning process is primarily to obtain a well trained classifier (e.g. neural network). The problem, which arises at this point is that the process of cross-validation on the output gives only efficiency measures and many differently trained classifiers [3]. Usually this problem is omitted by training another classifier or ensemble of classifiers on the whole input data [1], [4]. This method requires additional computations and maybe another algorithms. Second resolution of the problem is to select one of the classifiers already trained by cross-validation. Choosing the *right* one is, however, not a trivial problem, and may not be even possible.

In this paper the intermediate solution to these problems is proposed and tested.

2. PROBLEM DEFINITION

At the input we have an artificial intelligence algorithm and a set of test data. One wants to carry out a cross-validation on these data, and as a result receive an algorithm, whose efficiency is comparable to the efficiency predicted by cross-validation.

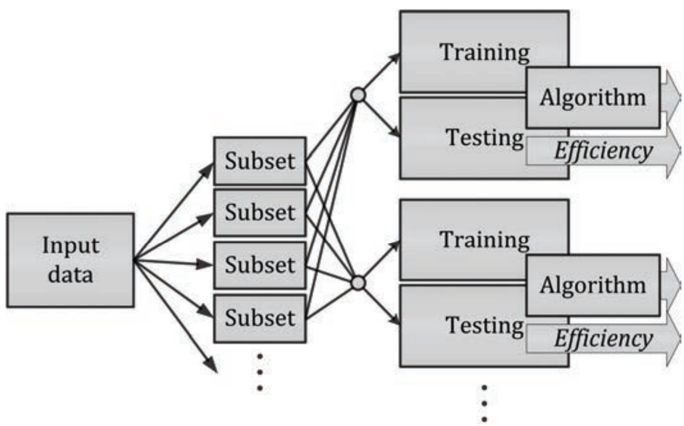


Fig. 1. Cross-validation process

In the literature [1], [2] the most frequently several types of cross-validation are found. Each of them repeat the learning-testing procedure *x* times:

1. 2-fold cross-validation – the data set is divided into two equal parts; $x = 2$;
2. k-fold cross-validation – the data set is divided into two k equal parts, $x = k$;
3. *leave-one-patient-out* cross-validation – the data set is divided into p parts (where p is the number of patients in the whole data set), each containing data belonging to exactly one patient; $x = p$;
4. *leave-one-sample-out* cross-validation – the data set is divided into n singletons (where n is the number of samples in the whole data set), $x = n$.

As can be seen, the x is usually a fairly large number. Thus, after finishing the testing process, one has x variously trained classifiers, where each classifier has a particular effectiveness on its test set (which may *not* correspond to the *actual* effectiveness on the whole input data!).

At this point, the problem is to choose one of them as that which would have the highest effectiveness of generalization. It is not possible to assess this on the basis of only cross-validation. For this reason, often the additional learning of another classifier on the entire data is performed, completely bypassing cross--validation [1]. However, in this case, one loses information about the effectiveness of the algorithm [5].

Another solution is to select one (e.g. best, random or average) algorithm trained by the cross-validation. In this case, however, we do not have at all certainty of its actual effectiveness (in the case of large x, one can be almost sure that the generalization effectiveness will be much lower than the result of testing, just because of the accident) [3].

3. THE PROPOSED ALGORITHM

In this paper, another solution to this problem is proposed. Instead of choosing one output algorithm, the method takes them all, and averages their outputs. So having x instances of trained classifiers, one must classify the data with all of them, and as a final answer choose the average of their responses.

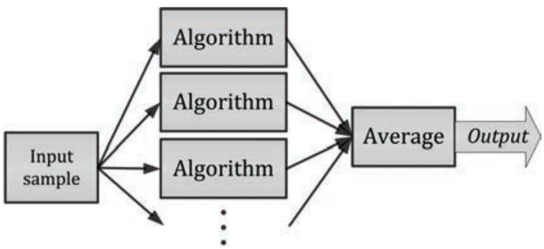


Fig. 2. Proposed algorithm

For set of x algorithms obtained through cross-validation $A = \{a_1, a_2, ..., a_x\}$, the final output of proposed method is:

$$\text{output} = \sum_i \frac{\text{output}(a_i)}{x} \tag{1}$$

Also, method's efficiency may be approximated by the average of each algorithm's efficiency obtained during cross-validation:

$$\text{efficiency} \approx \sum_i \frac{\text{effifiency}(a_i)}{x} \tag{2}$$

4. EXPERIMENT

4.1. TEST DATA

As the input data for training and testing of the recognition algorithms, several dozen of real-life endoscopic videos were taken. Every frame of every video had given (by an expert) a set of labels about its contents. Labels included: *blurred frame, sharp frame, healthy tissue, abnormal tissue, cancer, Crohn's disease* and many others.

The data set was split into cross-validation sets with algorithm described in [6]. Because of high computational complexity of methods 3 and 4, in the experiments only k-fold cross-validation was examined, wherein $k = 16$. This allows for more accurate testing of the algorithms, while still maintaining the clarity and generality of results.

During experiments, two groups of tests were performed:

1. recognition of blurred frames (4541 chosen frames from 18833 total) *vs.* sharp frames (4864/34688),
2. recognition of cancer (1477/23704) *vs.* healthy tissue (1281/10974).

The data came from real endoscopic examinations of 178 patients. To balance the data sets, from each video a maximum of 30 frames were chosen for each recognized object (i.e. max 30 blurred + max 30 sharp + max 30 cancer...). Except this data, additional 8 videos were used for results visualization.

4.2. RECOGNITION ALGORITHMS

In order to evaluate the operation of the proposed method, it was tested on two versions of a good recognition algorithm designed for recognizing large bowel disorders: *Kodogiannis* + ANN [6] and *Kodogiannis* + SVM [7]. Each of these algorithms was a subject of cross-validation with data sets described in the previous paragraphs.

After cross-validation, final validation test was performed on the another set of videos – for each of the videos a full analysis by every test algorithm was performed (there was no restriction on the number of frames). Output of each of the algorithms was recorded and compared with the expected result. The results of these tests are shown in the next section.

5. RESULTS

In this section, efficiency results of recognition algorithms on exemplary videos (not included in training nor testing sets) are presented.

On the charts, *Y* axis denotes sensitivity, specificity (top and bottom numbers) and efficiency score (middle number).

On *X* axis, first 16 columns are the cross-validation algorithms results, *Output* is the result of the proposed method, and (for comparison) the last result is an algebraic average of all 16 cross-validation results.

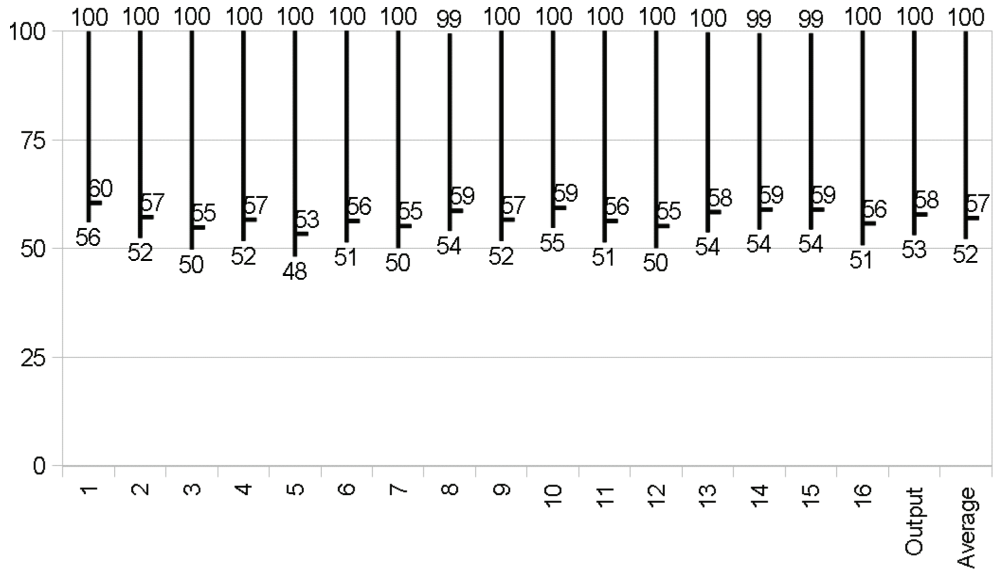


Fig. 3. Algorithm with ideal sensitivity but low specificity

Figure 3 presents a disease recognition algorithm trained for recognizing mainly positive samples (i.e. with a disease present). In this case, the algorithm recognizes every occurrence of abnormality (has almost 100% sensitivity), but performs almost randomly when given negative sample on the input. The proposed method gives results comparable to them.

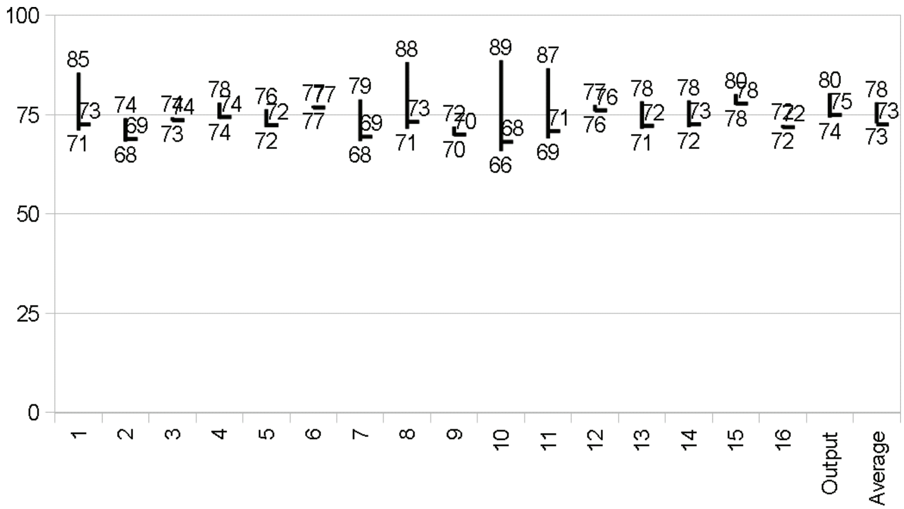


Fig. 4. Algorithm with high sensitivity and specificity

Figure 4 shows results of blur recognition algorithm with high both specificity and sensitivity (in this case, usually specificity is a little higher), but after cross-validation it is difficult to choose best one, as they perform differently – some of them have higher specificity, some higher sensitivity – i.e. they are quite well but differently trained. Proposed method minimizes variety of the output (specificity and sensitivity rates are quite the same) and performs better than any other algorithm. This is the case where this method is useful the most.

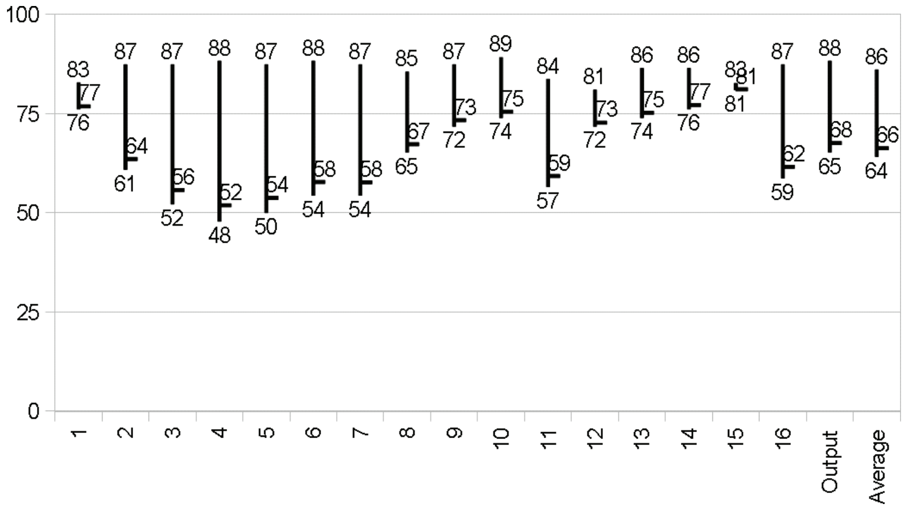


Fig. 5. Algorithm with low sensitivity but high specificity

Figure 5 shows a situation where there are many cross-validation algorithms that have low sensitivity. As a result, proposed method does not perform as well as the best cross-validation algorithm, but still preserves high efficiency.

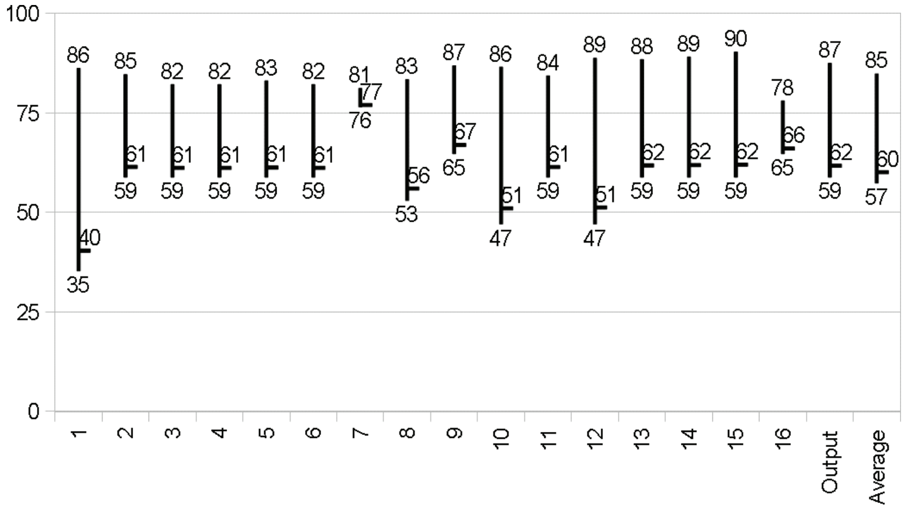


Fig. 6. Algorithm with moderate sensitivity and specificity

Figure 6 shows a situation where cross-validation algorithms perform quite similarly to each other and have moderate sensitivity and specificity. The proposed method preserves this behaviour and performs a little better than them on the average.

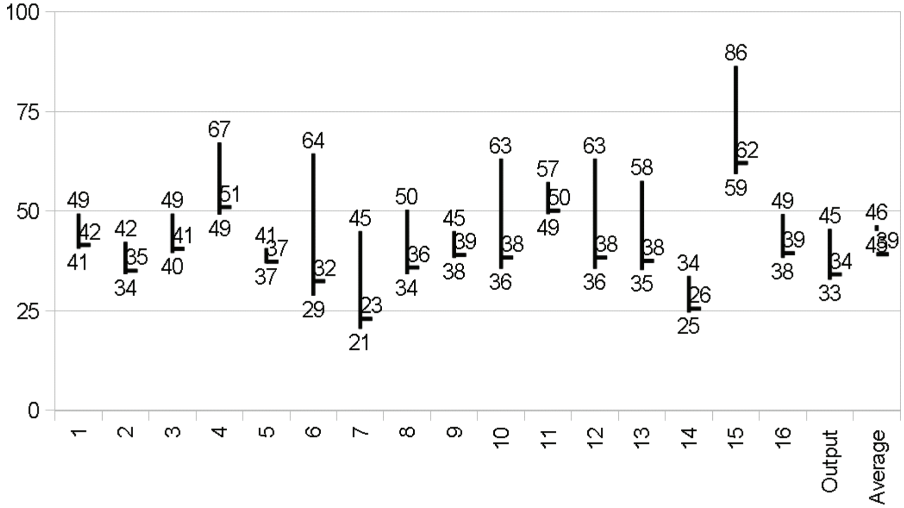


Fig. 7. Poorly trained algorithm (almost random) with a few good versions

Figure 7 shows a poorly trained algorithms, but where few of cross-validation versions of it are trained much better than the others. In this case, the proposed method gives bad results, much worse than optimal. Such situation is the worst case for this method. Before applying this method to a cross-validated artificial intelligence system, it must be assured that the process of classifier training does not produce such results.

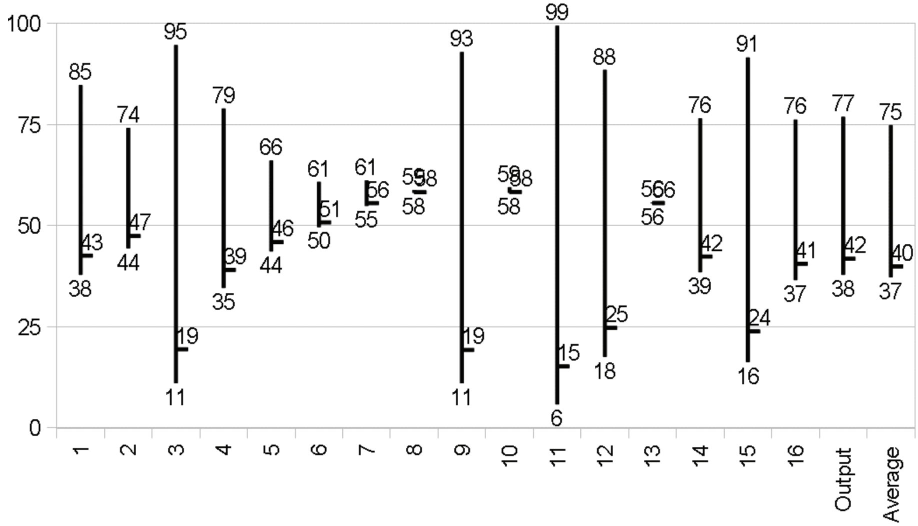


Fig. 8. Poorly trained algorithm (almost random)

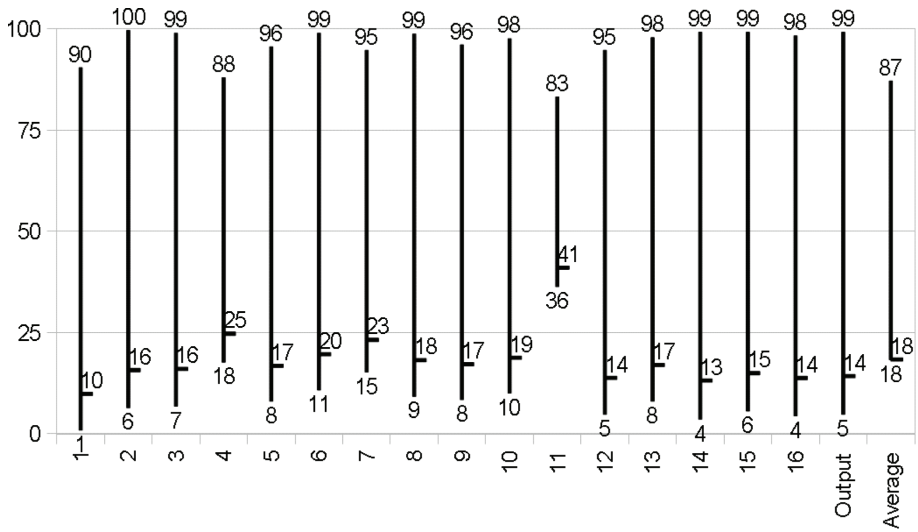


Fig. 9. Almost not trained algorithm

Figure 8 shows a poorly trained algorithms, where outputs seem to be almost randomly scattered regardless the input. The proposed method allows for recovering some information from them (of course, the output is still not very useful). On the other hand, it is likely that there are some cross-validation algorithms that perform much better than proposed method (but it is still difficult to guess which ones).

The last figure shows an algorithm that is almost not trained. It gives quite random outputs (sometimes specificity is almost 100%, sometimes reversely: sensitivity reaches ~100%). In this case, the proposed method does not help at all, which is a predicted behaviour as it is impossible to get useful information from such untrained algorithms.

6. CONCLUSION

Results of experiments presented in this article show that the proposed method is a good choice when it is important to obtain a well-trained classifier after performing cross-validation, with efficiency almost equal to efficiency predicted by cross-validation. This method does not require any additional computations nor algorithms (besides cross-validation itself). Classifiers obtained by proposed method perform usually a little better than algorithms trained through cross-validation process on the average. Nonetheless, if cross-validation algorithms perform variously (i.e. many of them are poorly trained) this method does not improve overall efficiency in such situation – final output is often much worse than the best algorithm trained through cross-validation.

ACKNOWLEDGEMENTS

This research was partially funded by grants from National Centre for Research and Development (PBS2/A3/17/2013, *Internet platform for data integration and collaboration of medical research teams for the stroke treatment centers*).

REFERENCES

[1] LIEDLGRUBER M., UHL A., *Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review,* IEEE Reviews in Biomedical Engineering, 2011, Vol. 4, 73–88.

[2] KARARGYRIS A., BOURBAKIS N., *Wireless Capsule Endoscopy and Endoscopic Imaging: A Survey on Various Methodologies Presented*, IEEE Engineering in Medicine and Biology Magazine, 2010, Vol. 29, No. 1, 72–83.

[3] LIU Y.L.Y., *Create Stable Neural Networks by Cross-Validation*, The 2006 IEEE International Joint Conference on Neural Network Proceedings, 2006.

[4] KOHAVI R., *A study of cross-validation and bootstrap for accuracy estimation and model selection,* International Joint Conference on Artificial Intelligence, 1995, 1137–1145.

[5] CYCHNERSKI J., BRZESKI A., BLOKUS A., *Performance of endoscopic image analysis algorithms in large bowel videos processing*, [in:] ICT Young 2013, 2013, no. 10.
[6] CYCHNERSKI J., *Anytime polynomial heuristic algorithm for partitioning groups of data with preserving class proportions for cross-validation*.
[7] CYCHNERSKI J., *Artificial intelligence support for disease detection in wireless capsule endoscopy images of human large bowel*, [in:] Proceedings of the 5th International conference of Young Scientists: Computer Science & Engineering 2011, November 2011, 24–27.
[8] KODOGIANNIS V., BOULOUGOURA M., *An adaptive neurofuzzy approach for the diagnosis in wireless capsule endoscopy imaging*, International Journal of Information Technology, 2007, 13(1).

cross-validation, data set partitioning, grouping, classes, anytime heuristic algorithm, greedy algorithm

Jan CYCHNERSKI*

ANYTIME POLYNOMIAL HEURISTIC ALGORITHM FOR PARTITIONING GROUPS OF DATA WITH PRESERVING CLASS PROPORTIONS FOR CROSS-VALIDATION

The article describes a problem of splitting data for k-fold cross-validation, where class proportions must be preserved, with additional constraint that data is divided into groups that cannot be split into different cross-validation sets. This problem often occurs in e.g. medical data processing, where data samples from one patient must be included in the same cross-validation set.

As this problem is NP-complete, a heuristic anytime polynomial algorithm is proposed and described in the article. Also, it is experimentally compared to two other, simpler algorithms.

1. INTRODUCTION

Efficiency and quality of most of artificial intelligence algorithms must be measured experimentally. Additionally, many of them have some input parameters which need to be carefully tuned to achieve maximum possible efficiency and consistency. To perform such tuning and efficiency testing, when only a limited set of data can be provided (as in most real-life cases), it is essential to use an efficient strategy allowing to make maximal usage of test data. Cross-validation is one of the most used methods to achieve this goal [1]. The main idea of cross-validation is to perform training–testing procedure many times, each time on a different subset of given data (subsets may overlap). The mostly used form of cross-validation is *k-fold cross-validation*. In this method the whole test data is split into k non-overlapping subsets. Training is performed on $k – 1$ subsets, and testing on the last subset. The process is repeated k times, each time with different testing subset.

* Department of Computer Architecture, Gdańsk University of Technology, Poland, e-mail: jan.cychnerski@eti.pg.gda.pl

2. PROBLEM DEFINITION

However, in many situations (e.g. in medicine [2], [3]) there are some dependencies between data samples, which must be considered during data splitting process. One of such dependencies is that some data samples are very similar to other (e.g. the same movie fragment, the same patient, etc.) – in such cases splitting these groups of similar data into training and testing sets may lead to high overestimation of algorithm's efficiency. To remove this effect, groups of similar data must be always put into the same cross-validation set.

Moreover, if the artificial intelligence problem is multi-class, it is often crucial that proportions of data classes in each cross-validation set should be similar, and also the total amount of data in each set should be almost equal [2].

In this article, a heuristic algorithm that try to fulfill all of these requirements is proposed. An example of the problem and its approximate solution is presented in table 2.

Table 2. Sample data and sets

	SET S1	**SET S2**	**SET S3**	**SET S4**	
	group G1	**group G2**	**group G3**	**group G4**	
Classes	10, 5, 5, 0	5, 5, 5, 5	0, 10, 15, 5	10, 20, 30, 5	
$\sum$	20	20	30	65	
	group G5	**group G6**	**group G7**		
Classes	0, 5, 15, 5	5, 10, 25, 0	0, 15, 15, 0		
$\sum$	25	40	30		
	group G8				
Classes	0, 10, 10, 5				
$\sum$	25				
	TOTAL				**MEAN**
Class c1	10	10	0	10	7,5
Class c2	20	15	25	20	20
Class c3	30	30	30	30	30
Class c4	10	5	5	5	6,25
$\sum$	70	60	60	65	63,75

3. ALGORITHMS

The input consists of the whole data set $D = \{d_1, d_2, ..., d_n\}$ and data classes set $C = \{c_1, ..., c_{m.}\}$. Each data sample belong to at least one class $d_i \in \{c_{1,1}, c_{1,2}, ...\}$. The data is divided into separate groups $G_1, G_2, ..., G_g$, $G_i = \{d_{i,1}, d_{i,2}, ...\}$, where each data sample d belongs to exactly one group G.

The task is to find such apportionment of the groups G_i into k sets, that each set would have similar number of data samples d, similar number of groups G, and similar proportions of data classes.

Let n be the total number of data samples in the input, g be the total number of groups in the input, $|c_j|$ be the total number of data samples belonging to class c_j, and g_j be the total number of groups that have at least one element of class c_j.

Now, let $|S_i|$ be the total number of data samples in set S_i, groups(S_i) be the total number of groups in set S_i, class$_j$(S_i) be the total number of data samples belonging to class c_j in set S_i, and groupclass$_j$(S_i) be the total number of groups containing data belonging to class c_j in set S_i.

Now, let $\hat{n} = \frac{n}{k}$ be the ideal desired number of data samples in each output set, $\hat{g} = \frac{g}{k}$ be the ideal desired number of groups in each output set, $\hat{c}_j = \frac{|c_j|}{k}$ be the ideal desired number of data samples of class c_j in each output set, and $\hat{g}_j = \frac{|g_j|}{k}$ be the ideal desired number of groups containing samples belonging to class c_j in each output set.

In this definition, the task is to find optimal allocation of groups G_i into sets S_i, that would fulfill simultaneously following requirements:

1. $\forall_i |S_i| \approx \hat{n}$;
2. $\forall_i \text{groups}(S_i) \approx \hat{g}$;
3. $\forall_{i,j} \text{class}(S_i) \approx \hat{c}_j$;
4. $\forall_{i,j} \text{groupclass}_j(S_i) \approx \hat{g}_j$.

3.1. PARTITIONING QUALITY METRIC

In most cases, these four requirements are contradicting (depending on input data). Therefore, an apportionment quality metric is required for balancing them. This metric must take into account all beforementioned aspects, and allow to compare different apportionments. This metric must be separately adapted for specific problem, as in some cases it is more important to equalize proportions of classes, but in other cases it is essential to equalize sizes of sets, etc.

Let $Q(S_1, S_2, ...) = \sum_i Q(S_i)$ be such a metric, where $Q(S_i)$ is arbitrary, and $Q \leq 0$, where 0 is quality value of ideal apportionment.

In this article, following function Q is proposed:

$$Q(S_i) = f_1(S_i) + 2 f_2(S_i) + f_3(S_i)$$

$$f_1(S_i) = -\frac{\left||S_i| - \hat{n}\right|}{\hat{n}}$$

$$f_2(S_i) = -\sum_j \frac{\left|\text{class}_j(S_i) - \hat{c}_j\right|}{j\hat{c}_j}$$

$$f_3(S_i) = -\sum_j \frac{\left|\text{gropupclass}_j(S_i) - \hat{g}_j\right|}{j\hat{g}_j}$$

3.2. NAÏVE ALGORITHM

The simplest and naïve solution of the problem is to randomly split G into k equally sized subsets (i.e. each set consists of the same number of groups). This algorithm gives acceptable results only for big number of groups with small number of equally distributed classes.

3.3. GREEDY ALGORITHM

The more *intelligent*, but still simple algorithm, is a greedy algorithm with scoring. It simply assign each group to the set which score will improve the most (or worsen the least). Additionally, groups are sorted by their decreasing size, which allows to consider bigger, more important groups before smaller.

Algorithm 1 Greedy algorithm with scoring

```
S_1 ← ∅, S_2 ← ∅, ...
G = {G_1, G_2, ...}
sort G by descending group size
for all g ∈ G do
    I = argmax_i Q(S_i ∪ {g}) − Q(S_i)
    S_I = S_I ∪ {g}
end for
```

The greedy algorithm works well if there are many groups, and groups are quite similar to each other. Computational complexity is $O(glogg + nk)$ (because of sorting and Q calculations).

3.4. ANYTIME ALGORITHM

Anytime algorithms are the algorithms which can be interrupted before its end, but still can give an approximate, correct result [4]. As algorithm works longer, this result becomes better and better. It this article, an extension of simple greedy algorithm to its anytime version is proposed as a better solution for aforementioned group partitioning problem. To simplify the notation, let $\varphi(\{x_1, x_2, x_3, ...\}) = \{\emptyset, \{x_1\}, \{x_2\}, \{x_3\}, ...\}$.

Algorithm 2 Anytime algorithm

1: init $S_1, S_2, ...$ with greedy algorithm with scoring
2: **loop**
3: $\quad \hat{S} \leftarrow \{S_1, S_2, ...\}$
4: $\quad$ sort $\hat{S}$ by ascending $Q(S_i)$
5: $\quad \bar{S} \leftarrow \begin{cases} (\hat{S}_1, \hat{S}_k), (\hat{S}_1, \hat{S}_{k-1}), ..., (\hat{S}_1, \hat{S}_2), \\ (\hat{S}_2, \hat{S}_k), ..., (\hat{S}_2, \hat{S}_3), \\ ..., \\ (\hat{S}_{k-1}, \hat{S}_k) \end{cases}$
6: $\quad a, b \leftarrow \emptyset$
7: $\quad$ **for all** $(A, B) \in \bar{S}$ **do**
8: $\quad\quad a, b \leftarrow \underset{a\in\varphi(A), b\in\varphi(B), Q(A\cap a\cup b)>Q(A)}{\operatorname{argmax}} Q(A \cap a \cup b) + Q(B \cup a \cap b)$
9: $\quad\quad$ **if** $a \neq \emptyset$ **or** $b \neq \emptyset$ **then**
10: $\quad\quad\quad$ **break for**
11: $\quad\quad$ **end if**
12: $\quad$ **end for**
13: $\quad$ **if** $a = \emptyset$ **and** $b = \emptyset$ **then**
14: $\quad\quad$ **break loop**
15: $\quad$ **else**
16: $\quad\quad A \leftarrow A \cap a \cup b$
17: $\quad\quad B \leftarrow B \cup a \cap b$
18: $\quad$ **end if**
19: **end loop**

In words, the algorithm tries to iteratively improve the solution returned by a greedy algorithm, by exchanging or moving some elements of selected sets in order to improve quality Q of the worst set for which it is possible.

The algorithm can be interrupted at any time after step 1. The algorithm is a deterministic anytime algorithm. Pessimistic computational complexity is $O(k^2 g^3)$.

4. EXPERIMENT

In this section, results of performed experiments are presented. All three beforementioned algorithms were tested on 22 random input data sets:

For each test repetition, the data was generated in two modes:

- number of data samples in each group was random (uniform distribution). Results of this mode are presented on charts (a), (c), (e).
- each group contains data samples from exactly one class. Number of samples is random (uniform distribution). Such data is typical for many applications (e.g. medical).

For each generated data set, three partitioning algorithms were tested:

- naïve algorithm – charts (a), (b);
- greedy algorithm – charts (c), (d);
- anytime algorithm – charts (e), (f).

Table 1. Input data sets

Test ID	Number of groups	Number of classes	Max samples per class in one group	Test repetitions
1-5	1000	1-5	100	100
6-10	100	1-5	100	100
11-15	50	1-5	100	1000
16-20	20	1-5	100	1000
21	100	3	1000,100,50	100
22	1000	3	1000,100,50	100

4.1. RESULTS

In this section results of all performed experiments are presented. To understand the charts, see figure 1.

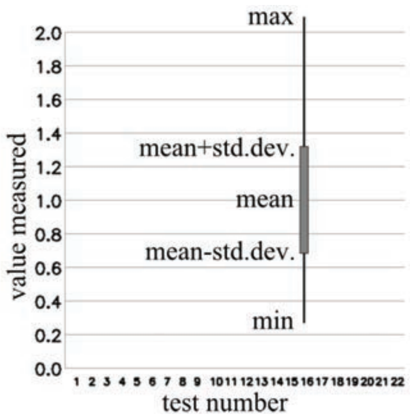


Fig. 1. Description of the charts

Figure 2 shows the sizes of output sets. Values on Y axis indicate proportion between actual set sizes, and ideal desired set sizes, i.e. $y = \frac{|S_i|}{\hat{n}}$. In this case, anytime algorithm outperformed the other algorithms.

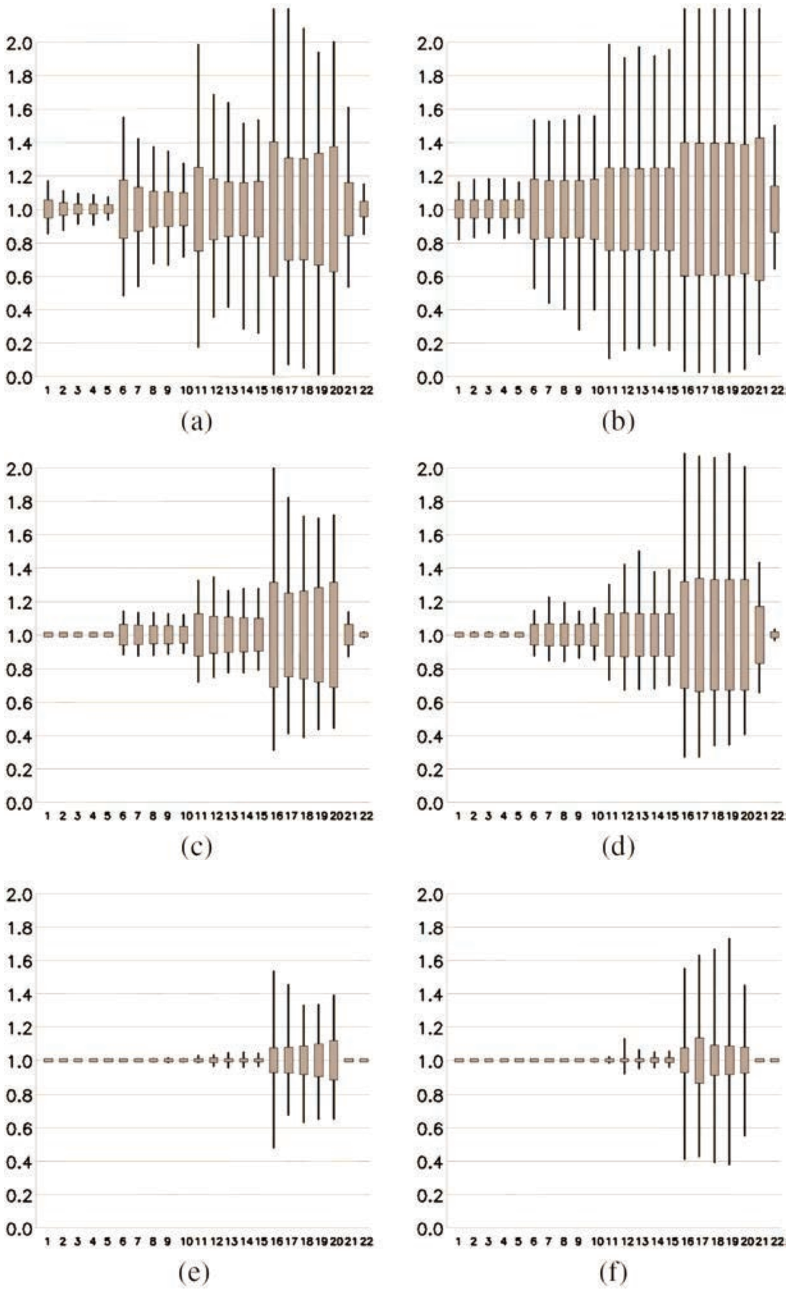


Fig. 2. Output set sizes, $y = \frac{|S_i|}{\hat{n}}$

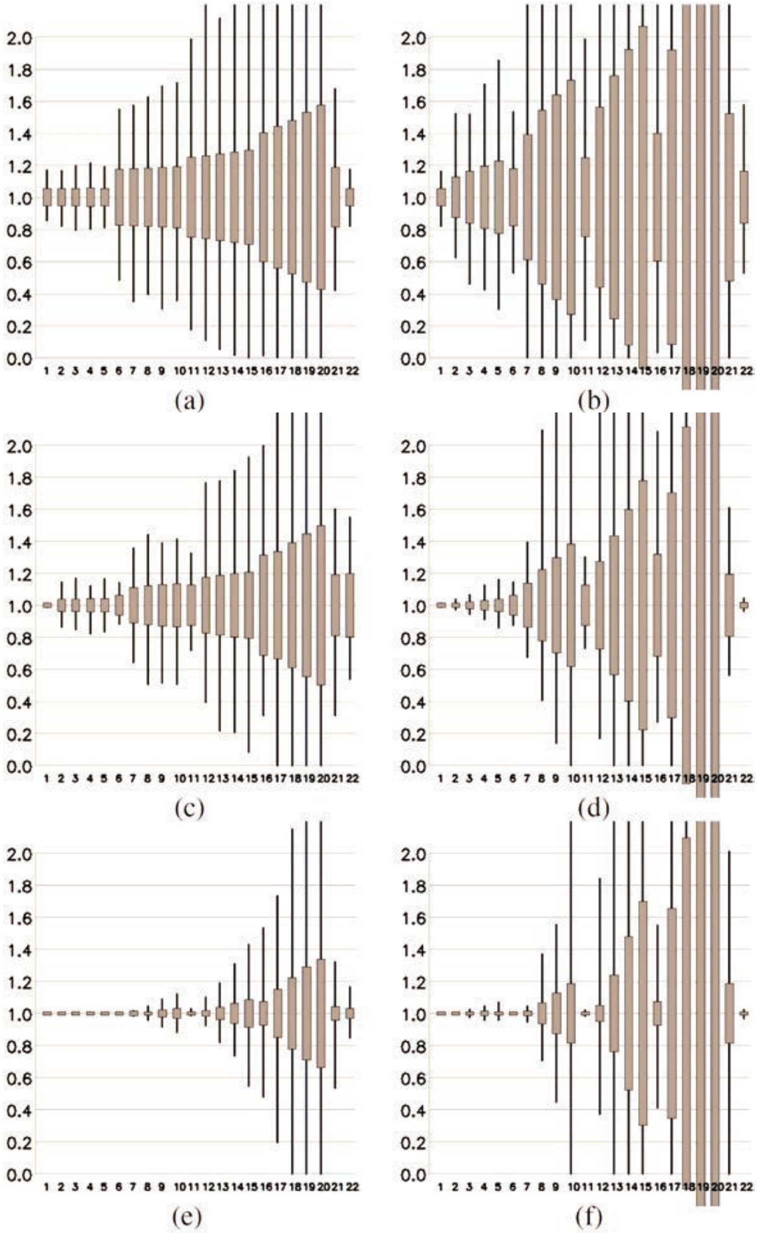


Fig. 3. Data samples for each class in output sets, $y = \frac{\text{class}_j(S_j)}{\hat{c}_j}$

Figure 3 shows the number of data samples by classes in output sets. Values on *Y* axis indicate proportion between actual and desired number of data samples for each class, i.e. $y = \frac{\text{class}_j(S_j)}{\hat{c}_j}$. In this case, the anytime algorithm definitely outperformed all other algorithms.

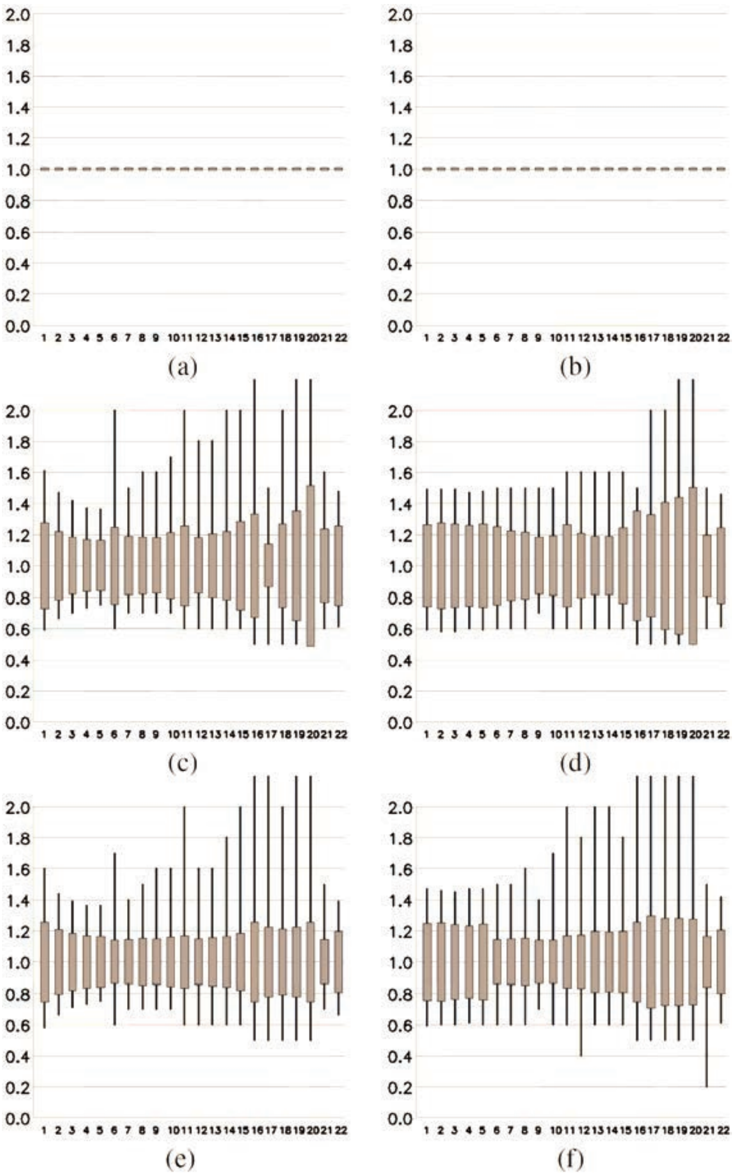


Fig. 4. Total number of groups in output sets, $y = \frac{\text{groups}(S_i)}{\hat{g}}$

Figure 4 shows the number of groups in output sets. Values on *Y* axis indicate proportion between actual and desired total number of groups in each output set, i.e. $y = \frac{\text{groups}(S_i)}{\hat{g}}$. In this case, naive algorithm gives ideal results (because of its construction) – which does not matter, as this algorithm is much worse in all other tests. Ignoring the naive algorithm, the anytime algorithm performed best.

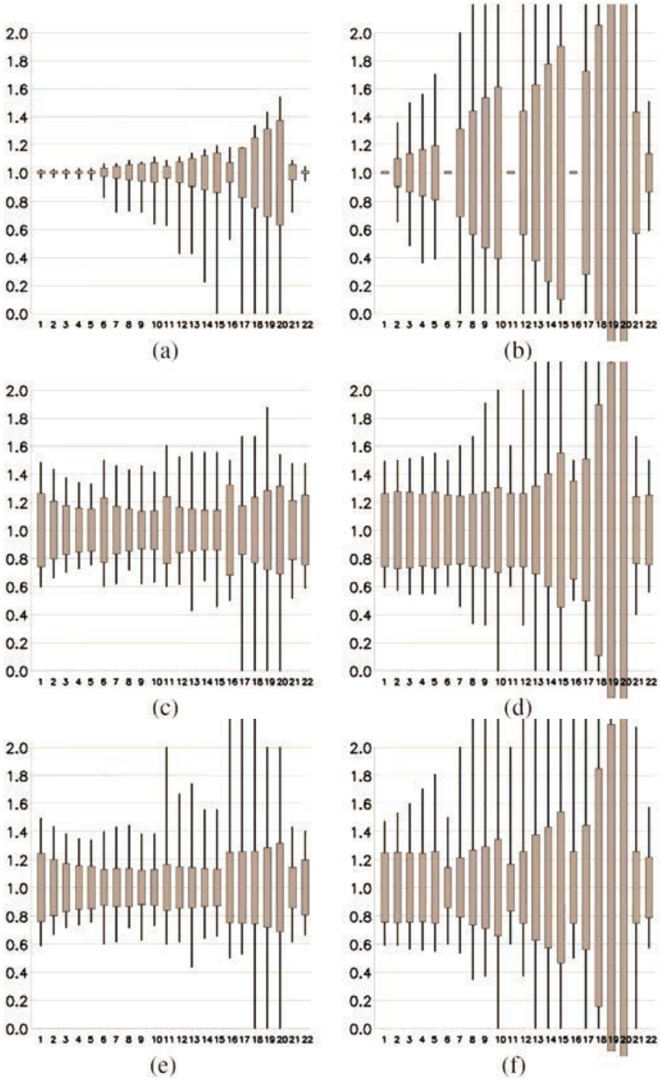


Fig. 5. Number of groups for each class in output sets, $y = \frac{\text{groupclass}_j(S_i)}{\hat{g}_j}$

Figure 5 shows the number of groups of given class. Values on *Y* axis indicate proportion between actual and desired number of groups for each class in output sets, i.e. $y = \frac{\text{groupclass}_j(S_i)}{\hat{g}_j}$. In this case, the anytime algorithm usually give best results, but differences between algorithms are smaller than in other tests. It's mainly because of fact that the task of balance number of groups by classes is more difficult, but less important than other balancing tasks in this problem.

5. CONCLUSION

Concluding, the proposed anytime algorithm gives acceptably good results for described randomized input data (which is similar to many typical problems), especially for small number of classes uniformly distributed through groups. Output cross-validation data sets fulfill quite well all four described requirements, which allows them to be used in practical applications.

ACKNOWLEDGEMENTS

This research was funded by grants from National Centre for Research and Development (PBS2/A3/17/2013, *Internet platform for data integration and collaboration of medical research teams for the stroke treatment centers*).

REFERENCES

[1] BURMAN P., *A Comparative Study of Ordinary Cross-Validation, v-Fold Cross-Validation and the Repeated Learning-Testing Methods*, Biometrika, 1989, Vol. 76, No. 3, 503–514.

[2] LIEDLGRUBER M., UHL A., *Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review*, IEEE Rev. Biomed. Eng., 2011, Vol. 4, 73–88.

[3] KARARGYRIS A., BOURBAKIS N., *Wireless Capsule Endoscopy and Endoscopic Imaging: A Survey on Various Methodologies Presented*, IEEE Eng. Med. Biol. Mag., 2010, Vol. 29, No. 1, 72–83.

[4] ZILBERSTEIN S., *Using Anytime Algorithms in Intelligent Systems*, AI Magazine, 1996, 17(3), 73–83.

PART 4

EXPERT AND SOFTWARE SYSTEMS DESIGN

Decisional DNA, Set of Experience Knowledge Structure, Virtual Engineering Objects

Syed Imran SHAFIQ*, Cesar SANIN*,
Edward SZCZERBICKI**, Carlos TORO***

VIRTUAL ENGINEERING OBJECTS (VEO): DESIGNING, DEVELOPING AND TESTING MODELS

In this article, the development and implementation of the concept of Virtual Engineering Object (VEO) is described. A VEO is a computerized real world representation of an engineering object. VEO will act as a living representation of the object capable of adding, storing, improving and sharing knowledge through experience, in a way similar to an expert of that object. In this paper, it is shown through test models how the concept of VEO can be implemented with the Set of Experience Knowledge Structure (SOEKS) and Decisional DNA. The SOEKS/DDNA is a flexible and standard knowledge representation structure to acquire and store experiential knowledge. A test case study for three different drilling machines, drilling tools and the working holding devices is developed to test and demonstrate the implementation of VEO. The test model confirmed that the concept of VEO is able to capture and reuse the experience of the engineering artifacts, which can be beneficial for efficient decision making in industrial design and manufacturing.

1. BACKGROUND

A large percentage of the product's life cycle time is spent on routine tasks; it takes up to 80% of the design time. It is noted that around 20% of the designer's time is spent in searching for and absorbing information, and 40% of all design information requirements are currently met by designer's personal records/stores, even though more suitable information may be available from other sources. This implies that design information and knowledge is not represented in a shared and easily accessible knowledge base [1].

* The University of Newcastle, University Drive, Callaghan, 2308, NSW, Australia.
** Gdansk University of Technology, Gdansk, Poland.
*** Vicomtech-IK4, San Sebastian, Spain.

Wang et al. [2] provided a good example of a new car being designed where 40–50% of its parts are entirely the same as existing ones, 30–40% require slight modification of existing ones and only 10–20% of the components are entirely new. Evoking a previously solved case to solve a new one is the basis of the design reuse theory; thus, past experience can play a vital role.

It is evident that manufacturing planning at the conceptual or early design stage is the key for designers to evaluate manufacturability in terms of criteria and metrics such as costs and time. However, there are not many techniques and software tools for conceptual manufacturing planning.

Cyber Physical System (CPS) is emerging as a must have technology needed by industry [3, 4]. CPS are integrations of computation with physical processes [4, 5]. Embedded computers and networks monitor and control the physical processes, usually with feedback loops where physical processes affect computations and vice versa. In the physical world, the passage of time is inexorable and concurrency is intrinsic. Neither of these properties is present in today's computing and networking abstractions [4]. CPS aims to integrate knowledge and engineering principles across the computational and engineering disciplines (networking, control, software, human interaction, learning theory, as well as electrical, mechanical, chemical, biomedical, material science, and other engineering disciplines) to develop new CPS science and supporting technology. Scalable CPS architectures for adaptive and smart manufacturing systems which dynamically enable the continuous design, configuration, monitoring and maintenance of operational capability, quality, and efficiency are, in fact, a requirement for the industry [6]. According to the European commission under the Horizons 2020 programme, the self-learning closing feedback loop between production and design should be included in future factories for optimizing energy expenditure and minimizing waste as a direct relation to the enhancement in control and immediate information processing that a CPS will provide.

Many knowledge-based techniques have been used in past that aim to organize past, present, and future information. Some of the important objectives of these techniques are sharing the information, forecasting, and generating new knowledge. Knowledge-based techniques used in the past had limited success as they were having some shortcomings like they were time consuming, not very intelligent, etc. Moreover, most of these knowledge systems are designed for a specific domain. The applicability of these systems significantly decreases in any other area. This makes them less flexible and less versatile. They do not have any standard knowledge representation. Most systems lack the capability for information sharing and exchange [7–9]. Another important limitation of current knowledge-based techniques is that they do not take into consideration formal experience. However, in recent times, SOEKS and DDNA, a smart knowledge representation technique, have emerged as strong general purpose solutions to the above-mentioned shortcomings that other knowledge based systems are facing. They are a combination of organized information obtained from formal decision events and has been successfully implemented in a variety of domains [10].

2. INTRODUCTION TO SOEKS AND DECISIONAL DNA

Sanin and Szczerbicki [11–16] introduced for the first time the Set of experience knowledge structure (SOEKS) and Decisional DNA (DDNA). SOEKS is a new, multi-domain knowledge representation technique, not only capable of gathering experience and formal decisions but also a tool for decision making process. Unlike other knowledge based systems, DDNA is more focused on extracting and evolving knowledge through experience, and reusing this knowledge to support decision-making.

2.1. SET OF EXPERIENCE KNOWLEDGE STRUCTURE (SOEKS)

The SOEKS is a dynamic structure, which feeds on the formal decision events and uses them for the representation of the experiential knowledge. A formal decision is defined as "a choice [decision] made or a commitment to act that was the result [consequence] of a series of repeatable actions performed in a structured manner" [17]. A SOE (short form of SOEKS) has four components; variables (V), Functions (F), Constraints (c) and Rules (R) as shown in Fig. 1. Each formal decision is represented and stored in a unique way based on these four components.

The variables are the source of the other components of SOEKS and are the centre root or the starting point of the structure. The functions are based upon the relationships and associations among the variables. They create links between dependent and non-dependent variables constructing multi-objective goals. The third components of SOE are constraints, like functions and are connected to variables. They specify limits and boundaries and provide feasible solutions. Rules, the fourth component of SOEKS, are conditional relationships that operate on variables. Rules are relationships between a condition and a consequence connected by the statements "IF-THEN-ELSE".

2.2. DECISIONAL DNA (DDNA)

The concept of DDNA is the metaphor of human DNA. A group of SOE of the same category comprises of a kind of chromosome, as DNA does with genes. These chromosomes or groups of SOE make a category, and they are bases for making decisions. Each module of chromosomes forms an entire inference tool, and creates a Decisional DNA as shown in Fig. 1.

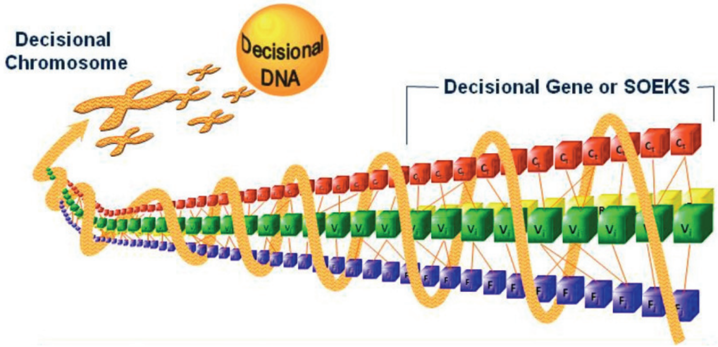


Fig. 1. Decisional DNA [18]

3. VIRTUAL ENGINEERING OBJECT (VEO)

VEO development together with Decisional DNA aim at enhancing industrial design and manufacturing. The need to carry out this development and rationale to adopt SOEKS is already discussed in the previous sections. In order to improve industrial design and manufacturing, we aim to capture the experience and knowledge of engineering artefacts, than re-use this knowledge for better decision making. To achieve this goal we conceived Virtual Engineering Objects (VEO).

A VEO is a representation of an artefact which can behave like an expert of that artefact and can help the practitioners in effective decision making based on the past experience. The concept of Virtual Engineering Object (VEO) is powered by SOEKS and DDNA; it is designed to have all the knowledge of the engineering artefact along with the associated experience embedded in it.

VEO provides a standard knowledge representation format and eventually forms various networks of VEOs based on their past manufacturing experience. These networks of VEOs form a part of a bigger Cyber Physical Systems (CPS) umbrella.

3.1. FORMULATION AND ARCHITECTURE OF VEO

A VEO can encapsulate knowledge and experience of every important feature related with an engineering object. This can be achieved by gathering information from six different aspects of an object viz. Characteristics, Functionality, Requirements, Connections, Present State and Experience as illustrated in Fig. 2.

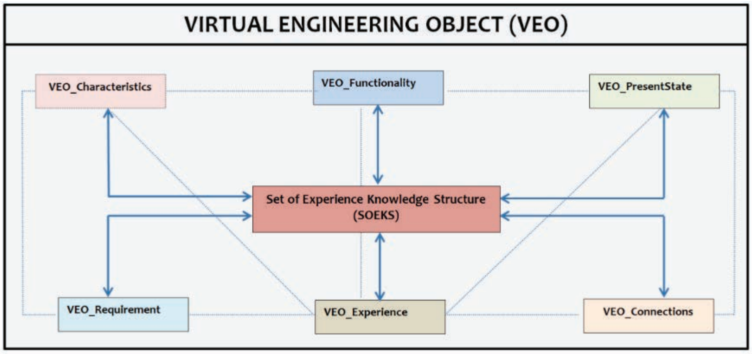


Fig. 2. Proposed VEO Structure [19]

The main features of a VEO (shown in Fig. 2) are as follows:

Characteristics describe the set of physical features and expected benefits offered by the artefact represented by the VEO. Not only the information like its geometry dimensions, appearance, weight etc. are captured in this module but also the possible advantages like "versatility" and the "ease of operation" can also be achieved from this. Knowledge stored in Characteristics assists in better decision making like which VEO is best suited for a given physical condition and also when more than one VEO of a similar kind are available it helps to decide which is the best in the given situation.

Functionality describes the basic working of the VEO and principle on which it accomplishes its operation. Knowledge related with the functioning and operation of an object like the time consumed, its working boundary limits and the outcome of the process that is performed are stored in Functionality. This module of the VEO assists in storing, selecting and reusing the functional/operational details of the object.

Requirements describe the set of necessities of the VEO for its precise working. Information like type and range of the power source, the space required and the extent of user expertise necessary for operating a VEO can be stored.

Connections describe how the VEO is related with other VEOs. Many engineering objects work in conjunction with other objects. These connecting VEOs may be a "part" or may be a "need" of each other. This module of VEO structure is essential for the scaling up and establishing the interconnections of VEOs in manufacturing scenario.

The ***Present State*** of the VEO highlights parameters of the VEO at the current moment. It is like taking a picture and storing information of that particular moment. It also gives an idea about the background of the VEO like its "reliability" and "precision" up till now.

The ***Experience*** of the VEO deals with the knowledge and information which is dynamic in nature, which keeps on changing with each new decision, operation or event. In other words every formal decision related to the VEO is stored in the Experience. This element of the VEO keeps on updating with every activity that is done on the VEO.

3.2. IMPLEMENTATION OF A VEO

For the purpose of implementation of VEO, we integrated it with the Decisional DNA. As discussed in section 2.1, SOEKS consists of Variables, Functions, Constraints and Rules. Moreover in section 3.1, we also discussed that a VEO structure include elements like Characteristics, Functionality, Requirements, Connections, Present State and Experience. SOEKS for each element of the VEO in the system are created individually. The goal behind this was to provide a more scalable setting, similar to the one that would be found in describing a diverse range of engineering objects. Weights are assigned to the attributes of the variables of an artefact, and then the six sets of SOEKS are generated. These individual SOEKS are combined under an umbrella (VEO), representing experience and knowledge.

3.3. DESIGN OF TEST CASE STUDY

As a case study, we considered a manufacturing set up having three different drilling machines, three drilling tools and three work holding devices. Figure 3 shows the framework for the case study, information and specifications about these above mentioned engineering objects were gathered from standard sources and data is stored according to the SOEKS format. Moreover, every formal decision taken is also stored as a SOE, which leads to the formation of interconnected VEO's.

The objective of this study is not only to develop VEO's for engineering artefacts but also demonstrate that different VEO's connect and forms a network. Furthermore to prove that the experience captured from this VEO network can be reused for better future decision making and efficient utilization of resources.

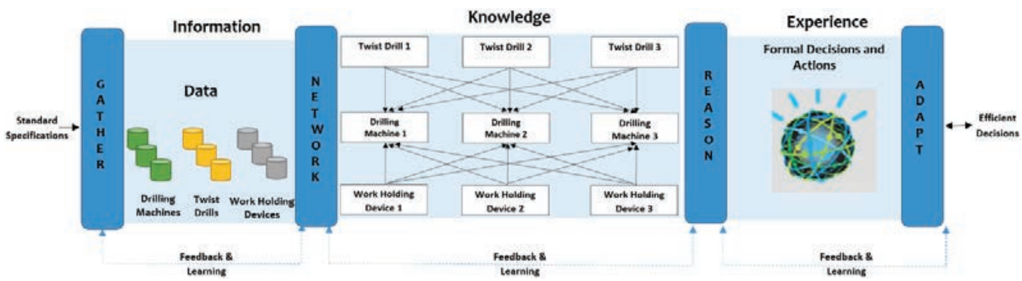


Fig. 3. Framework for the case study

A detailed VEO structure for a drilling machine used in the test case study is discussed and shown in Fig. 4. Effort is made to capture and store all the relevant information of the VEO adhering to the format of the SOEKS.

In the ***Characteristics*** section of drilling machine, VEO physical parameters like area, volume, maximum capacity, manufacturer details, service details are stored.

Furthermore rules are laid to extract knowledge about the VEO like “ease of operation” and “adaptability”. In ***Functionality***, variables related with the functioning of a drilling machine like cutting speed, feed, depth of cut, drilling diameter, drilling depth etc. are defined along with their operational limits. In addition to this knowledge, the outcome of drilling operation like quality of surface finish and machining precision can also be represented in the form of rules. How much Space is required? What and how much power source is required? What kind of expertise of the operator is necessary? All these information for each and every operation can be stored in the ***Requirements*** section of the VEO.

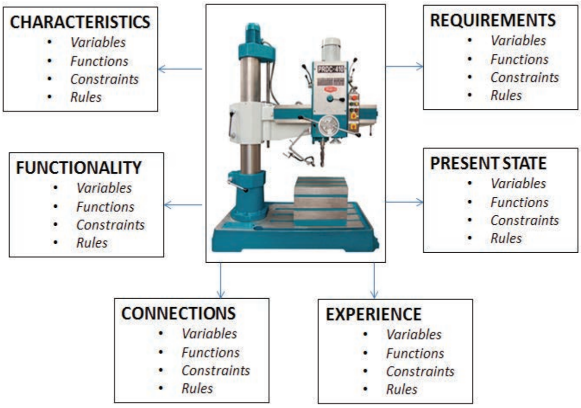


Fig. 4. Test Case VEO Architecture for Drilling Machine

We considered drilling machine, the machining tool, e.g. Twist Drill and the work holding device, e.g. vice, as separate artefacts/VEOs. And Information of these and their relation with main drilling machine VEO are stored in the ***Connections***. In the ***Present State*** not only whether the VEO is free or idle is determined but also knowledge about VEO like its overall reliability and machining precision can be extracted. And, lastly, in the ***Experience*** all the dynamic information related to each operation performed and the formal decisions taken on the drilling machine are stored.

This VEO structure is implemented using JAVA programing language, the reason being, Decisional DNA developed in JAVA has been successfully applied in various other domains. Every Variable [20] is stored as a SOEKS variable. An illustration of a variable (*VEO Name*) stored as a SOEKS variable is as follows:

```
<variable>
<var_name>VEO_Name</var_name>
<var_type>CATEGORICAL</var_type>
<var_cvalue>DM1 </var_cvalue>
```

```
<var_evalue>DM1 </var_evalue>
<unit></unit>
<internal>false</internal>
<weight>0.0</weight>
<l_range >0.0</l_range>
<u_range >0.0</u_range>
<categories>
<category></category>
</categories>
<priority>0.0</priority>
</variable>
```


Six JAVA classes (Characteristics, Functionality, Requirements, Present State, Connections and Experience) for a VEO each having SOEKS Variables, SOEKS Functions, SOEKS Constraints and SOEKS Rules are developed [20]. SOE for each class are stored individually. In a separate class these SOEs are combined to form knowledge and experience repository of an entire VEO. From this knowledge base, manufacturing information related with the VEO can be extracted for future decision making.

Similar to above discussed VEO format of drilling machine, similar structure for twist drills and work holding devices are also developed.

The formal decisions that are taken with regard to the engineering objects are stored adhering to the structure of SOEKS and VEO. Thus, we are able to capture and store information of every operation that is performed and then update the knowledge base of the VEO. The gathered information is effectively and efficiently converted into Decisional DNA structure. The next step is to be able to query the VEO and based on the experience it can predict and suggest options available according to our need.

4. CONCLUSIONS

In this test case study, we presented an approach to represent engineering artefacts based on knowledge and experience. We described the architecture of our approach and implementation that uses SOEKS/DDNA to represent VEO. We demonstrated this approach through some initial tests. As the illustrative result shows, we can model and represent engineering artefact virtually.

We are able to capture and store information of every operation that is performed on the VEO and then update the knowledge base of the VEO. We developed an approach that allows a VEO to capture and reuse its own experiences. The SOEKS and DDNA based VEO proved to be a suitable and comprehensive tool for knowledge discovery. We designed the architecture of our approach and implemented and tested our concepts in the form of a case study. The next step in this research is to apply the

DDNA to the gathered experience of VEO which can be used as an effective prediction tool and can improve manufacturing decision making. Finally, a VEO can be seen as a specialized form of CPS that enables users to make their knowledge shareable, transportable, and easily understandable.

REFERENCES

[1] BAXTER D., GAO J., CASE K., HARDING J., YOUNG R., COCHRANE S. et al., *An engineering design knowledge reuse methodology using process modelling*, International Journal of Research in Engineering Design, 2007, Vol. 18, pp. 37–48.

[2] WANG M.J.F., DEVARAJAN V., *A conceptual approach to managing design resource*, Comput. Ind., 2002, Vol. 47, pp. 169–183.

[3] BAHETI R., GILL H., *Cyber Physical Systems*, Presented at the The Impact of Control Technology, 2011.

[4] LEE E., *Cyber Physical Systems: Design Challenges*, University of California, Berkeley 2008.

[5] LEE E.A., *Cyber-Physical Systems – Are Computing Foundations Adequate?*, Presented at the In Position Paper for NSF Workshop On Cyber-Physical Systems: Research Motivation, Techniques and Roadmap, Austin, TX, 2006.

[6] GARCIA-CRESPO A., RUIZ-MEZCUA B., LOPEZ-CUADRADO J.L., GOMEZ-BERBIS J.M., *Conceptual model for semantic representation of industrial manufacturing processes*, Computers in Industry, 2010, Vol. 61(9), pp. 595–612.

[7] DANILOWICZ C., NGUYEN N.T., *Consensus-Based Partitions in the Space of Ordered Partitions*, Pattern Recognition Letters, 1988, Vol. 21, pp. 269–273.

[8] QIU Y.F., CHUI Y.P., HELANDER M.G., *Cognitive understanding of knowledge processing and modeling in design*, Journal Of Knowledge Managementimited, 2008, Vol. 12, pp. 156–168.

[9] DUONG T.H., NGUYEN N.T., JO G.S., *Constructing and Mining – A Semantic-Based Academic Social Network*, Journal of Intelligent and Fuzzy Systems, 2010, Vol. 21, pp. 197–207.

[10] SHAFIQ S.I., SANIN C., Szczerbicki E., *Set of Experience Knowledge Structure (SOEKS) and Decisional DNA (DDNA): Past, Present and Future*, Cybernetics and Systems, 2014, Vol. 45, pp. 200–215.

[11] SANIN C., SZCZERBICKI E., *Set of Experience: A Knowledge Structure for Formal Decision Events*, Foundations of Control and Management Sciences, 2005, Vol. 3, pp. 95–113.

[12] SANIN C., SZCZERBICKI E., *Using XML for implementing set of experience knowledge structure*, [in:] R. Khosla, R.J. Howlett, L.E. Jain (Eds.), Lecture notes in Artificial Intelligence systems – 9th International Conference KES, Melbourne, Australia, Springer-Verlag, Berlin–Heidelberg, 2005, pp. 946–956.

[13] SANIN C., SZCZERBICKI E., *Extending set of experience knowledge structure into a transportable language extensible markup language*, Cybernetics and Systems, 2006, Vol. 37, pp. 97–117.

[14] SANIN C., SZCZERBICKI E., *Using Set of Experience in the Process of Transforming Information into Knowledge*, International Journal of Enterprise Information Systems, 2006, Vol. 2, pp. 45–62.

[15] SANIN C., SZCZERBICKI E., *Decisional DNA and the Smart Knowledge Management System: A process of transforming information into knowledge*, [in:] A.G. (ed.), *Techniques and Tools for the Design and Implementation of Enterprise Information Systems*, IGI Global, New York 2008, pp. 149–175.

[16] SANIN C., SZCZERBICKI E., *Towards Decisional DNA: Developing Holistic Set of Experience Knowledge Structure*, Foundation of Control and Management Science, 2008, Vol. 9, pp. 109–122.

[17] SANIN C., TORO C., HAOXI Z., SANCHEZ E., SZCZERBICKI E., CARRASCO E. et al., *Decisional DNA: A multi-technology shareable knowledge structure for decisional experience*, Neurocomputing, 2012, Vol. 88, pp. 42–53.

[18] SANIN C., SZCZERBICKI E., *Towards the Construction of Decisional DNA: A Set of Experience Knowledge Structure Java Class within an Ontology System*, Cybernetics and Systems, 2007, Vol. 38, pp. 859–878.

[19] SHAFIQ S.I., SANIN C., SZCZERBICKI E., TORO C., *Decisional DNA based framework for representing Virtual Engineering Objects*, [in:] B.A.N.-T. Nguyen, B. Trawinski, K. Somboonviwat (eds.), *Intelligent Information and Database Systems*. 6th Asian Conference, ACIIDS 2014, Bangkok, Thailand, April 7–9, 2014, Proceedings, Part I, Vol. 8397, Springer, Lecture Notes in Artificial Intelligence, 2014.

[20] SHAFIQ S.I., SANIN C., SZCZERBICKI E., TORO C., *Implementing Virtual Engineering Object (VEO) with the the Set of Experience Knowledge Structure (SOEKS)*, Presented at the 18th Annual KES Conference, Gdynia, Poland, 2014.

typed in Times New Roman, 9 points,
line spacing at least 11 points, italics

Bartosz KUCHARSKI*
Edward SZCZERBICKI**

EXPERIENCE VISUALIZATON

Set of Experience Knowledge Structure (SOEKS) has proved its capability of storing substantial part of experience at a particular decision point. The management of SOEKS record is also proposed in many papers for example referring to concepts of Decisional DNA knowledge representation, that is a collection of a number of SOEKS. The aim of this paper is to focus on visualization and practical aspects of presenting gathered knowledge to the end user.

1. EXPERIENCE RECORD

Experience consists of knowledge of or skill of something or some event gained through involvement in or exposure to that thing or event [1]. A record (also called a compound data) is a data structure to enable to manage the data in order to persist at some decision making state [2]. An Experience record is a data structure to capture information about experience. Depending on intentions and planed use of gathered data the records can focus on one or many aspects in different perspectives. For example they may answer the known from old journalism classes questions called “Five Ws and the H”. They constitute a formula for getting the complete story on a subject [3]. The Five Ws and H applied to experience will determine Who was involved, What happened, When, Where, Why and How. From data gathering perspective the first four answers are relatively easy to capture, but even partial information could be

* Gdańsk University of Technology, Faculty of Management and Economics, Narutowicza 11/12, 80-233 Gdańsk, Poland.
** Gdańsk University of Technology, Faculty of Management and Economics, Narutowicza 11/12, 80-233 Gdańsk, Poland.

useful for certain applications. For example human resources department often uses simplified experience record as shown on Figure 1 whiteout any attention to questions of Why or How.

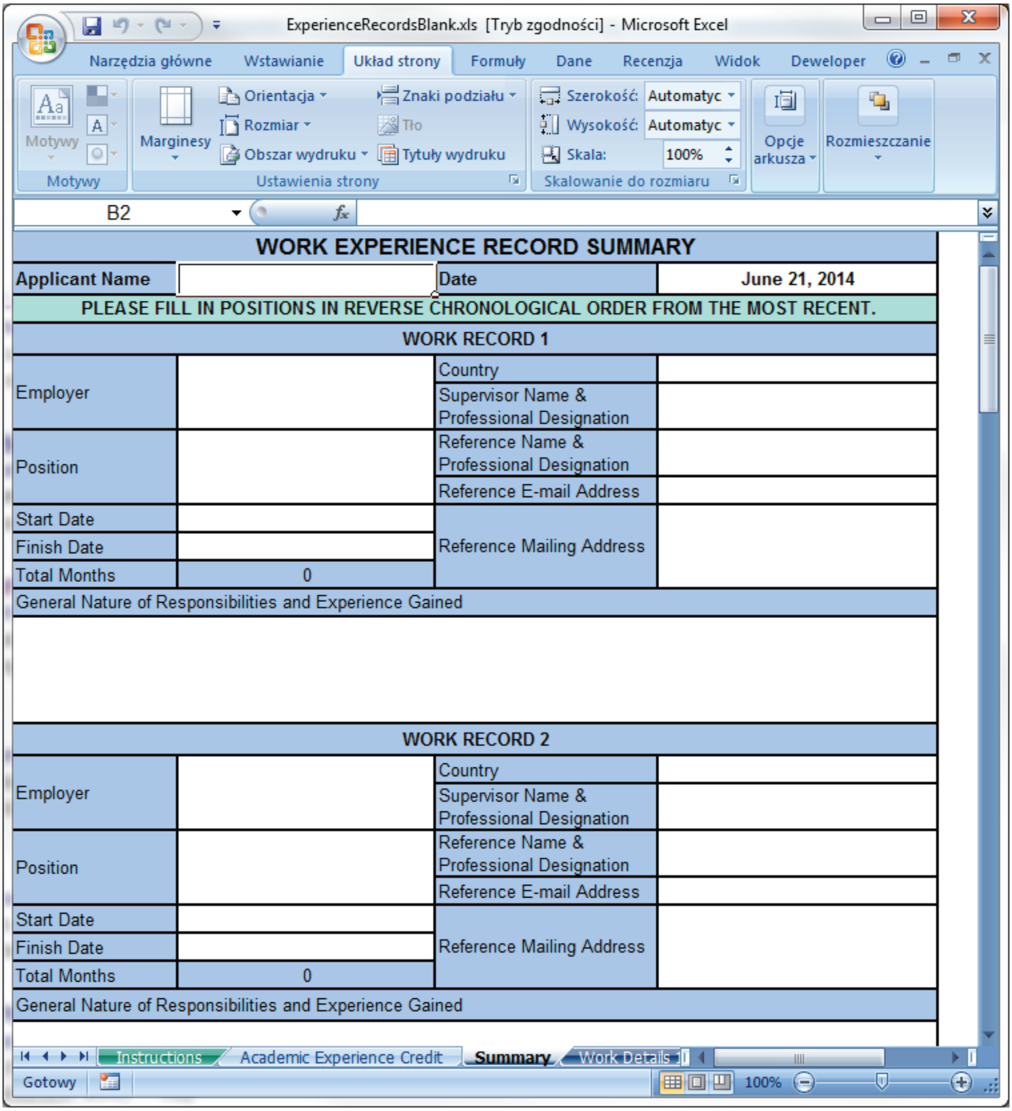


Fig. 1. Blank work experience record form used by Association of Professional Engineers and Geoscientists of Alberta [4]

On the other hand, there are forms that focus mainly on Why or How, for example design patterns defined by “Gang of Four” (Erich Gamma, Richard Helm, Ralph Johnson and John Vlissides) in software engineering [5]. Those patterns are common solutions to common problems with real live examples and expert insights into solution variants. The pattern structure consists of [5]:

- **Pattern Name and Classification:** A descriptive and unique name that helps in identifying and referring to the pattern.
- **Intent:** A description of the goal behind the pattern and the reason for using it.
- **Also Known As:** Other names for the pattern.
- **Motivation (Forces):** A scenario consisting of a problem and a context in which this pattern can be used.
- **Applicability:** Situations in which this pattern is usable; the context for the pattern.
- **Structure:** A graphical representation of the pattern. Class diagrams and Interaction diagrams may be used for this purpose.
- **Participants:** A listing of the classes and objects used in the pattern and their roles in the design.
- **Collaboration:** A description of how classes and objects used in the pattern interact with each other.
- **Consequences:** A description of the results, side effects, and trade offs caused by using the pattern.
- **Implementation:** A description of an implementation of the pattern; the solution part of the pattern.
- **Sample Code:** An illustration of how the pattern can be used in a programming language.
- **Known Uses:** Examples of real usages of the pattern.
- **Related Patterns:** Other patterns that have some relationship with the pattern; discussion of the differences between the pattern and similar patterns.

The above shows that experience could be recorded and generalized into general solutions in which the Who and When that initially was involved is secondary for future use.

Another example of experience record is lessons learned log used to experience sharing. The structure of this record could vary, but the most crucial parts are event, effect, cause, and recommendations. As shown in Figure 2 template for registering lessons learned in projects have some additional information related both to other project documentation like risk log or early warning flag. This experience record is meant to be managed by people and in PRINCE2 project management methodology, lessons learned log should be updated at the end of each stage at a minimum by the project manager [6]. The document contains records about management or quality processes that went well, badly or were lacking, but also abnormal events or notes about performance. There could be also recommendations for enhancements and indication what measurements and accuracy of estimations were used. In fact lessons learned are used not only in project management but also in other areas for example in operations to improve business processes.

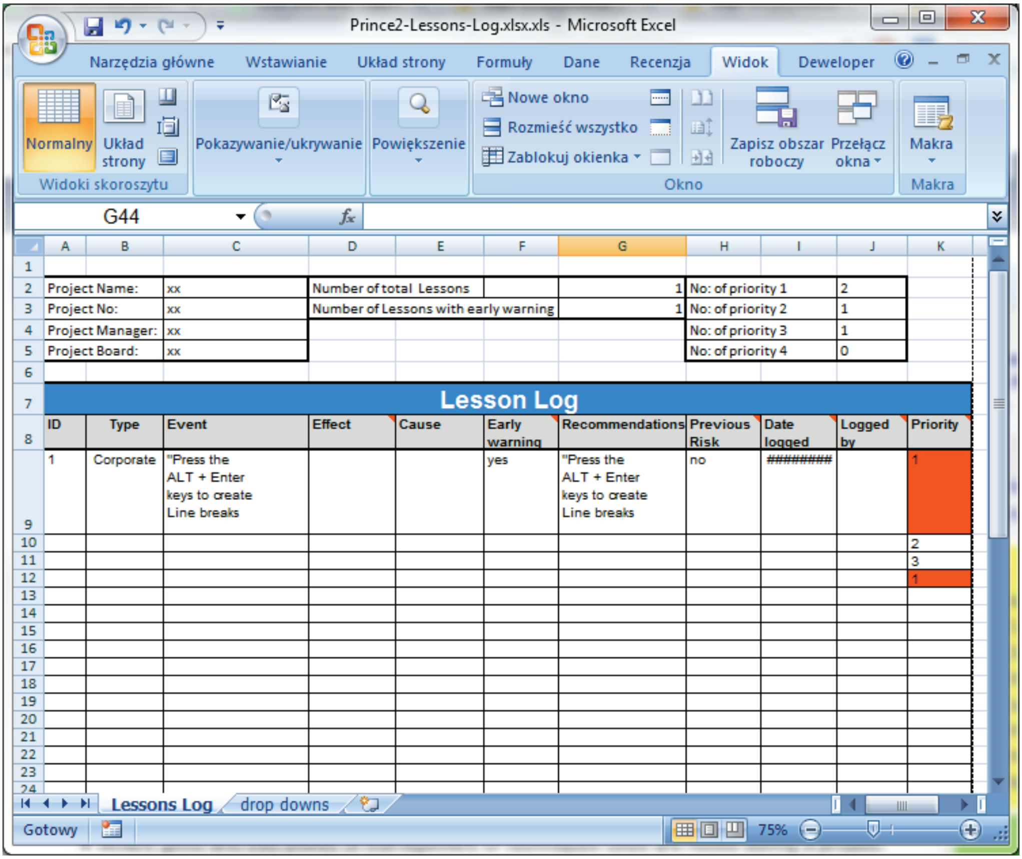


Fig. 2. Blank lesson log template used in PRINCE2 [7]

Another type of experience record is registering decision events in a formal way. The SOEKS structure contains variables to store state or values, functions to perform calculations, rules to define business logic, constrains to codify limitations [8]. The structure is formal, all elements have their notation business rules are similar to IF-THEN structures, variables have defined their types and range. Functions are represented by mathematical notation. All SOEKS elements could be interpreted directly by computer and its native backing representation is eXtended Markup Language (XML) [9].

2. DATA VISUALIZATION

The main goal of data visualization is presentation purposes. It could be done in many ways using plain text techniques or graphical once for presenting facts [10]. The data visualization could also be a tool for data analysis. Depending on analysis type: explorative or confirmative. Explorative analysis seeks data presentation that allows finding hypotheses about the data. Confirmative analysis aims for confirmation or rejection of the hypotheses. The data exploration adds user interaction perspective as shown in Figure 3.

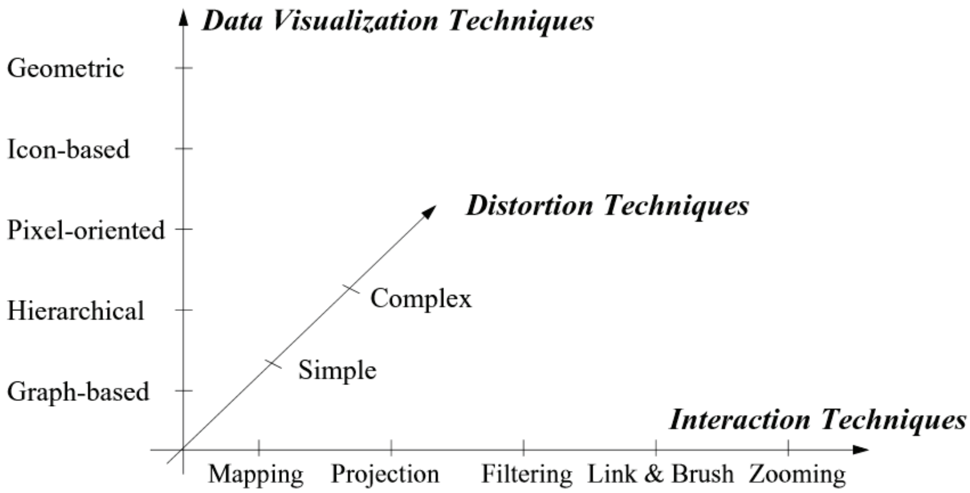


Fig. 3. Dimensions of Exploratory Data Visualizations [11]

Methods and techniques for visualization data can be applied to different levels of information. The basic level is atomic data about fact, when data is categorized and converted into information. There is no direct transformation between information and knowledge, but concepts could be also visualized as well as strategies. There could be a reason to present visualization that compound many levels in one picture. The confirmative or explorative requirements also impact presentation methods. The variety of possible methods confronted by its use were capture on one table by Ralph Lengler and Martin J. Eppler (Figure 4).

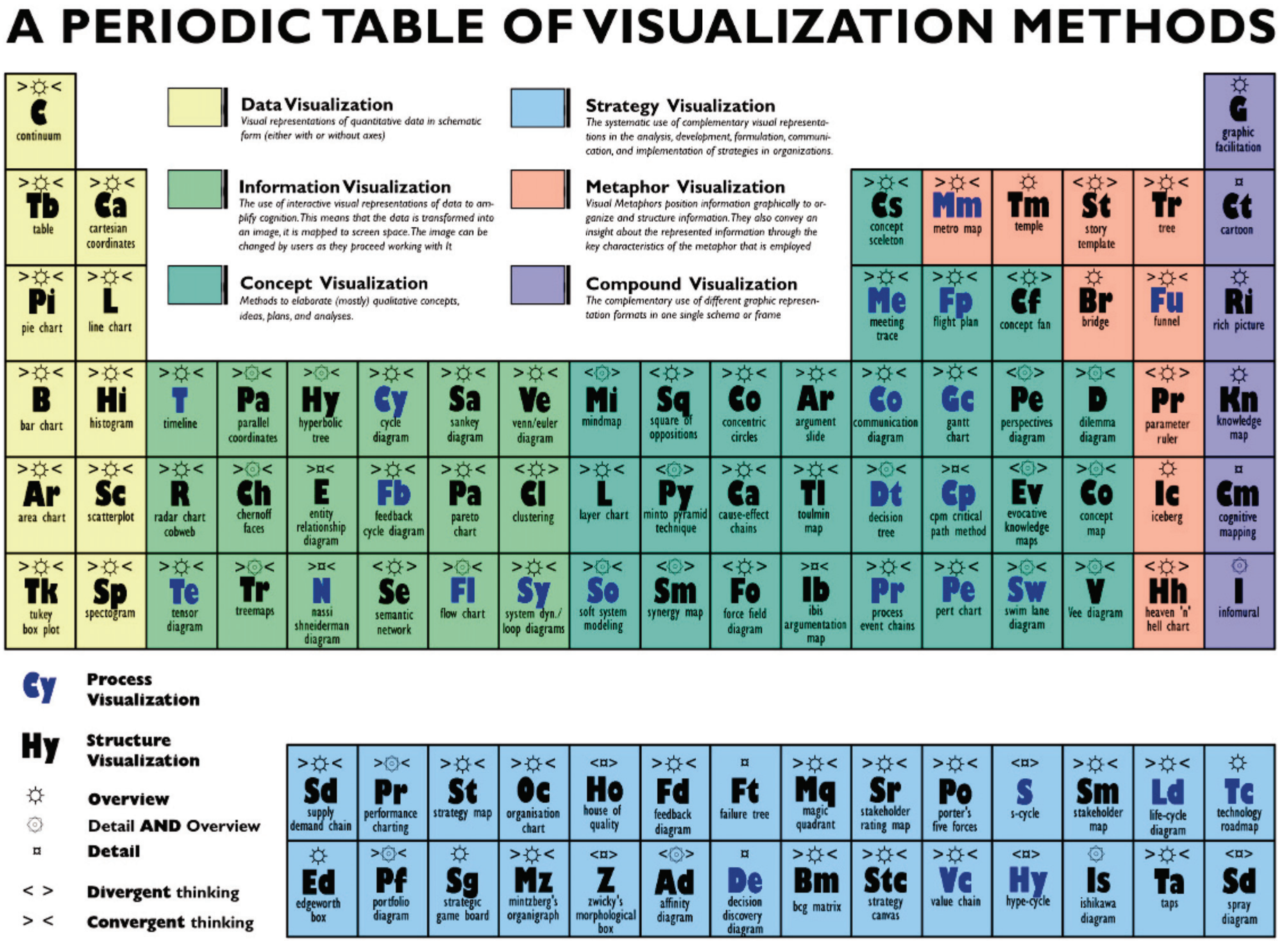


Fig. 4. A periodic table of visualization methods [12]

3. SOEKS VISUALIZATION

SOEKS that is selected as the most formal experience notation could be visualized in many ways using plain text techniques or graphical once for presenting facts. Figure 5 illustrates a perspective diagram technique applied to a particular SOEKS structure.

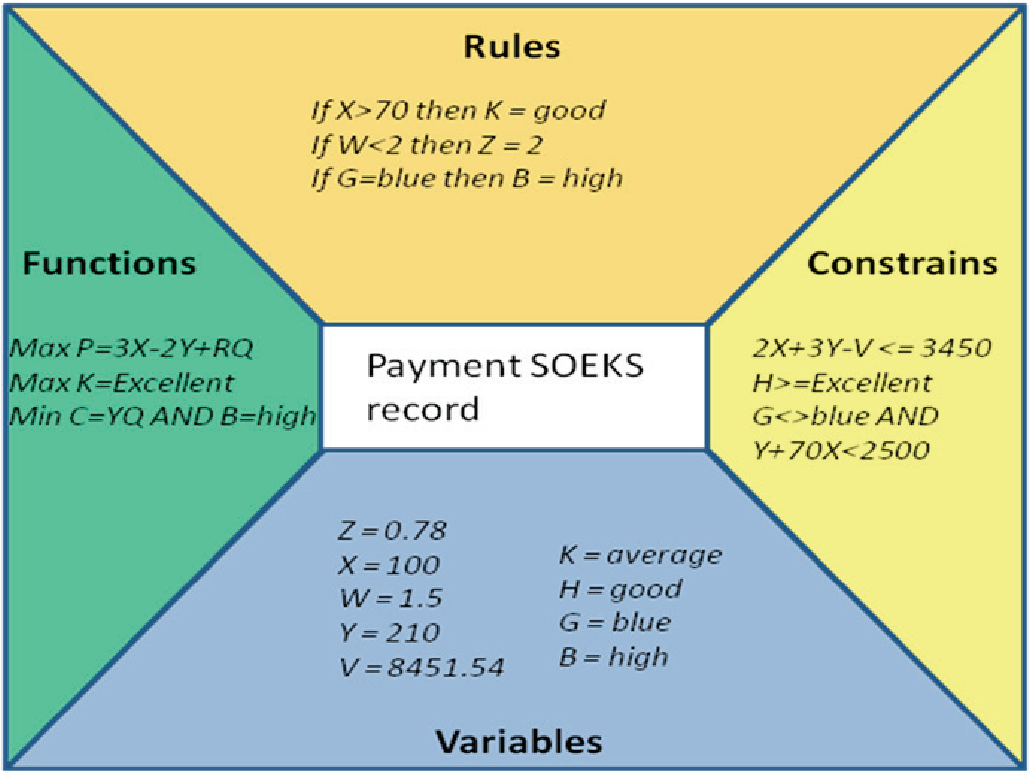


Fig. 5. Perspective diagram based on payment SOEKS record [13]

Perspective diagram is useful especially for presentation of concept and structure, but is limited in terms of user interaction and ability to cope with data.

For exploratory purposes a diagram based on graphs and its connection will be more usefull. Using terminology from the periodic table of visualization methods semantic network will apply. A similar approach was used by KERT (Knowledge Engineering Research Team) at the University of Newcastle to develop DDNA Manager application for managing SOEKS records (Figure 6).

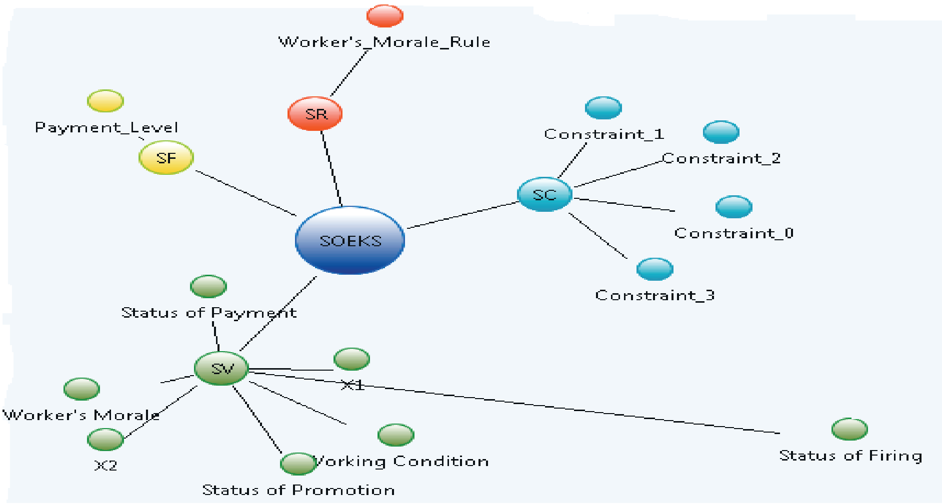


Fig. 6. DDNA Manager visualization of payment SOEKS record

Visualization of one record for complex structure is a challenge, but visualization of a large population of record needs some reference to an intuitive metaphor. In case of SOEKS it is a biological metaphor of Deoxyribonucleic Acid (DNA). Many experience records of similar decision event are grouped into bigger units like chromosomes. It is secondary that real chromosomes do not have to group similar genes, but only metaphor that fits the picture. Figure 7 shows decisional record for an organization in metaphor of tiny fragment of DNA. The idea is to show how stored decision events documents how organization works like DNA how an individual person is build.

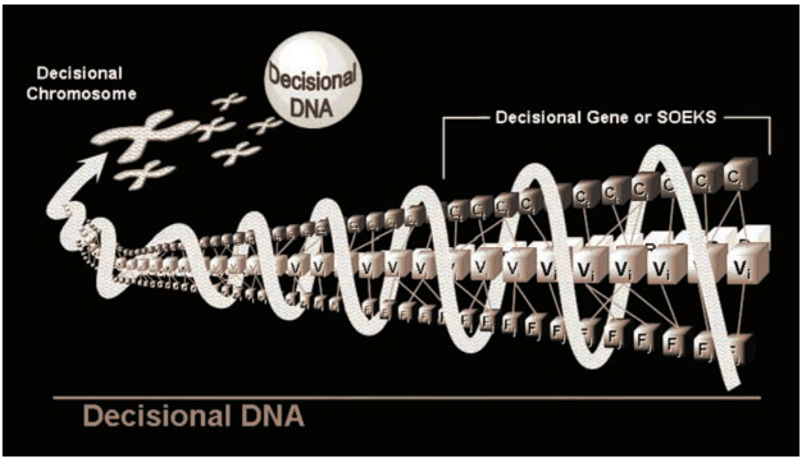


Fig. 7. Visualization of SOEKS database in DNA metaphor [14]

There are also different perspectives that above visualization does not address like for example an end user perspective. SOEKS is meant to be used in decision support system, but how we can express suggestions in graphical user interface that use an operator. The standard tool tips and hints are not enough when there are many relevant experiences. For applying complex information some tree based technique could be used (Figure 8).

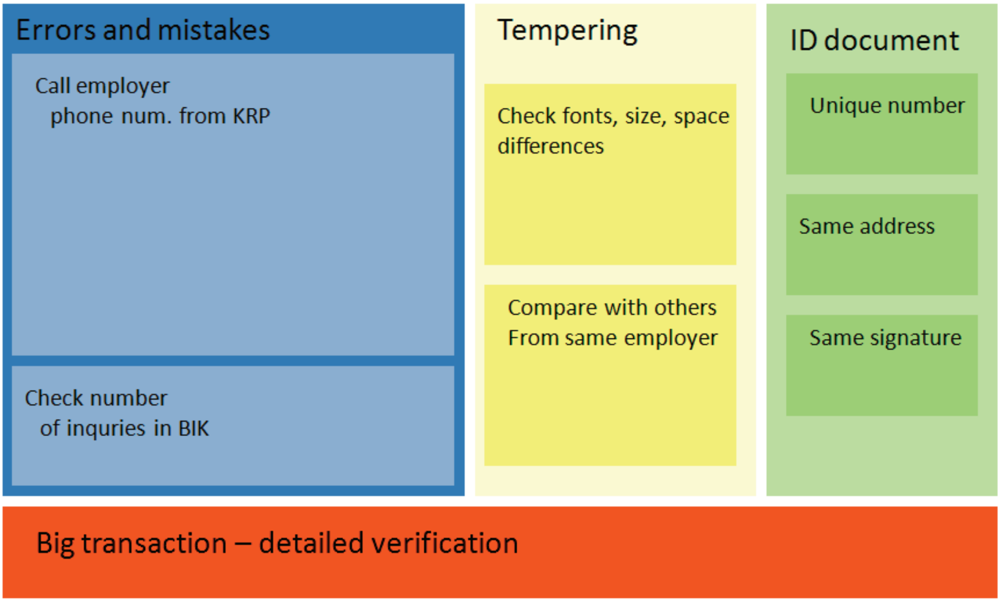


Fig. 8. Treemap visualization for recommended actions from SOEKS based decision support system in application verification task

In Figure 8 blocks refer to suggested action, based on previous experience records. The area refers to probability of fraud detection in simple loan process. Such visualization will be more user friendly then any math like formula, that is hidden in SOEKS functions or formal rule notation.

4. SUMMARY

This paper represents a general and very initial discussion of possible experience visualization that can be used in the process of experience-based knowledge structure representation. It addresses only a few most important aspects of this interesting and complex issue. Especially the aspect of presenting one experience to another person in comprehensive and easy to follow way needs to continued. The idea behind adding User Experience (UX) is a very true challenge in knowledge sharing, if something is complicated and hard to use large audience would not use it. Future research in this area may lead to finding attractive ways of presenting others experience that may popularize experience sharing and use of modern smart decision support systems. The main intention of the Authors of this paper is to use it as the means of fruitful and insightful discussion with expert ISAT community.

REFERENCES

[1] http://en.wikipedia.org/wiki/Experience 30.06.2014.
[2] DATE C. J., *Date on Database: Writings 2000–2006*, Apress, USA, 2006, 90.
[3] Journalism website. Press release: getting the facts straight. Work by Owen Spencer-Thomas, D. Litt, http://www.owenspencer-thomas.com/journalism/media-tips/writing-a-press-release. 30.06.2014.
[4] Work Experience Record revised March 2012, http://www.apega.ca/Applicants/pdf/P_Licensee/experiencerecord-summary.xls, 30.06.2014.
[5] GAMMA E., HELM R., JOHNSON R., VLISSIDES J., *Design Patterns Elements of Reusable Object-Oriented Software*, Addison Wesley Publishing Company, 1995, 6–9.
[6] *Managing Successful Projects with PRINCE2*™, TSO, UK, London 2009, 248.
[7] Lessons Learned Log, http://www.crazycolour.com/prince2/?title=Lessons_Learned_Log 30.06.2014.
[8] MANCILLA L., SANIN C., SZCZERBICKI E., *Smart knowledge sharing platform for e-Decisional Community*, Cybernetics and Systems: An International Journal, 2010, Vol. 41, pp. 17–30.
[9] SANIN C., SZCZERBICKI E., *Using Set of Experience in the Process of Transforming Information into knowledge*, International Journal of Enterprise Information Systems, 2006, Vol. 2 (2), pp. 40–55.
[10] SZCZERBICKI E., WHITE W., *System modelling and simulation for predictive maintenance*, Cybernetics and Systems: An International Journal, 1998, Vol. 29, No. 5, pp. 481–498.
[11] KEIM D.A., *An Introduction to Information Visualization Techniques for Exploring Large Databases*, Institute for Computer Science University of Halle, 2000, 6.
[12] LENGLER R., EPPLER M.J., *Towards A Periodic Table of Visualization Methods for Management*, [in:] GVE '07, Proceedings of the IASTED International Conference on Graphics and Visualization in Engineering, ACTA Press Anaheim, CA, USA, 2007, 88.
[13] SANIN C., SZCZERBICKI E., *Knowledge Supply Chain System: A Conceptual Model*, Knowledge Management: Selected Issues, ed. by A. Szuwarzynski, University Press, Gdańsk 2004, 79–97.
[14] SANIN C., TORO C., HAOXI Z., SANCHEZ E., SZCZERBICKI E., CARRASCOE., PENG W., MANCILLA-AMAYA L., *Decisional DNA: A multi-technology shareable knowledge structure for decisional experience*, Neurocomputing, Vol. 88, 1 July 2012, 42–53.

ICT systems, control, behavior, decision support, active testing, passive testing, supervisory control

Krzysztof M. BRZEZIŃSKI*

RECONSIDERING THE NOTION OF CONTROL: THE CASE OF PASSIVE TESTING

Passive testing is said to be uncontrolled. Although the obvious element of control (applying stimuli) is indeed absent, we argue that passive testing retains various other elements of control. They can fail, leading to false verdicts. Rather than proposing more robust ways of exercising control, we proceed by identifying the tacit elements of control and removing them altogether (so that they can no longer fail), at the cost of employing more complex testing algorithms. We argue that one of such elements of control is influencing, or even only assuming the characteristics of a test arrangement, and that relaxing control is in line with the current state of development of ICT systems. We further concentrate on the control assumption concerned with the placement of a passive tester within a distributed, asynchronous, message-passing system, in which messages in communication links experience delays. This serves as a case study that illustrates a more general approach towards control.

1. INTRODUCTION

The term “control” is so ubiquitous within technical domains (and beyond, e.g., in social and political sciences) that its use borders on *habit*. On the other hand, one does not normally contemplate what its multiple meanings are, and in particular – in which of these meanings the term is used on a particular occasion. This can easily lead to misunderstandings. The dictionary meaning of the term is “*to exercise authoritative or dominating influence over...*; *the power to influence behavior or the course of events*”, which connotes imposing, forcing, applying, stimulating, etc.

Also the notion of *testing* as a controlled activity is well established. Various definitions of testing, mainly of the operational kind, emphasize that its “*principle… is to apply inputs and…*” [1]. This is universally interpreted as the requirement that a tester should *control* an object under test by actively stimulating it, creating predefined circumstances, and soliciting responses. Such notion of *active testing* is then carried over

---

* Politechnika Warszawska, Instytut Telekomunikacji, Nowowiejska 15/19, 00-665 Warszawa.

to the notion of testing in general. On the other hand, there has always been some interest in a related activity, in which a tester does not apply any stimuli. This activity is clearly "uncontrolled" (in the same sense, in which active testing is "controlled"). Although it has been called "*passive testing*" [2], such designation is not universally accepted: if testing is necessarily controlled, then passive testing does not seem to fit the concept of testing at all. This dilemma used to be tentatively resolved by utilizing various euphemisms (such as "observer", "trace checker", passive monitor") [3].

In this paper, we look closer into the notion of control. Passive testing, with its conceptual status described above, suggests itself as a convenient case study. To further fix attention, we deal with ICT (Information and Communication Technology). This covers various information/computing systems, communication networks, and control systems. We concentrate on their behavior, construed as sequences or partial orders of events; this "optics" is known as DES (Discrete Event Systems).

We dissect "control" into two components, one of which is still retained by passive testing. This could help remove the nominal obstacle – in the revised sense, passive testing is *not* uncontrolled. On the other hand, we argue that control, in its various forms, gradually becomes incompatible with the current developments of the ICT technology and its applications. Designing more robust ways of exercising control has thus fundamental limitations. Instead, we propose to remove, as far as possible, the *need* for control. To illustrate this idea, we report on the passive testing approach and algorithm that *accommodates observational infidelity* resulting from the distributed nature of the testing arrangement, without the need for detailed information on the parameters of this arrangement; this amounts to relaxing the control requirements.

2. TESTING – THE BASIC SETTING

Testing is assessing an object, i.e., deciding whether it is "correct" w.r.t. some notion of correctness, based on experiments. The abstract notion of correctness $\mathbb{C}$, against which testing proceeds, may be conceived as a pair: $\mathbb{C}\langle \text{cor}, \text{Ref}\rangle$, where cor is a correctness relation, and Ref is a correctness reference. When a tester is realized, both need to be instantiated. The relation cor usually "disappears" within the design of a tester, and Ref may materialize as: a real behaving object (a reference implementation); a formalized description/model of correct behavior (a reference specification); a black-box entity (an oracle) that answers specific questions about the correctness of particular behaviors, etc. Another option is "operationalizing" (a notion due to Bridgman): transforming both cor and Ref into an imperative test program, more often referred to as a test suite, expressed in a suitable test language (such as TTCN-3).

The basic conceptual arrangement for testing, or a *test ensemble* (Fig. 1a), contains two "sides" in the object-subject relation: the entity in the role of a *tester* or testing system (T), and the entity (system *S*) subjected to test experiments, i.e., in the role of a Tut (*Thing under test*). These entities enter an empirical relation, i.e., communicate. Basing on this communication, and on a predefined notion of correctness, a tester issues a verdict *v* – a symbolic re-statement of the result of testing, which refers to the correctness of an *object of assessment*. In technical contexts this object is normally a *Tut*, but in general this is not necessarily so [4].

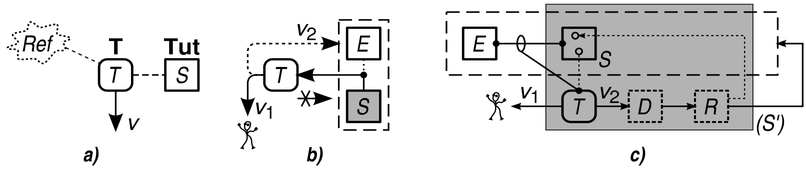


Fig. 1. Components and arrangements of a test ensemble

A verdict *v* is communicated to the users of a test ensemble, to help them decide on some context-dependent course of action. In fact, the sole pragmatic goal of testing is to provide *justification for a decision*. In the canonical case, justification is sought for selecting an alternative, i.e., one of just two mutually exclusive options. Accordingly, in testing there are two conventional verdict values: `PASS` and `FAIL`. They inform that justification was obtained for a generic *positive* resp. *negative* course of action, which may be translated into *accepting* a system that was found correct, resp. *rejecting* a system that was found faulty. The third conventional verdict value: `INC` (Inconclusive) indicates the absence of justification for any particular course of action.

A verdict itself is not, strictly speaking, *the* justification; this justification is hidden in the mechanism of arriving at a given verdict value that was implemented within a tester. It is then tacitly assumed that this mechanism can be *trusted*, and that no "false justifications" will be put forward. However, testing may be deficient – it may generate *false verdicts*: a tester might issue a `FAIL` when the behavior of *S* does not in fact exhibit a failure (which violates the *soundness* [5] or *unbias* [6] property normally required of testing), or it might *not* issue a `Fail` when it could have inferred that such failure did in fact occur (such testing is said to be *not adequate* [5] or *lax*).

False verdicts could be related to measurement uncertainty in metrology [7]. They may lead to suboptimal or wrong decisions, with ensuing damage and/or cost. The significance of such possibility should be judged against the *criticality* of a system being tested (as generally postulated by the dependability community [8]). To fix attention, however, let us assume that a technical goal is to avoid *any* false verdicts.

It is normally assumed that both directions of communication between *T* and *S* are used: for *stimuli* and *responses*. This corresponds to paradigmatic *active* testing. For *passive* testing [9], the communication link through which a tester could apply stimuli is closed (blocked), left unused, or simply absent (Fig. 1*b*). This makes passive testing remarkably different from active testing, and in many aspects – more demanding.

In active testing, *S* unfolds its behavior by communicating with a tester. In passive testing, the only way for a *reactive* system *S* to unfold its behavior is to communicate with some other entities in its environment, jointly depicted in Fig. 1b as one lumped entity *E*. This entity becomes a part of a test ensemble. In basic passive testing, it is common to consider only two verdict values: `FAIL` to indicate the occurrence of a behavioral failure of *S*, and `INC` in any other case. The `INC` verdicts may further be ignored (as non-actionable), or treated as a useful "heartbeat" indication.

Verdicts will normally be conveyed to some external (and usually human) decision-making entity (v_1). Within the design/development process, this entity decides on various corrective loops, or on blocking the shipping of a defective product. In the operational (use) phase of the system's life cycle, the need for continuous testing prevails. Such testing, which mostly uses passive testing techniques, is often referred to as *runtime verification* [10]. Verdicts may then still be conveyed to humans, but it is also possible to feed them back to *S* or to its environment (v_2), to effect a real-time online control loop. Such loops are researched under the tag of *supervisory control* [11], or (in the computer science tradition) *execution monitoring* [12]. To serve as a "controller", a passive testing module (*T* in Fig. 1*c*) needs to be complemented with the diagnosis module (*D*), and the mitigation module that produces appropriate reaction (*R*) [13]. In particular, this supervision machinery might be *embedded* in *S*, which then becomes a system S', with a set of new "self-*" properties (self-managing, self-healing). In this case, certain "non-standard" means of accessing *S* might be possible (as depicted by hollow circles), but most testing approaches discourage the use of such mechanisms, concentrating on the *external behavior* of a system.

In the sequel, what matters is the "quality" of verdicts produced by *T*, regardless of how these verdicts are further used. Accordingly, we only consider control of *testing itself*, not supervisory control that might internally use testing techniques.

3. CONTROL RECONSIDERED

Let us dissect the notion of control, as applied to testing itself, into two aspects:

- controlling, i.e., actively stimulating a system *S* by a tester, to solicit (enforce) its "interesting" behaviors; this is the defining element of *active* testing, and also the paradigmatic understanding of testing in general;
- having the "power to influence" various *other* circumstances of testing, which include the arrangement and characteristics of a test setup.

We propose that "controlling" in the latter sense should also include *knowing* and *assuming*. Having control over the test setup would then mean: purposely and directly setting up the components of a test architecture, but also: knowing/believing/assuming that this setup is as intended. Pragmatically different notions of "assuming" are often confused. In *engineering*, to assume some characteristics normally means to know (i.e., to be justified in believing) that these characteristics hold, with the underlying "meta-assumption" that all the necessary arrangements have been undertaken to this effect, and need not be further cared about. When this assumption fails, a technical operation in question may proceed in unexpected ways. Failure of *such* control over the test setup does not differ in any significant way from failing to "personally" set up the test ensemble as intended. According to function theorists, a designated function that cannot fail is void, and thus not interesting [14]. Therefore, the notion of control cannot rely on its being absolutely infallible. On the other hand, *logical* assumptions (i.e., premises) have little to do with belief or trust, and their violation (falsity) may result, e.g., in vacuous truth. It is known that testing is meaningful only under certain assumptions of *this* kind about the structure and/or behavior of a *Tut*, called *test hypotheses* [6]: one posits some properties of a *Tut*, in order to infer its other properties.

Testing used to be applied to ICT systems mostly in circumstances in which it can be controlled in *both* senses. In general, more control translates into testing algorithms that are simpler, and thus cheaper. As one of many optimization criteria, it is natural to employ the cheapest possible algorithm that would still be adequate (i.e., in our case, that would *not issue false verdicts*) under a given level of control. Refraining from using all the available control would be a self-imposed handicap. Justification for such understanding, however, tends to erode. Traditional ideas about testing being controlled, philosophically anchored in David Hume's reductionism, are becoming incompatible with the developments in ICT technology. For example, current systems-of-systems are not built at once, so there is no notion of their being "finished", and no clear point at which they might be factory-tested in a lab; they exhibit emergent (unanticipated) behavior; they may evolve at run-time; their parts or subsystems may not allow certain patterns of external access (e.g., for legal or security-related reasons); and they may also have no fixed geometry, which bears on delay properties (in particular, when components move). Adjusting the concepts and techniques of testing to these changing circumstances is inevitable. Various aspects of control will then become inconvenient, impossible to realize, or incompatible with other features of a system. By fully surrendering control of the first kind, passive testing directly addresses a number of these concerns. It still, however, remains controlled in the second sense, which makes calling it "uncontrolled" unwarranted. In fact, generic passive testing algorithms are based on *very strict assumptions*, and thus require tight control regarding the characteristics of the test setup – much tighter than in active testing.

The particular problem of control of the second kind on which we further focus concerns *delays in a test setup*. Its dual formulation is: (*a*) *placing* a passive tester at a particular position within a test ensemble (which may not succeed, or the very notion of a "particular position" may be elusive), and (*b*) *assuming* that a tester is currently positioned within a particular area (but it may have already "wandered" beyond that area). For fundamental reasons, the physical positioning of a tester within a test setup bears on *delays* – a combination of propagation delays and various lumped delays additionally introduced by "invisible" (abstracted out) elements, such as buffers.

Designing more robust ways of exercising control over delays has fundamental limitations. Therefore, we propose to remove (or considerably reduce) the *need* for such control, so that in a particular instance of testing a tester would not "accidentally" violate the delay-related assumptions.

4. THE CONTROL PROBLEM – INFIDELITY OF OBSERVATIONS

When conceived as a distinct entity in a distributed system, a (passive) tester can only perceive another entity *S* through a distorted image of its behavior; this fact is captured by the notion of *observational infidelity*. Infidelity factors include losing, adding, reordering, changing, and delaying the observations; we further deal only with delays. Basing on such distorted observations, a tester may issue false verdicts.

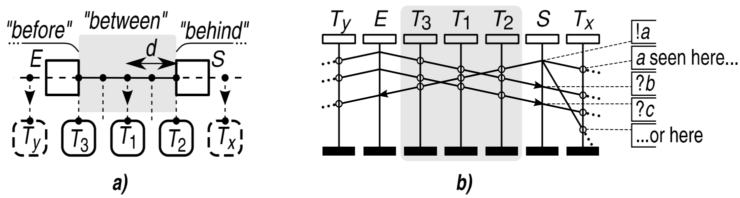


Fig. 2. Delay phenomena in a test architecture

Figure 2a expresses in more detail the idea that a passive tester *T* "spies" on a communication link connecting *E* and *S* – it observes the behavior of *S* by "tapping" the link, and registering messages travelling to and from *S* (note, however, that *T* is *not* a proxy between the communicating system elements). In this view, the "tap" is assumed to be perfect and delayless. It transpires that *T* must be attached somewhere between *E* and *S*. This common *betweenness assumption* translates to attaching a tester *T* (via a tap) at a time-distance d_{T1} from *S*, where $0 \leq d_{T1} \leq d_E = d_{\max}$. The corresponding requirement on a tester is that its operation anywhere within the "betweenness" area should be *sound*. In other words – a tester should not issue a `FAIL` verdict for a correct logical behavior of *S* (i.e., behavior that manifests itself at the boundary of *S* as a correct sequence of events), for *any* observation of this behavior that is possible within the "betweenness" area. On the other hand, it is known that finite delays alone may result in many different observations of the behavior of *S*. A combination of *message crossovers* is shown in the MSC (Message Sequence Chart) diagram in Fig. 2b. Three events that constitute the behavior of *S* are perceived by the testers placed within the (shadowed) betweenness area in three different orders.

One of them (that of T_2) corresponds directly to the behavior of *S*, but others do not; this is how infidelity manifests itself. A given *T* is unable to autonomously decide if the order of its observations is, or is not the "direct" and "true" one. If the sequence of events at *S* is "objectively correct", then, for soundness, each of the considered testers: T_1, T_2, and T_3 is required *not* to issue a `FAIL` verdict. Note the circles on the instance axes of testers; they are non-standard for MSC, and denote eavesdropping of messages.

The betweenness assumption appears to be the weakest control element (of the second kind) routinely applied to passive testing. This assumption may be *strengthened* (which allows *weaker* and *cheaper* testing algorithms), by requiring that the actual value of *d* lie within a known sub-range of values between 0 and d_E, or even be known precisely. It would then be possible to decide *a priori* if a particular tester is appropriate, or to "tune" its algorithm accordingly. As a special case, $d_T = 0$ corresponds to *co-locating* a tester (T_2), which, in popular (but misguided) interpretation, allows a tester to directly perceive the behavior of *S* as it happens, with no infidelity.

The betweenness assumption might also be further *weakened*, which would require *stronger* testing algorithms. As can be seen in Fig. 2*a* and *b*, there are additional areas in which testers could possibly be positioned. Tester T_x is placed "after" (or behind) *S*, and tester T_y – "before" *E*. We would then have $d_{Tx} < 0$ or $d_{Ty} > d_{\max}$, respectively. A part of the reasoning embedded in testing algorithms developed under the betweenness assumption is: "...message from *E*, observed (recorded) by *T*, has not been received by *S* yet". In case of T_x this does not hold, as "...message from *E*, observed (recorded) by *T*, must have been already received by *S*".

Now, the aim is to consider the possibilities of relaxing control (of the second kind) over passive testing, so that a tester might "land anywhere", and remain *sound*.

There are loose ends in the foregoing discussion. It remains to be established what an *observation* is, *how* a tester observes, and *what* it can observe. For this, formal models would have to be used, which is a main premise of *model-based testing* [15]. We have developed two such complementary models: an architectural model (AM), and a behavioral model (BM). Due to space limitations, and in order not to burden this predominantly conceptual discussion with detail, we only provide their brief synopsis.

Architectural models of a distributed system, well suited to active testing, were found in [16] to be inadequate, or even inconsistent, when applied to passive testing. A very simple architectural model was then proposed, in which a system is construed as a family of entities *A*, each with exactly two ports: one output port *A*! and one input port *A*? (notation reminiscent of the TTCN-2 test language), which may be connected by unidirectional point-to-point FIFO channels, in configurations depicted in Fig. 3a and b. The application of this model to a test ensemble is shown in Fig. 3c. A single passive tester *T* uses its single input port to input a merged stream of signals sent by *S* towards *E*, and those sent by *E* towards *S*; the "copies" of these signals travel towards a tester in channels c_1 and c_2, respectively. By suitably adjusting the delay of these channels, the position of T may be perceived as "between E and S". One or both of these channels may, however, be arbitrarily long, as in Fig. 3d. The betweenness area is thus no longer distinguished topologically, or favored in any way. Also a "tap" disappears – there is no need, and no possibility, of representing it directly.

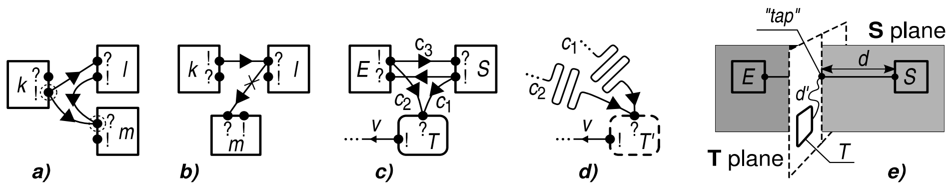


Fig. 3. The architectural model of a system

In Figure 3e a test architecture is presented as the intersection of two planes: the system plane S, and the T plane that contains any "testing machinery". A testing system will normally be itself a distributed system. It is thus natural to expect additional delay d' within the T plane. In our model, this additional delay is already included in the delay properties of channels c_1 and c_2. This also shows that the "co-location" case is elusive. For d' to disappear (to be exactly 0), co-location would have to be taken literally: the roles of a tester and a Tut would have to be played by *the same* entity, which is not normally considered.

As a behavioral model, we use a class of Behavior Machines – simple structures akin to Finite State Machines (FSM) and Labelled Transition Systems (LTS). There is nothing particularly new in their formulation: Z is a finite set of states; states are connected by transitions labelled with output events, input events, and silent events. BMs are used in two capacities: to consistently model the behavior of entities in a test ensemble (including a tester), and as a reference model of the correct behavior of a *Tut*. This model is accessed by a testing algorithm during the operation of a tester.

It may now be stated that, at the model level, a passive tester *makes an observation* when it fires a transition labelled by a respective input event. The behavior machine of T is *trivial*, and almost entirely independent of any particular testing algorithm. It contains a loop for observing a signal, with the optional output transition for the `FAIL` verdict to be presented at the output port. The whole passive testing algorithm is hidden behind the decision to fire, or not to fire this output transition.

Let B be a reference Behavior Machine, against which the behavior of S is assessed. The essence of a large class of passive testing algorithms that directly use B as a *Ref* (i.e., unlike in active testing, that do not involve the intermediate stage of deriving, or generating, a program-like test suite) is to hypothesize about a state in which S may possibly be, according to observations performed so far. As it may not be possible to identify with certainty any such single state, a tester's hypothesis is a set H of states from B. The operation of testers of this class consists in updating H after each observation. Two border cases are: $H = Z$ ("nothing is known about the state of *S*"), and $H = \{\}$ ("according to a *Ref*, there is no state in which *S* can possibly be"). In the latter case, a tester issues a `FAIL`. A generic passive testing algorithm of this kind, which takes observations at face value, and uses them directly to drive the evolution of *H*, may be called SHE (Simple Hypothesis Elimination). It is attributed to [2], although the idea had been known earlier.

SHE clearly does not tolerate infidelity effects; it is only effective in the "co-located" case, which makes it practically useless. Two generic ways of dealing with delay-related infidelity of observations are: (a) keeping SHE and equipping it with an additional, separate *preprocessing module*, which suitably reorders events so that they would be observed in their "right" order, and (b) extending the passive testing algorithm itself, so that it would deal with infidelity internally, at the cost of its complexity. The former approach, conceptually similar to *signal conditioning* widely applied in metrology, was found to be very limited. The generic idea of the latter approach is to keep a particular signal, once observed, within the "processing horizon" of the algorithm, until a tester is certain that this signal, or any intertwining of this signal with consecutive observations, can no longer influence the set *H*, and thus – the verdict. This idea forks into two sub-classes, depending on how the horizon is understood: as an event-window (defined by the number of signals to be considered jointly), or as a time-window (defined by the time, during which an input signal cannot be discarded). Complex algorithms of the general "hypothesis elimination" class can thus be classified into EWE (Event-Window hypothesis Elimination), and TWE (Time-Window hypothesis Elimination).

Passive testing algorithms that are more complex than SHE (and thus tolerate weaker control of the second kind) routinely assume the processing horizon that corresponds to the betweenness area. This is also the case for the early algorithm of the TWE class by Wvong [17], which we analyzed and implemented to gain more insight. We have been able to extend it so that it remains sound also beyond the betweenness area, which demonstrates the feasibility of the idea of relaxing control over testing.

5. CONCLUDING REMARKS

To restate, two main points of this paper are: that there is more to "control" than just issuing stimuli, and that there may be important reasons for *relaxing* control, rather than trying to strengthen and perfect it. Note, however, that problems of control re-appear (fractal-like) in various functional parts and at various levels of abstraction, and there is apparently no single, universal guideline for dealing with them all.

The idea that assumptions made in engineering contexts are a kind of control seems novel. The observation that various deficiencies of testing might be attributed to failed assumptions was made earlier (for active testing) in [18], but was not further systematically discussed.

The early stages of the present work were conducted with the cooperation of Michał Goliński, who later independently provided his own, informal account of the Wvong's algorithm and its proposed extensions [19].

REFERENCES

[1] CACCIARI L., RAFIQ O., *Controllability and observability in distributed testing*, Information and Software Technology, 1999, Vol. 41, No. 11–12, 767–780.

[2] LEE D., NETRAVALI A., SABNANI K., SUGLA B., JOHN A., *Passive Testing and Applications to Network Management*, [in:] ICNP, 1997, 113–122.

[3] BRZEZIŃSKI K., *Active – Passive: On Preconceptions of Testing*, Journal of Telecommunications and Information Technology, 2011, 63–73.

[4] BRZEZIŃSKI K., GUMIENIAK A., JANKOWSKI P., *Passive Testing for Reverse Engineering: Specification Recovery*, [in:] PDCN, 2008, 27–32.

[5] NETRAVALI A., SABNANI K., VISWANATHAN R., *Correct Passive Testing Algorithms and Complete Fault Coverage*, [in:] FORTE, 2003, 303–318.

[6] GAUDEL M.-C., *Testing Can Be Formal, Too*, [in:] TAPSOFT, 1995, 82–96.

[7] BRZEZIŃSKI K., *Pomiary a testowanie: przypadek biernego testowania zachowania systemu reaktywnego*, [in:] PPM, 2014, 163–166.

[8] AVIZIENIS A., LAPRIE J.-C., RANDELL B., LANDWEHR C., *Basic Concepts and Taxonomy of Dependable and Secure Computing*, IEEE Trans. Dependable and Secure Computing, 2004, Vol. 1, No. 1, 11–33.

[9] BRZEZIŃSKI, K., *Towards the Methodological Harmonization of Passive Testing Across ICT Communities*, [in:] Engineering the Computer Science and IT, Soomro S. (Ed.), In-Tech, 2009, 143–168.

[10] CHUPILKO M., KAMKIN A., *Runtime Verification Based on Executable Models: On-the-Fly Matching of Timed Traces*, [in:] MBT, 2013, 67–81.

[11] RAMADGE P., WONHAM W., *The Control of Discrete Event Systems*, Proc. IEEE, 1989, Vol. 77, No. 1, 81–98.

[12] SCHNEIDER F., *Enforceable Security Policies*, ACM Trans. Inf. Syst. Secur., 2000, Vol. 3, No. 1.

[13] LEUCKER M., SCHALLHART C., *A brief account of runtime verification*, J. Logic and Algebraic Programming, 2009, Vol. 78, 293–303.

[14] MILLIKAN R., *In Defence of Proper Functions*, Philosophy of Science, 1989, Vol. 56, 288–302.

[15] UTTING M., PRETSCHNER A., LEGEARD B., *A taxonomy of model-based testing*, Univ. of Waikato, 2006.

[16] BRZEZIŃSKI K., *On Embeddable Passive Testing*, [in:] IFAC PDeS, 2010, 113–118.

[17] WVONG R., *New Methodology for OSI Conformance Testing Based on Trace Analysis*, Master's Thesis, Univ. British Columbia, 1990.

[18] JARD C., JERON T., MOREL P., *Verification of Test Suites*, [in:] TestCom, 2000, 3–18.

[19] GOLIŃSKI M., *Algorytmy i architektury testowania biernego*, Warsaw University of Technology, 2012.

consistency, business process, process modelling,
software architecture, BPMN

Stanisław Jerzy NIEPOSTYN*

THE SUFFICIENT CRITERIA FOR CONSISTENT MODELLING OF THE USE CASE REALIZATION DIAGRAMS WITH A BPMN DIAGRAM

BPMN diagrams are more and more often used to visualise scenarios in use case driven IT projects. The verification of consistency of BPMN diagrams is subsequently needed to identify errors in requirements at the early stage of the development process. The consistency verification is challenging due to a semi-formal nature of BPMN diagrams. We examine the BPMN diagram to workable in automatic verification of consistency of scenarios of the visualized use cases. Moreover we consider whether the BPMN diagram enables simultaneous modelling of the functionality, of the structure and of the behaviour of the target system model. Those properties should enable to develop consistent and complete models and thus to generate automatically complete workflow applications without any manual programming. In this paper we examine how BPMN diagram fulfils those properties.

1. INTRODUCTION

In object-oriented software development, the UML notation has become the major standard for the software architecture modelling at different stages of the life cycle and at different views of the software system, including the requirements specification. Thus in the majority of projects using UML diagrams [1], use case diagrams are developed at the beginning of software development to describe the main functions of the software-based system. The class diagrams are created then to show the structure of the system while, state machine diagrams are drafted to show the behaviour of system elements ([2]). Subsequently BPMN diagram can be used to verify consistency of other diagrams. This kind of diagram can be used to visualize scenarios as so called a “use case realization diagram”. Notably, the BPMN diagram has not been formally

* Institute of Computer Science, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland.

integrated either with the class diagram or with the state machine diagram. In this paper we consider to reverse these activities with the BPMN diagram: firstly we develop the BPMN diagram, then we check if adequate class, use case and state machine diagrams could be derived.

An early consistency check of the use case realization diagram seems crucial for the consistency and completeness of the software architecture, but proves to be challenging due to the informal nature of activity specifications. By the sufficient criteria for consistent modelling of an use case realization diagram we mean that: all flow paths of an use case realization diagram can be performed; a diagram describing the structure aspect of the system can be generated; a diagram describing the behaviour aspect of the system can be generated; all elements of the use case realization diagram are mapped onto generated diagrams.

We examine whether the BPMN diagram fulfils above criteria. The aim of the consistency analysis is to validate that all flows are connected. Next we propose to check that each activity and each instance of object in our activity diagram has link with each element of the subsequent generated UML diagrams.

The object pseudo-code can be used to formalize this problem and to provide a tool to support the analysis. The idea with Z formalisation was presented in [3] to keep the consistency and the satisfactory completeness between class, state machine and use case diagrams based on DCD Diagram. This idea was extended in [4] for class, state machine and use case diagrams based on FSB UML diagram. Such object pseudo-code may be easily implemented in Java coding tools.

Based on our previous work cited above, this article presents our analysis of the workability of the BPMN diagram to keep sufficient criteria for consistent modelling and to enable automatic generation of complete workflow applications without any manual programming. In order to verify the proposed criteria we provide a semantics for the BPMN diagram with the object pseudo-code.

The automated checking of rule sequence applicability has been integrated into the Topcased tool and published on the site http://www.project-media.pl/bpmn2.php.

The paper is organized as follows: in section 2 the related works on consistency checking of UML diagrams are described, then completeness and types of inconsistencies are described. In section 3 the rule-based method for consistent modelling of the use case realization diagrams with a BPMN diagram is described. The rationale of applying the BPMN diagram to derive the complete and consistent UML model is given in section 4. Section 5 concludes the paper.

2. RELATED WORKS

Different software models describe the same system from different points of view, at different levels of abstraction and granularity, possibly in different notations. They may represent the perspectives and goals of different stakeholders. Usually some inconsistencies between models are arising [5]. Inconsistencies in models reveal design problems. If these problems are detected at the early stages of the design, costs of fixing them are much lower than if they are detected at later stages of software design and implementation.

In software development, the BPMN has become the standard notation for the business processes modelling. Wong and Gibbons [6] check the consistency of business process models at different levels of abstraction. Dijkman et al. [7] check BPMN diagram in term of "completion of a path". Weidlich and Mendling [8] analyse behavioural consistency (as the absence of contradictions) between process models. Ways of identification of consistency problems for processes modelled with BPMN were proposed by Andreas Rogge-Solti et el. [9]. Authors check consistency of data dependencies and make sure that no data anomalies will be manifested through.

We propose to analyse the consistency of the BPMN diagram in a similar way as for the consistency the UML diagrams. As shown in [10], there are several methods to verify consistency in UML diagrams: meta-model-based method, graph-based method, scenario-based method, constraint-based methods and knowledge-based methods. We are focusing here on the constraint-based methods and on graph-based methods.

Our research focuses attention to the consistency of class, state machine, and use case UML diagrams derived from the BPMN diagram. We consider the sufficient criteria for consistent modelling of use case realization diagram to generate consistent class and state machine diagrams. Specifically the method we propose here is the use case realization diagram based on the BPMN standard.

3. CRITERIA OF CONSISTENCY

According to Functional-Structure-Behaviour (FSB) framework introduced by John Gero [11] the purpose of the design description is to transfer sufficient information about target system so it can be constructed. The description must at least enable to incorporate a function, a structure, and a behaviour of the target system. Therefore the development of software in which one cannot take into account these three dimensions are "doomed to fail". Truyen [12] described a model, in major MDA concepts, as a formal specification of the function, structure and behaviour of a system. He claims that model must be represented by a combination of UML diagrams. Spanoudakis and Zisman [5] described this as a situation, in which model inconsistencies may arise.

In this section we explain informally the model consistency, which we subsequently apply in the modelling of the use case realization diagram. Then we present our concept of the dimensions of the software architecture which form the consistent description of software architecture.

3.1. MODEL INCONSISTENCIES

To assert that something is consistent we have to declare what it is consistent with. Software models describe a system from different points of view, at different levels of abstraction and granularity, in different notations. They may represent viewpoints and goals of different stakeholders. Usually inconsistencies between diagrams are arising. Inconsistencies reveal design problems. The roots of consistency can be found in formal methods. The research on consistency models was outlined by Finkelstein [13]. Finkelstein stated, that inconsistency is not necessarily a bad thing, and should be evaluated after the translation of the model specification into a formal logic. UML is not a formal language so often UML models are translated into more formal notation. In UML inconsistency between UML diagrams [14] are studied. Inconsistencies arise because some models are overlapping [5].

We think that the use case realization model should be created firstly and should be already in a consistent form before supporting logical rules for creation next models. Such approach should prevent complicated and complex detection of inconsistencies during the model construction.

3.2. THE SUFFICIENT CRITERIA FOR CONSISTENT UML ACTIVITY DIAGRAM

In the majority of projects using UML diagrams [1], use case diagrams are developed at the beginning of software development to describe the main functions of the software-based system. Then class diagrams are created to show the structure of the system, and state machine diagrams are built to show the behaviour of system's elements ([2]). Subsequently activity or sequence diagram can be used in order to verify consistency of other diagrams. These diagrams are also using visualizing scenarios i.e. – use case realization diagrams.

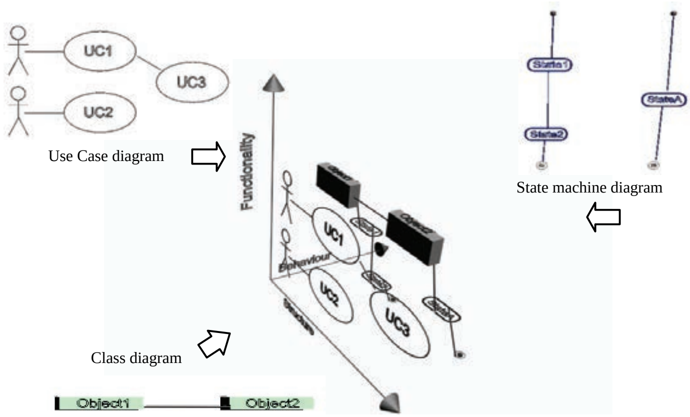


Fig. 1. Three dimensions of the software architecture view

Activity diagram enables to associate activities with objects (instantiate classes), and use-cases ([2]). It can be noticed that Use Case, Class and State Machine diagrams are orthogonal (Fig. 1), and enable to derive use case realization diagram. A model, which adequately integrates these diagrams thus enables to keep the consistency and the sufficient completeness of the whole system because these three diagrams do not have any common element.

Activity diagram must have all flow paths connected. It means that all flow of an activity diagram can be performed. Activity diagram may compare as the directed graph with connected vertices. A graph is said to be connected if every pair of vertices in the graph is connected.

Considering the criteria mentioned above, we define sufficient criteria for consistent Activity Diagram in the following conditions:

– the Activity Diagram is a connected graph;
– the Activity Diagram describes the structure aspect of system;
– the Activity Diagram describes the behaviour aspect of system;
– sufficiently consistent activity diagram enables to create subsequent class and state machine diagrams;
– all elements of the Activity Diagram must be mapped onto generated class and state machine diagrams.

4. CONSISTENCY OF THE BPMN DIAGRAM

The BPMN diagram has many elements to describe the three dimensions of software: functional, structural and behavioural. In Figure 2 an example of a routine task in an office modelled by BPMN diagram is shown. Complete and consistent UML class, UML state machine and UML use case diagrams could be derived from the sufficiently consistent BPMN diagram.

Figure 2 presents a request of a service from an office. The mappings between BPMN diagram and UML diagrams are also shown there. The BPMN diagram is in simple and unambiguous relationships with state diagram (behaviour), and use case diagram (functionality) based on consistency rules. But relationships with class diagram (structure) and with state machine diagram are more complex.

Each BPMN object element in BPMN diagram corresponds to only one class from the class diagram, and this object element (*Data Object*) is interpreted as an instance of a proper class. The associations between classes are derived from BPMN diagram with horizontal control flow between tasks, to which one the adequate element objects (*Data Object*) are connected with *DataAssociations*. Moreover, each BPMN element object has simple and unambiguous state of the state machine diagram. Each BPMN element object with its state corresponds to only one state in the state diagram, but

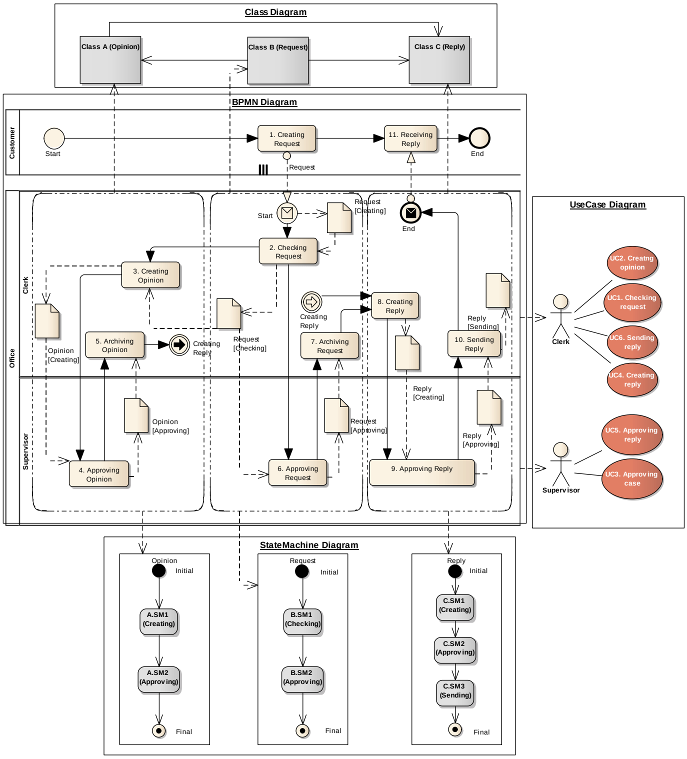


Fig. 2. The Use Case Realization Diagram based on BPMN standard

```
CLASS bpmnDiagram
ATTRIBUTES:
    lanes: List<Actor>                  // functional dimension
    events: List<Events>                // behaviour dimension
    tasks: List<Activity>               // behaviour and functional dimension
    dataobjects: List<Instances>        // structure dimension
    sequenceflows: List<ControlFlow>    // behaviour dimension
    dataassociations: List<ObjectFlow>// structure dimension
METHODS:
    checkConsistency(BPMNDiagram) RETURN result
    checkCompleteness(BPMNDiagram) RETURN result
    createCLDiagram(BPMNDiagram) RETURN void
    createSMDiagram(BPMNDiagram) RETURN void
    createUCDiagram(BPMNDiagram) RETURN void
```

Fig. 3. Object Pseudo-Code of the FSB UML diagram

transitions between states are not simple and unambiguous. Transitions in this State Chart are derived from BPMN diagram with vertical sequence flow between tasks, to which one adequate element objects are connected with *DataAssociations*. In a more simpler way the BPMN tasks can be mapped onto use case diagram.

A few BPMN tasks are realized by one use case, and each use case is associated with an actor in Use Case diagram. The Actor is derived from the BPMN Lane, which is grouping particular BPMN tasks. In order to improve the readability of Fig. 2, not all dependencies between diagrams are visible.

The BPMN diagram was presented in Object Pseudo-Code in Fig. 3.

4.1. SUFFICIENT COMPLETENESS OF THE BPMN DIAGRAM

We described in Section 3 the sufficient criteria for the consistency modelling. Among other criteria our diagram must describe the function, structure and behaviour aspects of the system. It means that BPMN diagram contains the elements, which enable to describe function, structure and behaviour. This property we called the sufficient completeness.

```
METHOD checkCompleteness(bpmnDiagram)
   IF bpmnDiagram has no start or bpmnDiagram has no end THEN
      RETURN false
   END IF
   IF bpmnDiagram has no activity or bpmnDiagram has no sequenceflow THEN
      RETURN false
   END IF
   IF bpmnDiagram has no dataobject or bpmnDiagram has no dataassociation THEN
      RETURN false
   END IF
```

Fig. 4. Object Pseudo-Code of the completeness checking of the BPMN diagram

Figure 4 presents the simplified checking method for completeness of the BPMN diagram. The dimension of functionality describes *Lanes*, and *Activities*. *DataObjects* and *DataAssociations*, and *SequenceFlows* with *Tasks* represent the dimension of structure, and the dimension of behaviour comprises of *Starts*, *Ends*, *Events*, *SequenceFlows* and *Tasks*.

The common elements for the three dimensions are *Tasks* and *DataObjects*. Those elements could be used to integrate the three dimensions of software architecture within this BPMN diagram. Other elements of the BPMN diagram model fully describe the three dimensions so the BPMN diagram is sufficiently complete.

4.2. SUFFICIENT CONSISTENCY OF THE SIMPLE BPMN DIAGRAM

Other sufficient criteria for consistency presented in Section 3 were related with connectivity of the BPMN diagram. If we assume that all nodes of the BPMN diagram are vertices of the graph then if the graph is connected then the BPMN diagram is also connected. In Figure 5. the Depth-first search (DFS) algorithm for searching graph paths is shown. The original algorithm was extended with eliminating duplicated paths, but with vertex and edges covering. This algorithm is recursive. If there is lack start or end nodes or the graph is not connected then the algorithm return error.

```
METHOD checkConnectivity(bpmnDiagram)
   IF current vertex in scenario is the last vertex THEN
      Add the main flow to scenarios list
      RETURN 0
   END IF
   Add current vertex to the unique vertices list
   Search for next vertex connected with current vertex
   IF no new next vertex THEN RETURN 0 END IF
   FOR founded vertices
      Push founded vertex to the scenario
      CALL checkConnectivity METHOD WITH founded vertex
      IF result no 0 THEN RETURN result END IF
      Pop founded vertex from the scenario
   END FOR
   RETURN 0
```

Fig. 5. Object Pseudo-Code of the sufficient consistency of the FSB UML diagram

It should be noted that the BPMN standard has a very large number of symbols, and there are many BPMN symbols, and BPMN constructions, which induce unconnected BPMN diagram to create. More details are described in next section.

```
METHOD createSMDiagram(smDiagram)
   FOR founded BPMN DataObject
      Create UML State WITH Name of BPMN DataObject state
      Relate UML State WITH UML Class
      Relate UML State WITH other UML States Linked by BPMN Activities
   END FOR
   RETURN
```

Fig. 6. Object Pseudo-Code of the method to create state machine diagram from the BPMN diagram

Latter sufficient criteria showed in Section 3 were related to mapping all elements of the BPMN diagram onto subsequent generated class, use case and state machine diagrams. Because inconsistencies arise between elements belonging to several models therefore the best method to avoid these inconsistencies is to create the subsequent diagrams based on the single consistent diagram. This property implies that the corresponding UML models (use case diagram, state machine diagram, class diagram) are consistent too. Below we proposed adequate methods in the object pseudo-code.

```
METHOD createCLDiagram(clDiagram)
   FOR founded BPMN DataObject
      Create UML Class WITH Name of BPMN DataObject
      Relate UML Class WITH founded BPMN DataObject
   END FOR
   RETURN
```

Fig. 7. Object Pseudo-Code of the method to create class diagram from the BPMN diagram

```
METHOD createUCDiagram(ucDiagram)
   FOR founded BPMN Lane
      Create UML Actor WITH Name of Lane
      Relate UML Actor WITH BPMN Lane
   END FOR
   FOR founded BPMN Task
      Create UML UseCase WITH Name of BPMN Task
      Relate UML UseCase WITH Lane contains founded BPMN Task
   END FOR
   RETURN
```

Fig. 8. Object Pseudo-Code of the method to create use case diagram from the BPMN diagram

4.3. SUFFICIENT CONSISTENCY OF THE FULL BPMN DIAGRAM

The BPMN standard operates large number of symbols. The more symbols are there to remember, the greater the cognitive load and potential errors. Several BPMN symbols like *Signal Events*, *Link Events*, or *Escalation Events* are not required to have ingoing or outgoing sequence flows. E.g. one uses links to allow the process designer to graphically divide a long and/or complex process sequence and thus to link up parts that are not next to each other on the diagram. Even *Message Events* do not have to be linked with the source of messages. Moreover there are many BPMN constructs, which cannot be transformed to the correct Petri nets [7]. Therefore the full BPMN diagram contains many different symbols, and many process constructions should be not connected.

In Figure Fig. 9. a business process was shown, in which there are two paths, but only one seems to be connected. If we assume that there is no link between the *IntermediateLinkThrow* event, and *IntermediateLinkCatch* event, then the second (with *Link Event* symbols) path is not connected.

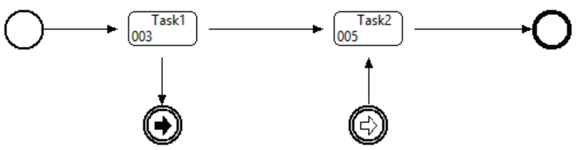


Fig. 9. BPMN diagram with unconnected path

The above example is trivial, and could be applied in BPMN tools, but there are many BPMN symbols, and constructs, which one could interpret in ambiguous ways or that cannot be transformed to any Petri net. Thus the BPMN diagrams with such symbols like *Link Events*, *Cancelation Events*, *Escalation Events* or *Signal Events* could be not consistent.

5. CONCLUSIONS

In this paper we have examined a BPMN diagram if it enables to keep the sufficient consistency and completeness of the business process model. The BPMN diagram allows to automatically generate complete workflow applications with no need for any "manual" programming, but with a few constraints. In a case that BPMN diagram has not comprised of certain BPMN elements then derived UML diagrams are consistent. In a case when the BPMN diagram contains such BPMN elements as *Signal Event*, *Escalation Event*, *Cancelation Event* or *Compensation Event* then BPMN diagram cannot be connected and that is why it cannot be consistent.

REFERENCES

[1] CHOI H., YEOM K., *An Approach to Software Architecture Evaluation with the 4+1 View Model of Architecture*, [in:] Ninth Asia-Pacific Soft. Eng. Conf., IEEE Computer Society, 2002, pp. 286–293.

[2] ISSA A., ABU RUB F.A., *Performing Early Feasibility Studies of Software Development Projects Using Business Process Models*. Proceedings of the World Congress on Engineering 2007, Vol. I, WCE 2007, July 2–4, London, U.K., 2007.

[3] BLUEMKE I., NIEPOSTYN J., *From Three Dimensional Document Circulation Diagram into UML*, [in:] Computing in Science and Technology 2011, Kwater Tadeusz et al. (eds.), 2012, Warsaw University of Life Sciences, ISBN 978-83-7583-378-2, pp. 7–22.

[4] NIEPOSTYN S. J., BLUEMKE I., *The Function-Behaviour-Structure Diagram for Modelling Workflow of Information Systems*, [in:] Advanced Information Systems Engineering Workshops, M. Bajec, J. Eder (eds.), Lecture Notes in Business Information Processing, Springer, 2012, Vol. 112, pp. 425–439 (EOMAS 2012).

[5] SPANOUDAKIS G., ZISMAN A., *Inconsistency management in software engineering: Survey and open research issues*, [in:] S.K. Chang (ed.), *Handbook of software engineering and knowledge engineering*, World Scientific Publishing Co., London 1999, 1, 329–380.

[6] WONG P.Y., GIBBONS J., *A process semantics for BPMN*, [in:] Proceedings of the 10th International Conference on Formal Methods and Software Engineering, Lecture Notes in Computer Science, 2008, Vol. 5256, pp. 355–374.

[7] DIJKMAN R.M., DUMAS M., OUYANG C., *Semantics and analysis of business process models in BPMN*, Information and Software Technology (IST), 2008, 50 (12), 1281–1294.

[8] WEIDLICH M, MENDLING J., *Perceived consistency between process models*, Inf. Syst., 2012, 37(2), 80–98.

[9] ROGGE-SOLTI A., KUNZE M., AWAD A., WESKE M., *Business process configuration wizard and consistency checker for BPMN 2.0*, [in:] T. Halpin, S. Nurcan, Krogstie.

[10] PAIGE R.F., BROOKE P.J., *Metamodel-Based Model Conformance and Multi-View Consistency Checking*, ACM Transactions on Software Engineering and Methodology, 2007, Vol. 16, Iss. 3.

[11] GERO J.S., *Design prototypes: a knowledge representation schema for design*, AI Magazine, 11(4), 26–36.

[12] TRUYEN F., *The Fast Guide to Model Driven Architecture*, The Basics of Model Driven Architecture, Cephas Consulting Corp., January 2006.

[13] FINKELSTEIN A., GABBAY D., HUNTER A., KRAMER J., NUSEIBEH B., *Inconsistency Handling in Multi-Perspective Specifications*, IEEE Computer Society Press, Transactions on Software Engineering, August 1994, 20(8), 569–578.

[14] Workshop on "Consistency Problems in UML-based Software Development II" 2003, L. Kuzniarz et al. (eds.), San Francisco, USA, Blekinge Institute of Technology, Research Report 2003:06.

Networked Control Systems,
Optimal Control, Time-Delay Systems

Michał MORAWSKI*
Przemysław IGNACIUK*

IMPROVING ROBUSTNESS OF LQ OPTIMAL NETWORKED CONTROL SYSTEMS IN THE PRESENCE OF DATA LOSS AND DELAY

The paper presents an algorithmic approach to improve robustness of linear-quadratic (LQ) optimal control strategies acting in networked environment. While providing smooth and efficient performance under nominal operating conditions, optimal controllers show sensitivity to parametric uncertainties and perturbations. In networked control systems the situation aggravates due to the delay in feedback information exchange among the system elements and possibility of data loss. By using additional control information, the proposed algorithm increases the stability margin without downgrading dynamical properties of LQ optimal control solution. The robustness to networked-induced uncertainties is evaluated experimentally.

1. INTRODUCTION

The problem of local plant control may be considered well investigated, with many efficient solutions implemented in practice [1]. However, the control of systems involving information transfer delay remains a challenging research topic [2]. The delay may constitute an integral characteristic of the controlled object, as is the case of communication [3, 4] and logistic [5, 6] networks, or appear as a consequence of introducing network as the signal distribution medium [7, 8]. Thus, in addition to the usual problems of plant model uncertainty, in networked control systems (NCSs) one needs to answer the challenges of information sharing between spatially separated entities. On the one hand, performing control actions on a distant object involves input delay – the time to transfer the controller command to the object. On the other hand, obtaining measurements to generate new controller commands involves output delay

* Institute of Information Technology, 215 Wólczańska St., Lodz University of Technology, 90-924 Łódź, Poland, e-mail: michal.morawski@p.lodz.pl, przemyslaw.ignaciuk@p.lodz.pl

– the delay on the return path from the object to the controller. The relevant information may not arrive on time, be lost, or discarded due to errors, augmenting the system uncertainty.

In local plant control, linear-quadratic (LQ) optimal strategies offer good quality and balanced dynamics. Unfortunately, networked-induced effects downgrade those favorable properties obstructing wider deployment of optimal controllers in NCSs. First of all, the presence of delay significantly complicates the procedure of finding the optimal solution [9]. The classical approaches, e.g. [10], result in sophisticated design expressions that are not adequate for real-time control under processing constraints, frequently encountered in NCSs. Recently obtained explicit solutions [9] relieve computational complexity. However, those results are provided in continuous time domain and thus are not directly applicable to the systems with finite rate of information exchange in networked environments. It has been shown in [11] that in discrete-time domain the optimal controller can be represented in predictor dynamics of low order. The reduced complexity and efficient performance are obtained under the assumption of constant, known delay. In order to successfully implement the optimal strategies in NCSs one needs to answer the communication process uncertainty related to information loss and delay jitter.

In this paper, an efficient implementation scheme of LQ optimal control strategies in NCSs is proposed. By appropriate treatment of control data history, the developed algorithm mitigates the impact of network-induced perturbations. As a result, robustness can be improved without throttling the system dynamics. The algorithm performance is tested experimentally in NCS implementation of remote steering of a cable-suspended load.

2. SYSTEM MODEL

Consider the system depicted in Fig. 1. The plant is to be controlled remotely according to a suitably chosen regulation strategy. The signals between the controller and the plant are exchanged via a communication network. Such configuration, while flexible and cost efficient, introduces additional phenomena that need to be accounted for in proper control system design.

First of all, even if fast analog-to-digital and digital-to-analog converters are applied to manipulate the signals at the plant, handling the events at the network interfaces causes extra delay that limits the achievable sampling rate. At the input interface, this delay is associated with capturing the arriving packets, extracting the relevant information, and applying the retrieved information for the control purposes. Similarly, at the output interface, finite time is needed to form and transmit packets through the network device. Therefore, unless the dominant time constant of the controlled process significantly exceeds the sampling period, in an NCS the effects of finite sampling rate should be addressed. Here, it is assumed that the system is time driven with period T and signals are applied through zero-order hold.

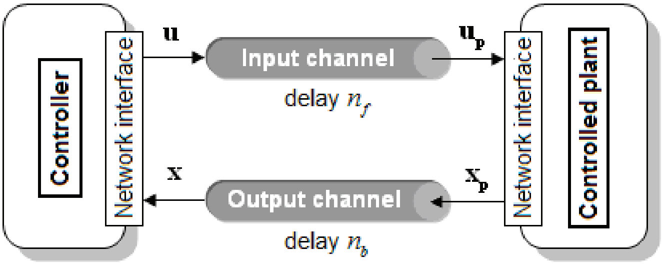


Fig. 1. Remote plant control

Secondly, the data transfer through a network channel is never instantaneous. Hence, unless only local signal exchange is considered with the transmission delay incorporated into the discretization period, the controller command $\mathbf{u}$ will be received at the plant with forward (input) delay $n_f > 0$, $\mathbf{u_p}(k) = \mathbf{u}(k - n_f)$, and the state measurement will be delivered with backward (output) latency $n_b > 0$, $\mathbf{x_p}(k) = \mathbf{x}(k - n_b)$, where $k = 1, 2, 3,...$ is the independent variable representing the evolution of time. The overall loop delay $\tau = n_f + n_b$.

The system dynamics are given by

$$\mathbf{x_p}(k+1) = \mathbf{F}\mathbf{x_p}(k) + \mathbf{G}\mathbf{u_p}(k) + \mathbf{d}(k) = \mathbf{F}\mathbf{x_p}(k) + \mathbf{G}f[\mathbf{x_p}(k-\tau)] + \mathbf{d}(k), \tag{1}$$

where $\mathbf{x_p} \in \Re^n$ is the plant state vector, $\mathbf{u_p} \in \Re^m$ is the plant input vector, $\mathbf{F} \in \Re^{n\times n}$ and $\mathbf{G} \in \Re^{n\times m}$, for m,n positive integers, $m \le n$, and $\mathbf{d}(k) \in \Re^n$ represents the cumulative effect of parametric uncertainties and external disturbances acting on the plant. The plant input $\mathbf{u_p}(k) = f[\mathbf{x_p}(k - \tau)]$ is influenced by network-induced effects of data loss and delay. $\mathbf{F}$ is assumed nonsingular and the pair ($\mathbf{F}$, $\mathbf{G}$) controllable.

3. LQ OPTIMAL CONTROL PROBLEM

The objective is to find a function $\mathbf{u}(k)$ so that the quadratic cost functional

$$J(\mathbf{u}) = \frac{1}{2}\sum_{k=0}^{\infty}\left[\mathbf{x_p}^T(k)\mathbf{Q}\mathbf{x_p}(k) + \mathbf{u}^T(k)\mathbf{R}\mathbf{u}(k)\right] \tag{2}$$

is minimized under nominal operating conditions ($\mathbf{d}(k) = 0$) of system (1). In (2), $\mathbf{Q} \in \Re^{n\times n}$ is semipositive definite state weighting matrix and $\mathbf{R} \in \Re^{m\times m}$ is positive definite input weighting matrix. In order to obtain unique stabilizing solution, **Q** should be chosen in such a way that the pair (**F**, $\mathbf{Q}_r$), where $\mathbf{Q} = \mathbf{Q}_r^T \mathbf{Q}_r$, is detectable.

3.1. CLASSICAL APPROACH

The presence of delay in (1) significantly complicates the search for optimal control sequence **u**(*k*) in problem (2). In order to circumvent this issue, a different model description involving a number of additional state variables to capture the effects of delay, can be applied. As a result, the original problem (2) is transformed into a standard, delay-free one, amenable to the classical design methods.

The system dynamics in the extended state space can be represented as

$$\mathbf{x_e}(k+1) = \mathbf{F_e}\mathbf{x_e}(k) + \mathbf{G_e}\mathbf{u}(k) + \mathbf{D_e}\mathbf{d}(k), \tag{3}$$

where $\mathbf{x_e} \in \Re^{n_e}$, $n_e = n + \tau m$, is the extended state vector,

$$\begin{aligned} &\mathbf{x_e}(k) = [\mathbf{x_p}(k) \quad \mathbf{w}_1(k) \quad \dots \quad \mathbf{w}_\tau(k)]^T, \\ &\mathbf{w}_i(k) = \mathbf{u}(k - \tau + i - 1) \text{ for } i = 1, \dots, \tau, \end{aligned} \tag{4}$$

$\mathbf{F_e} \in \Re^{n_e \times n_e}$ is extended state matrix, $\mathbf{G_e} \in \Re^{n_e \times m}$ and $\mathbf{D_e} \in \Re^{n_e \times m}$ are input matrices,

$$\mathbf{F_e} = \begin{bmatrix} \mathbf{F} & \mathbf{G} & 0 & \dots & 0 \\ 0 & 0 & \mathbf{I}_m & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{I}_m \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}, \mathbf{G_e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \mathbf{I}_m \end{bmatrix}, \mathbf{D_e} = \begin{bmatrix} \mathbf{I}_n \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \tag{5}$$

and $\mathbf{I}_l$ denotes $l \times l$ identity matrix. Thus, through transformation (4) and (5), instead of time-delay system (1) a delay-free system of increased order is obtained.

Solving problem (2) is equivalent to finding the minimum of

$$J(\mathbf{u}) = \frac{1}{2}\sum_{k=0}^{\infty}\left[\mathbf{x_e}^T(k)\mathbf{Q_e}\mathbf{x_e}(k) + \mathbf{u}^T(k)\mathbf{R}\mathbf{u}(k)\right] \tag{6}$$

where $\mathbf{Q_e} = diag\{\mathbf{Q},\ 0\}_{n_e \times n_e}$, subject to (3). Since delay is no longer present in (3) in an explicit form, the minimization of (6) amounts to solving a standard optimization problem, i.e. to calculate the gain matrix $\mathbf{K_e}$ for $\mathbf{u}(k) = -\mathbf{K_e}\mathbf{x_e}(k)$ from

$$\begin{aligned} &\mathbf{K_e} = (\mathbf{R} + \mathbf{G_e}^T\mathbf{P_e}\mathbf{G_e})^{-1}\mathbf{G_e}^T\mathbf{P_e}\mathbf{F_e}, \\ &\mathbf{P_e} = \mathbf{Q_e} + \mathbf{F_e}^T\mathbf{P_e}\mathbf{F_e} - \mathbf{F_e}^T\mathbf{P_e}\mathbf{G_e}(\mathbf{R} + \mathbf{G_e}^T\mathbf{P_e}\mathbf{G_e})^{-1}\mathbf{G_e}^T\mathbf{P_e}\mathbf{F_e}. \end{aligned} \tag{7}$$

Using transformation (3)–(6) one avoids handling the effects of delayed argument in problem (2). However, the alternative description leads to complexity $O(n_e^2)$.

3.2. PREDICTOR-BASED APPROACH

In [11] it was formally shown that instead of using the standard approach involving extended space (6), one can introduce an appropriately chosen predictor

$$\mathbf{v}(k) = \mathbf{F}^{\tau}\mathbf{x}_{\mathbf{p}}(k) + \sum_{i=1}^{\tau}\mathbf{F}^{i-1}\mathbf{G}\mathbf{u}(k-i),\ \mathbf{v} \in \Re^n, \tag{8}$$

and find the optimal control sequence for the original system (1) by solving the alternative optimization problem

$$J(\mathbf{u}) = \frac{1}{2}\sum_{k=0}^{\infty}\left[\mathbf{v}^T(k)\mathbf{Q}\mathbf{v}(k) + \mathbf{u}^T(k)\mathbf{R}\mathbf{u}(k)\right]. \tag{9}$$

The optimal input sequence is then given by $\mathbf{u}(k)$ = $\mathbf{K}\mathbf{v}(k)$ with the gain determined from

$$\begin{aligned}\mathbf{K} &= (\mathbf{R} + \mathbf{G}^T\mathbf{P}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{P}\mathbf{F},\\ \mathbf{P} &= \mathbf{Q} + \mathbf{F}^T\mathbf{P}\mathbf{F} - \mathbf{F}^T\mathbf{P}\mathbf{G}(\mathbf{R} + \mathbf{G}^T\mathbf{P}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{P}\mathbf{F}.\end{aligned} \tag{10}$$

The computational complexity reduces to $O(n^2)$, i.e. the order of the controlled plant. The delay does not affect the problem complexity.

4. ROBUST IMPLEMENTATION ALGORITHM

It follows from the analysis given in [11] that for sufficiently small $\mathbf{d}(k)$, Lyapunov stability of LQ optimal control system is guaranteed for constant, known delay. In the actual networked implementation, however, one cannot expect the delay to remain fixed, equal to the estimate obtained at the connection setup. In addition to experiencing variable, uncertain delay, the information transmitted in the network may be lost or corrupted, thus increasing the overall system uncertainty. As a result, one would need to throttle the system dynamics to preserve stability.

While potentially causing problems due to extra uncertainty source, however, the control system implementation in networked environment offers also algorithmic means of reducing that uncertainty influence. In order to improve robustness to network-induced hazards, it is proposed to apply the following implementation algorithm involving computation of multiple control values, established for different loop delays.

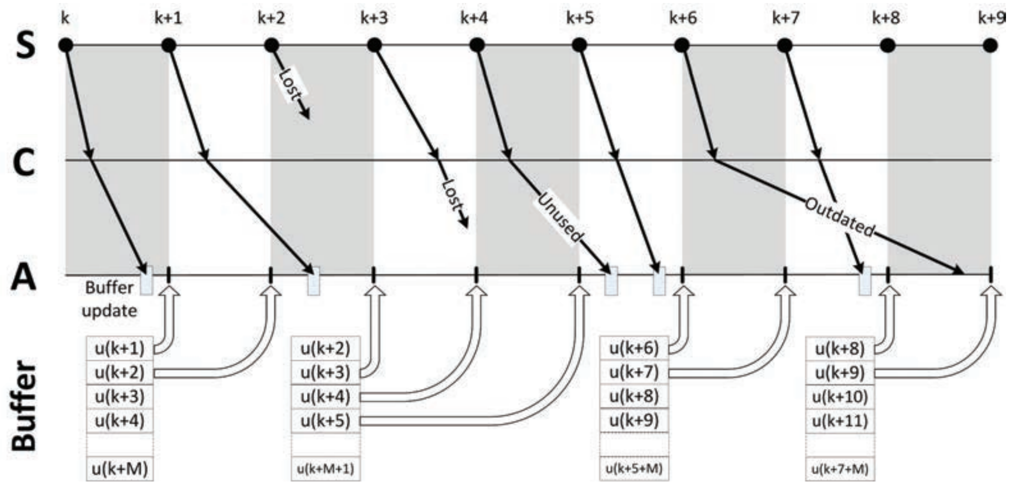


Fig. 2. NCS implementation scheme under various transfer conditions. Data produced by sensor (**S**) at instant *k* is processed by controller (**C**) and the calculated sequence of *M* control values reaches actuator (**A**) by instant *k* + 1 (nominal case); the timestamp comparison indicates that the received information is up-to-date which allows for buffer update; the first element from the buffer is applied by **A** at instant *k* + 1. Between instants *k* + 1 and *k* + 2, **A** receives no data, so at *k* + 2 the second element from the buffer is used for the control purposes. The buffer is updated once the packet sent by **S** at instant *k* + 1 reaches **A** by *k* + 3. The packets issued by **S** at instants *k* + 2 and *k* + 3 are lost, so **A** performs no buffer update and extracts subsequent elements from the buffer at *k* + 4 and *k* + 5. In the interval (*k* + 5, *k* + 6] **A** receives two packets; both are accepted, yet the latter overwrites the buffer contents since it contains more recent information. In turn, the packet issued at *k* + 6 arrives after the one sent at *k* + 7 and is ignored by **A** since it carries outdated information

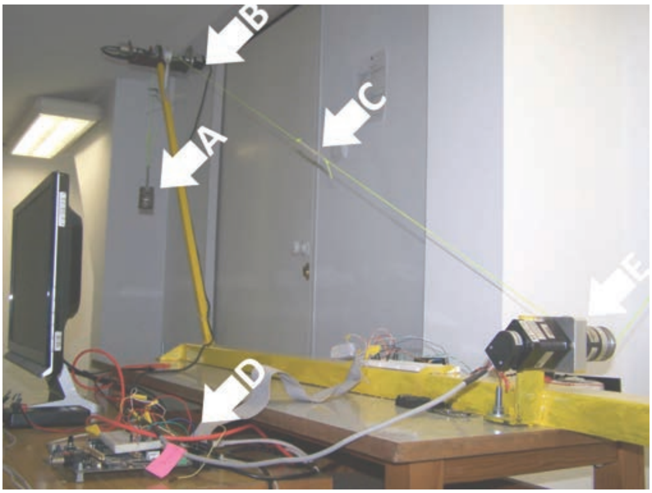


Fig. 3. Experimental setup: A – load, B – measurement head with encoder, C – elastic element, D – MCU, E – motor with gearbox and encoder

The algorithm operation is illustrated in Fig. 2. At the plant input interface a vector of M input values, M being a positive integer, is maintained in a buffer. The value corresponding to current time instant is applied to the actuator. Together with the plant state measurement, the contents of the buffer are sent to the controller. The controller operates in event-based mode and clock synchronization with the plant is not required. Instead of calculating a single $\mathbf{u}$ for (nominal) loop delay τ, the controller computes M values for delay spanning the range from 1 to M using the data obtained from the plant. All M values are sent in a single packet to the plant. Once the packet arrives at the plant, the buffer contents are updated if, based on timestamp comparison, the packet contains the recent information (note that route changes and statistical scheduling lead to delay fluctuations and packet reordering that may disturb the sequence of events in network transmission). With the proposed algorithm applied, the effect of network-induced component in $\mathbf{d}(k)$ is reduced, thus making stability easier to maintain without downgrading the dynamics.

5. PERFORMANCE EVALUATION

In order to assess the performance of proposed implementation algorithm, a series of experiments is conducted. The test setup, illustrated in Fig. 3, models vertical motion of a lift suspended on elastic line. The position of the lift is acquired using a standard quadrature encoder. The mass and its counterweight motion is driven by a step motor with step 1.2° and gearbox 5:1. Another encoder is connected to the motor shaft. The motor torque is large enough to neglect the motor dynamics, and high friction between the line and the motor shaft (with the wrap angle ≈500°) allows one to disregard the slip effect. The control signal accuracy (linear displacement at the motor shaft tip) is 0.0605 mm and the achievable accuracy of the position measurements is 0.2 mm.

The nominal plant dynamics are modeled as a 2^{nd}-order system

$$\dot{\mathbf{x}}_{\mathbf{p}}(t) = \begin{bmatrix} 0 & 1 \\ -\alpha/m & -\beta/m \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} b \\ 0 \end{bmatrix} \mathbf{u}_{\mathbf{p}}(t), \tag{11}$$

where x_1 is the mass position, x_2 mass velocity, m = 0.791 kg, α = 240.9906 kg/s^2, β = 3.5595 kg/s, b = 0.0605 mm/s. The sampling period T = 10 ms. Parameters α and β depend on the temperature and other environmental factors and may change by 10% from their nominal values. In order to further strain the test and evaluate the robustness to disturbances and measurement errors, the mass velocity is obtained through noisy numerical differentiation performed by a moving average filter with coefficients [1, –1].

In order to implement the network functionality, the motor is driven by NXP LPC1768 microcontroller unit (MCU), and the same MCU interfaces the encoder using built-in hardware component. The MCU works in time-driven fashion. It sends the current load and shaft positions together with the history of $M = 10$ control values applied to the plant in $T = 10$ ms intervals using a general purpose IP-based network to a PC implementing the control logic. The standard UDP transmission without acknowledgements, which requires about 8 kbps throughput in either direction, is used. As soon as the PC receives the packet, it computes a set of M control values for loop delays in the range 1 to M and sends this information to the actuator (through MCU). The PC operates in event-based mode. The packets experience variable delay and may be dropped in either direction (controller-to-actuator and sensor-to-controller). The MCU selects appropriate value from the set extracted from the received packet according to the logic described in Section 4. If the timestamp comparison indicates delay exceeding M, the system is stopped on emergency basis.

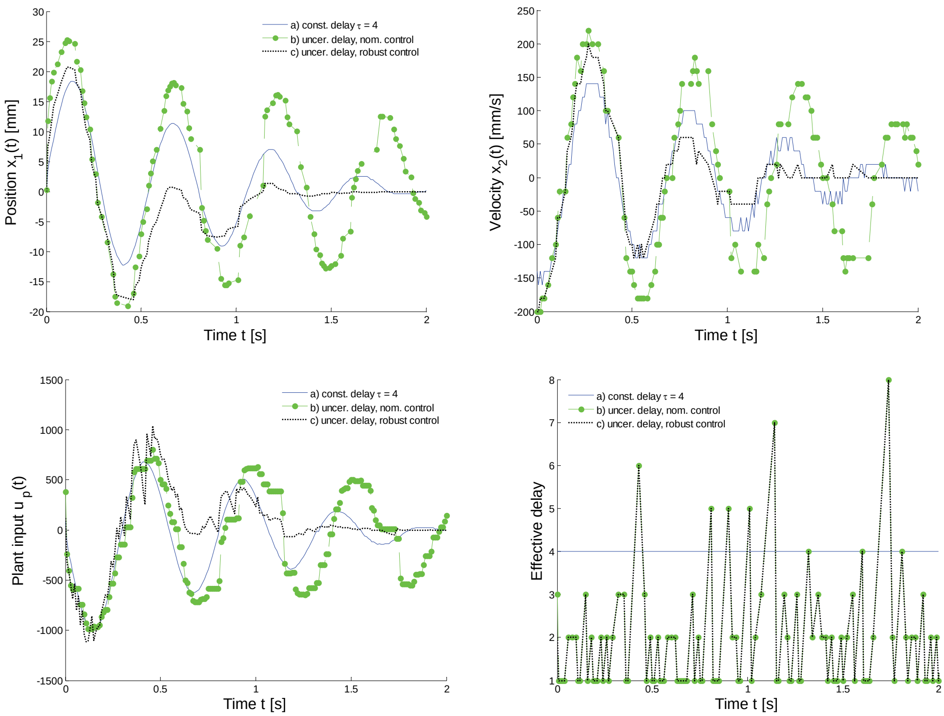


Fig. 4. Test results – mass position, velocity, input signal, and effective delay obtained for: a) nominal system, b) uncertain system with nominal control, c) uncertain system with robustness enhancement

The purpose of the experiment is to accelerate the mass to a desired velocity, set as 3000*0.0605 ≈ 180 mm/s, and stop the movement when the position is close to a demand one. Three scenarios are considered. Scenario a) reflects the nominal conditions – no packet drop, constant delay $\tau = 4$, whereas in scenario b) and c) the packets are dropped randomly which results in variable (effective) delay depicted in Fig. 4. Due to intended short stabilization interval in NCSs, in order to accurately reflect short-term networking fluctuations, the drop rate is modeled as a Markov process [12] with average drop probability 0.5. In experiment b) no compensation of channel uncertainty is applied, whereas the proposed robustness enhancement described in Section 4 is turned on in experiment c). The weighting matrices in optimal control problem are set as $\mathbf{Q} = diag\{400, 40\}$ and $\mathbf{R} = 1$. It follows from the graphs depicted in Fig. 4 that the network-induced uncertainty (curve b) significantly disrupts the nominal control process (reflected in curve a). The convergence time increases and stability may be lost. In turn, the proposed robustness enhancement (case c) reduces the influence of network related perturbations providing nearly nominal system performance.

6. CONCLUSION

An efficient algorithm for implementation of LQ optimal controllers in NCSs was developed. The proposed scheme reduces the impact of network-induced perturbations (variable delay and loss of information), thus allowing for reduced conservatism in the gain selection and better dynamics. The algorithm can be implemented independently of the plant control strategy, thus offering robustness enhancement for various predictor-based control systems operating under uncertain signal distribution channels. In particular, the control system designer may flexibly adjust the sampling rate (and network overhead) according to the transfer capabilities of communication system or control quality requirements.

ACKNOWLEDGEMENT

This work has been performed in the framework of a project “Design and validation of control algorithms in networked dynamical systems” financed by the National Science Centre of Poland – decision number DEC-2012/05/D/ST6/03030. P. Ignaciuk is a scholarship holder of Ministry of Science and Higher Education for Outstanding Young Researchers.

REFERENCES

[1] NOF S.Y. (Ed.), *Springer Handbook of Automation*, Springer-Verlag, Berlin–Heidelberg 2009.
[2] VYHLÍDAL T., LAFAY J.-F., SIPAHI R. (Eds.), *Delay Systems. From Theory to Numerics and Applications*, Springer-Verlag, Cham 2014.
[3] KARBOWAŃCZYK M., ŻUK J., *Quantity-Based Flow Control Strategy for Connection-Oriented Communication Networks*, Bulletin of the Polish Academy of Sciences – Technical Sciences, 2012, Vol. 60, No. 2, 293–300.

[4] IGNACIUK P., BARTOSZEWICZ A., *Congestion Control in Data Transmission Networks – Sliding Mode and Other Designs*, Springer-Verlag, London 2013.
[5] IGNACIUK P., BARTOSZEWICZ A., *Linear-Quadratic Optimal Control of Periodic-Review Perishable Inventory Systems*, IEEE Transactions on Control Systems Technology, 2012, Vol. 20, No. 5, 1400–1407.
[6] IGNACIUK P., *LQ Optimal and Robust Control of Perishable Inventory Systems with Multiple Supply Options*, IEEE Transactions on Automatic Control, 2013, Vol. 58, No. 8, 2108–2113.
[7] MORAWSKI M., ZAJĄCZKOWSKI A.M., *Approach to the Design of Robust Networked Control Systems*, International Journal of Applied Mathematics and Computer Science, 2010, Vol. 20, No. 4, 689–698.
[8] ZHANG L., GAO H., KAYNAK O., *Network-Induced Constraints in Networked Control Systems – A Survey*, IEEE Transactions on Industrial Informatics, 2013, Vol. 9, No. 1, 403–416.
[9] BASIN M., *New Trends in Optimal Filtering and Control for Polynomial and Time-Delay Systems*, Lecture Notes in Control and Information Sciences, Springer-Verlag, Berlin–Heidelberg, 2008, Vol. 380.
[10] UCHIDA K., SHIMEMURA E., KUBO T., ABE N., *The Linear-Quadratic Optimal Control Approach to Feedback Control Design for Systems with Delay*, Automatica, 1988, Vol. 24, No. 6, 773–780.
[11] IGNACIUK P., MORAWSKI M., *Linear-Quadratic Optimization in Discrete-Time Control of Dynamic Plants Operated with Delay Over Communication Network*, 22nd Mediterranean Conference on Control and Automation, Palermo, Italy, June 2014, 412–417.
[12] NOGUEIRA A., SALVADOR P., VALADAS R., PACHECO A., *Modeling Self-Similar Traffic over Multiple Time Scales based on Hierarchical Markovian and L-System Models*, Computer Communications, 2010, Vol. 33, No. S1, S3–S10.

genetic algorithms, shortest paths,
multi-criteria optimisation

Łukasz CHOMĄTEK*

EFFICIENT ALGORITHM FOR MULTI-CRITERIA OPTIMISATION IN THE ROAD NETWORKS

However nowadays there exist many systems that allow Users to find the route between given points on the map, their specific preferences are usually not taken into account. Popular web services such as Google Maps give the possibility to find an itinerary for drivers but their own criteria for the routes cannot be provided. In example, one cannot specify that he would avoid sharp turns or traffic lights. Detailed information about the road network is present in numerous Geographical Information Systems (GIS), but it is not utilized in the sense of the multi-criteria optimization, where non-dominated solutions are supposed to be found. In this paper we present an efficient way to handle multiple criteria with the use of the genetic algorithm that takes into account the reduction of the search space. Conducted research shown that proposed algorithm performs better than algorithms known from the literature.

1. INTRODUCTION

However nowadays there exist many systems that allow Users to find the route between given points on the map, their specific preferences are usually not taken into account. Popular web services such as Google Maps give the possibility to find an itinerary for drivers but their own criteria for the routes cannot be provided. In example, one cannot specify that he would avoid sharp turns or traffic lights. Detailed information about the road network is present in numerous Geographical Information Systems (GIS), but it is not utilized in the sense of the multi-criteria optimization, where non-dominated solutions are supposed to be found. In this paper we present an efficient way to handle multiple criteria with the use of the genetic algorithm that takes into account the reduction of the search space. Conducted research shown that proposed algorithm performs better than algorithms known from the literature. The

* Institute of Information Technology, Lodz University of Technology.

paper is organized as follows: at first we discuss the known methods of the multi-criteria optimization of the paths in the graphs. Then we describe the proposed modifications of the operators of the genetic algorithm. Finally we present obtained results and conclude our work.

2. MULTI-CRITERIA PATH OPTIMIZATION

In this chapter we described known exact and stochastic approaches for handling optimal path problem. We focused on the multi-criteria problem, because modern algorithms for the single criterion allow to find the solution in the amount of time that is needed for a few table look-ups. Detailed review of the single criterion algorithms will be given in the section ... For the multi-criteria problem we are interested in finding the Pareto-optimal solutions, because we assumed that there are no weights assigned to the criteria.

Let $G = (V, E)$ be a directed graph, where V and E are a set of its nodes and edges, respectively. Let $c : E \rightarrow (\square^+ \cup \{0\})^d$ be a function that assigns a d-dimensional cost vector to each edge in the graph. Let P be a path that connects nodes $s, t \in G$. We denote cost of the path as C, and it is understood as a sum of the costs of the edges that belong to the path. Let P and P' be the paths that connect nodes s and t. We say that P dominates P' if and only if $C_i(P)\ C_i(P')$ for each $i = 1, ..., d$. In our case we search for the set of paths that are nondominated by any other path.

2.1. EXACT METHODS

There exist some exact methods for the multi-objective optimization of the paths in graphs. One of the first attempts was an algorithm named Multicriteria Label-Setting proposed in [1]. It is an extension of the Dijkstra's [2] algorithm where a tuple of non-dominated labels is kept for each vertex. Each entry in such a tuple represents cost of a single optimization criterion. The idea of the algorithm is to eliminate possibly dominated labels. Another approach based on the Bellmanns–Ford algorithm is a Multicriteria Label-Correcting where individual labels can be considered many times during one execution of the algorithm. Mentioned algorithms can be only applied to the small sized graphs because of the large search space.

SHARC [3] algorithm allow to reduce the search space significantly because it utilizes some properties of the graph. This algorithm needs a preprocessing phase, where graph is divided into partitions and for each partition unimportant vertices are identified. As an unimportant vertex authors understood such a vertex that can be omitted during the search (for example vertex with only one "in" and one "out" edge). Research conducted in [3] revealed its efficiency in the multi-criteria optimization of the paths, but the search space is still large.

2.2. STOCHASTIC METHODS

Many of the known stochastic methods such as tabu-search [4] or particle swarm optimization [5] are utilized for finding the optimal path. Methods that can find a sub-optimal solutions in a reasonable time can be successfully applied instead of the exact methods in most cases. In route planning the actual velocity in each moment of the trip cannot be predicted, so that such a solutions are acceptable.

Many researchers decided to utilize genetic algorithms for solving multi-criteria shortest path problem. The first work we have found regarded routing in the computer networks [6], but proposed algorithm can be easily applied to the road networks. In this work each individual represented a single path and genes of the individual were the nodes of the network. During the mutation, one of the nodes was removed from the path and new node was added instead. In the crossover phase, common nodes of the selected individuals were identified and then the parts of the path were interchanged between the chromosomes. General scheme of the algorithm was taken from the Simple Genetic Algorithm. As one can see the algorithm did not take into account the topology of the network, which can be useful because of the potential reduction of the search space. Presented algorithm was suitable for rather small graphs, because for larger instances it was unable to find a proper solution.

Another approach was presented in [7], where the schema of the algorithm was similar to one presented above, but some of the network features were modifying the value of the cost function. The authors suggested giving some penalty for: turns, traffic lights on the path and low road category in the administrative division. In this approach, the information about the network topology is considered only in the cost function, but not in the genetic operators.

The third way in the genetic algorithms designed to find the shortest path was to apply the idea of viruses [8]. The viruses are patterns that are known to have good influence to the value of cost function and in the case of path-finding they are precomputed shortest paths between some points on the map. Viruses influences the genetic operators but have two disadvantages: one have to compute the proper virus patterns and the enforced injection of some given patterns disturbs the idea of the genetic algorithm which is supposed to work in a random way.

Due to the conducted review of the literature, we noticed that there is no genetic algorithm, which takes into account the topology of the road network graph both in the cost function and genetic operators.

2.3. SEARCH SPACE REDUCTION

The main problem in finding the shortest path for large graphs is the size of the search space. It influences the amount of time needed to find a solution. Modern algorithms consist of two phases. In the first phase some kind of preprocessing is done, and its goal is to find some dependencies in the graph, which can help in reducing the search space. In the second phase of the algorithm actual queries for shortest paths are performed. It is assumed that the first phase of the algorithm can take significant amount of time if it will help to reduce the amount of time needed to execute a query.

There are several ways to reduce the search space for the single criterion optimization, which can be divided into five groups:

- pruning of nodes,
- decomposition of the problem,
- hierarchical division of edges,
- omitting unimportant nodes and edges,
- precomputing of some distances

The pruning methods applied for the graphs, where some coordinates can be added to their nodes, are based on the observations that in the real world examples one doesn't have to analyze the whole graph, but only a certain region containing the origin and the destination node. Such an attempt was made by Karimi [9] and Fu [10], where the search space was limited to a rectangle with specified dimensions.

Regarding the second groups of methods, Nicholson [11] proposed two independent searches for each query. One of them was a classical search from source to destination node; the second one was carried from destination to source. Direction of graph edges in case of backward search is opposite to the real one. To complete the query, at least one node must be visited by both forward and backward searches.

Algorithms that belong to the third group are inspired by the observation of real drivers behavior. When one has to take a long trip, he usually tries to reach the major road and get as close as possible to the destination. When he is close to the destination, he chooses smaller roads to reach his goal. When driver is traveling on big roads he is not considering leaving them while he is far from the destination – actually this means the reduction of the search space by pruning the edges. Administrative division of the roads into categories is not sufficient for such algorithms, as the roads on the certain level often are not a connected component. To overcome this problem, several authors [12], [13] proposed their own hierarchical division. This significantly reduced the querying time, but the hierarchy-building process was a time consuming task.

In the fourth group we found algorithms where authors [14] propose some rules to reduce the structure of the input graph by removing a groups of unimportant nodes and edges and substituting them by a single node. An example of the rule is to remove the node which has only one "in" and one "out" edge.

The last group contain the most efficient known algorithms like Hub Labeling. They are based on the preprocessing phase, in which distances between some nodes are calculated. Then, in the querying phase one only have to do few lookups in the distance table. The research [15] shown that Hub Labeling is only four times slower than in the case of lookup-table with all possible distances precomputed.

However considered methods are very fast, in most cases they can be only used in the single criterion scenario. On the other hand, our previous research shown that taking into account some methods of search space reduction in the genetic algorithms [16], can positively influence their convergence to the optimal result in the single criterion scenario [17]. We decided to generalize our method to the multi-criteria case.

3. PROPOSED APPROACH

In this section we describe the modification of the genetic operators which utilize the method of the search space reduction proposed in the Highway Hierarchies [12] algorithm. Highway Hierarchies is a two-phase method, in which in the first phase each edge has a hierarchy level assigned which corresponds to its importance. In the second phase, edges on the higher hierarchy level are automatically added to the search scope while these on the lower level are discarded. The brief description of the algorithm is given in the section ... In the following sections we show the proposed genetic operators.

3.1. HIGHWAY HIERARCHIES ALGORITHM

In HH, the graph G consist a set of nodes *V* and a set of edges *E*. It is divided into *L* subsets. The result is a sequence of graphs $G = G^0, G^1, \ldots, G^L$, with nodes appropriately: $V^0 = V, V^1 V^2 \ldots V^L$ and edges $E^0 = E, E^1 E^2 \ldots E^L$. For each edge, information on the maximum hierarchy level to which it belongs, is kept.

Highway Hierarchies algorithm requires two parameters: *H* – identifying the degree to which requests for the shortest way are met without coming to a higher level in the hierarchy and *L*, which represents the maximum permissible hierarchy level. The method used to iteratively generate a higher level with number *l + 1* for graph G^l is as follows:

1. For each node $v \in V$, build the neighborhood N_H^l for all nodes reached from *v* by using Dijkstra's algorithm in graph G^l, respecting the *H* constraint. Set the state of the node *v* to „active".
2. For each vertex $v \in V$:
   a. Build the partial tree *B*(*v*), and assign to each node its state. The state of the node is inherited from the parent node every time the node is reached or settled. Node becomes „passive" if on the shortest path $\langle v, u, ..., w \rangle$,where $v \neq u \neq w$:

$$\left| N_H^l(u) \cap N_H^l(w) \right| \leq 1 \tag{1}$$

   Partial tree is completed, when reached but not settled nodes don't exist.

b. For each node t, which is a leaf node in the tree $B(v)$ move each edge (u, w), where $u \in N_H^l(t)$, $w \in N_H^l(v)$ to the higher hierarchy level.

During the first stage, a highway hierarchy is constructed, where each hierarchy level G^l, for $l < L$, is a modified subgraph of the previous level graph G_{l-1} such that no canonical shortest path in G_{l-1} lies entirely outside the current level for all sufficiently distant path endpoints: this ensures that all queries between far endpoints on level $l - 1$ are mostly carried out on level l, which is smaller, thus speeding up the search [12]. In the querying phase one use a bidirectional search, and for each visited node only edges that belong to the higher or equal hierarchy level as the preceding edge. It is proved that when the search scopes meet, the optimal path is found.

3.2. INITIAL POPULATION

We assume that we have input graph processed by the Highway Hierarchies algorithm, which means that we have one division per criterion. We construct random paths in the similar way as in the Highway Hierarchies algorithm. We use a bidirectional search and do not allow to add to priority queue edges with a hierarchy levels dominated by one we used to visit the current node. The difference is that we add a node to the priority queue with a random priority (instead of cost of travelling from the source node to the visited node). This ensures that the path will be find and it is not violate the rules of the hierarchical search.

3.3. FITNESS FUNCTION

While evaluating a fitness function we take into account the structure of the path and its overall cost. The structure is encoded as follows:

$$s(P) = \sum_{i=1}^{d} \sum_{l=0}^{L-1} \frac{\text{count}(P,l,c_i)(L-l)}{\text{count}(P)(l+1)} \tag{2}$$

where:

count(P) – number of edges in the path,

count(P, l, c_i) – number of edges on level l in the path with respect to the i-th criterion,

L – highest hierarchy level,

d – number of criteria.

Value of the function $s(P)$ should be minimized. The structure of this function does not promote any criterion. The value of the function is lower when there are many edges on the high hierarchy level for each criterion. This helps us to promote the paths that are well constructed in the sense of all of the criteria.

3.4. CROSSOVER AND MUTATION OPERATORS

The use of the hierarchical division while generating the initial paths enforces to handle its rules during the whole run of the algorithm. The crossover operator can only be applied for two individuals when there exist at least one common node on the paths it represents. We also check before substitution of the path fragments if the next edge after a crossing point is not dominated by the preceding edge.

The mutation operator works as follows: we take two random nodes on the mutated path and generate new random path between them using the same algorithm as while generating the initial population, considering only edges that are not dominated by these preceding the nodes selected for mutation.

Adapted crossover and mutation operators allowed us to take into account the search space reduction in the remaining parts of the genetic algorithm.

4. OBTAINED RESULTS

In this section we present the results of our research regarding the modified operators of the genetic algorithm applied for solving the problem of the multi-criteria optimization of the path between two nodes of the given graph.

The research was carried out on the map of the city of Lodz, Poland, containinga-bout 7000 nodes and 17000 edges. Some of the nodes do not represent crossroads but are placed on the map to preserve geographical coordinates of the roads.

We have chosen two nodes on the map which represented the starting and the finishing point on the route that was supposed to be found. The first node was located in the suburbs and the second one was placed in the opposite side of the city. As the simulation environment for finding the optimal routes, we used Matsim environment integrated with our algorithms. The data was prepared using OpenStreetMap (spatial information) and the PostGIS plug-in (spatial functions to extract values for the different criteria). We tested our approach for three different criteria: the length of the road, inverted maximum speed limit and the distance to the buildings. We decided to minimize all of the criteria, so the optimal path will be a short drive through the highly urbanized area but with the highest possible speed limits.

We tested the modifications of the genetic algorithm presented in the former section on two different genetic algorithms Vector Evaluated Genetic Algorithm (VEGA) and Strength Pareto Evolutionary Algorithm (SPEA). We decided to choose VEGA because of its two phase nature: during the iterations, individuals are divided into some groups and in each group they’re optimized with respect to the single criterion. After that, the population is merged and the multi-criteria optimization is performed. Such a cycle can appear multiple times. For the multi-criteria optimization we chose the SPEA algorithm which takes into account the dominance relation between the solutions represented in chromosomes in the selection step.

Before starting the presented experiments we tuned the algorithms' parameters. We obtained the best results for: mutation probability – 10%, tournament selection with three randomly selected individuals and one winning individual, crossover probability – 90%.

4.1. GENETIC DIVERSITY OF THE POPULATION

In this section we would like to discuss the influence of our modifications of the genetic operators into the genetic diversity of the population. Preservation of the genetic diversity is needed to avoid the premature convergence of the algorithm, where the evolution stuck in the local minimum. We decided to keep 30 individuals in the population. In the case, where individuals are generated randomly without any information about the structure of the graph, the chromosome with the best fitness can dominate the population, because of the large search space. This situation is illustrated in the Fig. 1.

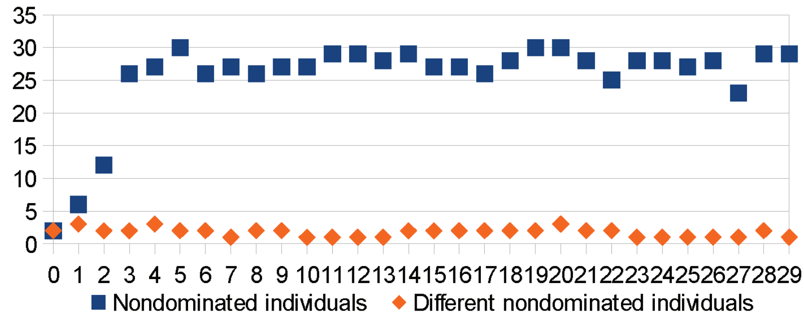


Fig. 1. Premature convergence of the VEGA algorithm

The situation is significantly different, when the search space reduction method is applied. We decided to perform hierarchical division into $L = 3$ levels with $N = 5$ neighborhood size. The result is shown in the Fig. 2.

On both figures we show the number of non-dominated individuals in the population and the number of different non-dominated individuals. In the best situation, the number of different non-dominated chromosomes should be high, which will show the exploration of the solution set. What is more, there should persist some dominated individuals because they can evolve into new optimal solutions. As one can see, application of proposed operators causes better diversity of the population, so that the search space can be explored in the more efficient way.

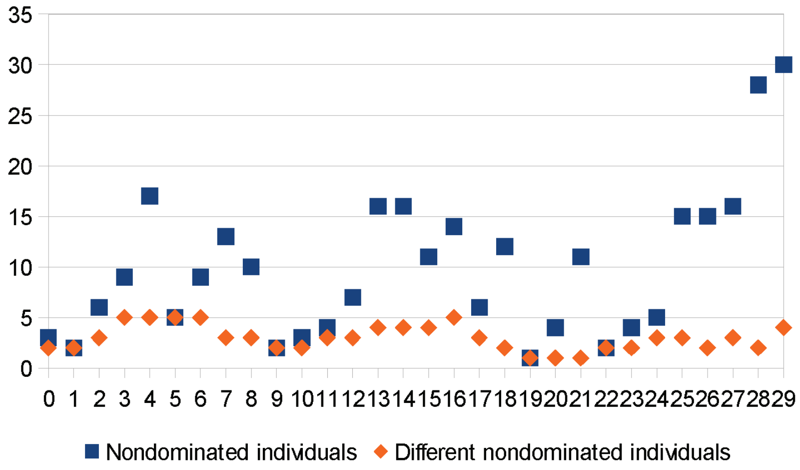


Fig. 2. Genetic diversity of the population with the search space reduction method applied

5. CONCLUSIONS

Conducted research shown that proposed modifications of the genetic algorithms significantly influences the optimisation process. The reduction of the search space results in their better performance while calculating the initial population. What is more, it can be a way to reduce the phenomenon of the premature convergence as it helps to maintain the genetic diversity of the population. Proposed algorithm can be successfully applied to the customizable applications where the search criteria are free to choose. In our future works we plan to extend our methods by proposing measures of the genetic diversity of the population. Such an attempt will be another way to reduce the influence of the premature convergence.

REFERENCES

[1] MARTINS E., *On a multicriteria shortest path problem*, European Journal of Operational Research, 1984, Vol. 16, No. 2, 236–245.
[2] DIJKSTRA E.W., *A note on two problems in connexion with graphs*, Numerische Mathematik, 1959, Vol. 1, 269–271.
[3] BAUER R., DELLING D., *SHARC: Fast and robust unidirectional routing*, Journal of Experimental Algorithmics, 2009, Vol. 14, No. 4.
[4] POTVIN J., NAUD M., *Tabu search with ejection chains for the vehicle routing problem with private fleet and common carrier*, Journal of the Operational Research Society, 2011, Vol. 62, No. 2, 326–336.
[5] HAMZHEEI M., REZA Z., HANNANEH R., *An ant colony-based algorithm for finding the shortest bidirectional path for automated guided vehicles in a block layout*, The International Journal of Advanced Manufacturing Technology, 2013, Vol. 64, No. 1, 399–409.

[6] MUNETOMO M., TAKAI Y., SATO Y., *A migration scheme for the genetic adaptive routing algorithm*, [in:] 1998 IEEE International Conference on Systems, Man, and Cybernetics, 1998, Vol. 3, 2774–2779.

[7] KANOH H., *Dynamic route planning for car navigation systems using virus genetic algorithms*, Int. J. Know.-Based Intell. Eng. Syst., 2007, Vol. 11, No. 1, 65–78.

[8] FU P., QIAO J., YIN H., *A virus evolutionary genetic algorithm usinglocal selection*, [in:] Innovative Computing, Information and Control, 2007, 572–582.

[9] KARIMI H.A., *Real-time optimal route computation: A heuristic approach*, ITS Journal, 1996, Vol. 3, No. 2, 111–127.

[10] FU L., *Real-time vehicle routing and scheduling in dynamic and stochastic traffic networks*, 1996.

[11] NICHOLSON J., *Finding the shortest route between two points in a network*, Computer Journal, 1966, Vol. 9, 275–280.

[12] SANDERS P., SCHULTES D., *Highway Hierarchies Hasten Exact Shortest Path Queries*, Lecture Notes in Computer Science, 2005, Vol. 3669, 568–579.

[13] DELLING D., SANDERS P., SCHULTES D., WAGNER D., *Engineering Route Planning Algorithms*, Lecture Notes in Computer Science, 2009.

[14] GEISBERGER R., SANDERS P., SCHULTES D., DELLING D., *Contraction Hierarchies: Faster and Simpler Hierarchical Routing in Road Networks*, Lecture Notes in Computer Science, 2008, Vol. 5038, 319–333.

[15] ABRAHAM I., DELLING D., GOLDBERG A.V., WERNECK R., *A hub-based labeling algorithm for shortest paths in road networks*, Lecture Notes in Computer Science, 2011, Vol. 5038, 319–333.

[16] CHOMATEK Ł., ZAKRZEWSKA D., *Search Space Reduction in the Combinatorial Multi-Agent Genetic Algorithms*, Intelligent Systems 2014 Conference, 2014.

[17] CHOMATEK Ł., *Application of genetic algorithms for route planning in the hierarchical road networks*, SMC 2013. XV International Conference. System Modelling and Control, 2013.

BIBLIOTEKA INFORMATYKI SZKÓŁ WYŻSZYCH

Information Systems Architecture and Technology. Advances in Web-Age Information Systems, pod redakcją Leszka BORZEMSKIEGO, Adama GRZECHA, Jerzego ŚWIĄTKA, Zofii WILIMOWSKIEJ, Wrocław 2009

Information Systems Architecture and Technology. Service Oriented Distributed Systems: Concepts and Infrastructure, pod redakcją Adama GRZECHA, Leszka BORZEMSKIEGO, Jerzego ŚWIĄTKA, Zofii WILIMOWSKIEJ, Wrocław 2009

Information Systems Architecture and Technology. Systems Analysis in Decision Aided Problems, pod redakcją Jerzego ŚWIĄTKA, Leszka BORZEMSKIEGO, Adama GRZECHA, Zofii WILIMOWSKIEJ, Wrocław 2009

Information Systems Architecture and Technology. IT Technologies in Knowledge Oriented Management Process, pod redakcją Zofii WILIMOWSKIEJ, Leszka BORZEMSKIEGO, Adama GRZECHA, Jerzego ŚWIĄTKA, Wrocław 2009

Information Systems Architecture and Technology. New Developments in Web-Age Information Systems, pod redakcją Leszka BORZEMSKIEGO, Adama GRZECHA, Jerzego ŚWIĄTKA, Zofii WILIMOWSKIEJ, Wrocław 2010

Information Systems Architecture and Technology. Networks and Networks Services', pod redakcją Adama GRZECHA, Leszka BORZEMSKIEGO, Jerzego ŚWIĄTKA, Zofii WILIMOWSKIEJ, Wrocław 2010

Information Systems Architecture and Technology. System Analysis Approach to the Design, Control and Decision Support, pod redakcją Jerzego ŚWIĄTKA, Leszka BORZEMSKIEGO, Adama GRZECHA, Zofii WILIMOWSKIEJ, Wrocław 2010

Information Systems Architecture and Technology. IT TModels in Management Process, pod redakcją Zofii WILIMOWSKIEJ, Leszka BORZEMSKIEGO, Adama GRZECHA, Jerzego ŚWIĄTKA, Wrocław 2010

Information Systems Architecture and Technology. Web Information Systems Engineering, Knowledge Discovery and Hybrid Computing, pod redakcją Leszka BORZEMSKIEGO, Adama GRZECHA, Jerzego ŚWIĄTKA, Zofii WILIMOWSKIEJ, Wrocław 2011

Information Systems Architecture and Technology. Service Oriented Networked Systems, pod redakcją Adama GRZECHA, Leszka BORZEMSKIEGO, Jerzego ŚWIĄTKA, Zofii WILIMOWSKIEJ, Wrocław 2011

Information Systems Architecture and Technology. System Analysis Approach to the Design, Control and Decision Support, pod redakcją Jerzego ŚWIĄTKA, Leszka BORZEMSKIEGO, Adama GRZECHA, Zofii WILIMOWSKIEJ, Wrocław 2011

Information Systems Architecture and Technology. Information as the Intangible Assets and Company Value Source, pod redakcją Zofii WILIMOWSKIEJ, Leszka BORZEMSKIEGO, Adama GRZECHA, Jerzego ŚWIĄTKA, Wrocław 2011

Information Systems Architecture and Technology. Web Engineering and High-Performance Computing on Complex Environments, pod redakcją Leszka BORZEMSKIEGO, Adama GRZECHA, Jerzego ŚWIĄTKA, Zofii WILIMOWSKIEJ, Wrocław 2012

Information Systems Architecture and Technology. Networks Design and Analysis, pod redakcją Adama GRZECHA, Leszka BORZEMSKIEGO, Jerzego ŚWIĄTKA, Zofii WILIMOWSKIEJ, Wrocław 2012

Information Systems Architecture and Technology. System Analysis Approach to the Design, Control and Decision Support, pod redakcją Jerzego ŚWIĄTKA, Leszka BORZEMSKIEGO, Adama GRZECHA, Zofii WILIMOWSKIEJ, Wrocław 2012

Information Systems Architecture and Technology. The Use of IT Models for Organization Management, pod redakcją Zofii WILIMOWSKIEJ, Leszka BORZEMSKIEGO, Adama GRZECHA, Jerzego ŚWIĄTKA, Wrocław 2012

Information Systems Architecture and Technology. Intelligent Information Systems, Knowledge Discovery, Big Data and High Pefomance Computing, pod redakcją Leszka BORZEMSKIEGO, Adama GRZECHA, Jerzego ŚWIĄTKA, Zofii WILIMOWSKIEJ, Wrocław 2013

Information Systems Architecture and Technology. Networks Architecture and Applications, pod redakcją Adama GRZECHA, Leszka BORZEMSKIEGO, Jerzego ŚWIĄTKA, Zofii WILIMOWSKIEJ, Wrocław 2013

Information Systems Architecture and Technology. Knowledge Based Approach to the Design, Control and Decision Support, pod redakcją Jerzego ŚWIĄTKA, Leszka BORZEMSKIEGO, Adama GRZECHA, Zofii WILIMOWSKIEJ, Wrocław 2013

Information Systems Architecture and Technology. Models of Decision Making in the Process of Management in a Risky Environment, pod redakcją Zofii WILIMOWSKIEJ, Leszka BORZEMSKIEGO, Adama GRZECHA, Jerzego ŚWIĄTKA, Wrocław 2013

Wydawnictwa Politechniki Wrocławskiej
są do nabycia w księgarni
plac Grunwaldzki 13, 50-377 Wrocław
budynek D-1 PWr., tel. 71 320 29 35
Prowadzimy sprzedaż wysyłkową
zamawianie.ksiazek@pwr.wroc.pl

ISBN 978-83-7493-857-0