Journal Subline

LNCS 10840

Transactions on

Computational Collective Intelligence XXIX

Ngoc Thanh Nguyen • Ryszard Kowalczyk
Editors-in-Chief

Springer

Lecture Notes in Computer Science 10840

More information about this series at http://www.springer.com/series/8851

Ngoc Thanh Nguyen · Ryszard Kowalczyk (Eds.)

Transactions on Computational Collective Intelligence XXIX

Springer

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISSN 2190-9288 ISSN 2511-6053 (electronic)
Transactions on Computational Collective Intelligence
ISBN 978-3-319-90286-9 ISBN 978-3-319-90287-6 (eBook)
https://doi.org/10.1007/978-3-319-90287-6

Library of Congress Control Number: 2018940913

Printed on acid-free paper

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

Preface

It is my pleasure to present Volume 29 of the LNCS *Transactions on Computational Collective Intelligence*. This volume includes ten interesting and original papers that have been selected after peer review.

The first paper, "Fuzzy Logic and PD Control Strategies of a Three-Phase Electric Arc Furnace" by Loredana Ghiormez, Octavian Prostean, Manuela Panoiu and Caius Panoiu, is devoted to the problem of fuzzy control and a conventional proportional derivative control for the electrode positioning system. The authors propose a comparison of the performance analysis of a conventional PD controller and a fuzzy-based intelligent controller. For this, they simulated the systems using Matlab/Simulink software. The results of the closed-loop systems show that the proposed fuzzy controller has better dynamic performance, rapidity, and good robustness compared with the proposed PD controller.

The second paper entitled "Bionic Hand Control in Real Time Based on Electromyography Signal Analysis" by Martin Tabakov, Krzysztof Fonal, Raed A Abd-Alhameed, and Rami Qahwaji describes fuzzy type-1 and type-2 models for the control of a bionic hand in real time. The experiments conducted show comparable results with respect to applied assumptions that give the confidence to implement the proposed concept into real-time control processes.

In the third paper, "Dynamic Decision Support in the Internet Marketing Management," Paweł Ziemba, Jarosław Jankowski, Jarosław Wątróbski solve the problem of selecting an advertisement variant on the basis of dynamically changing values of evaluation criteria. The authors have built a framework, used in an online environment, of a dynamic multi-criteria decision analysis, which is based on the PROMETHEE method that makes it possible to carry out a very detailed analysis of a decision process and obtained solutions. The simulation results are very interesting.

The fourth paper, "Biological Regulation and Psychological Mechanisms Models of Adaptive Decision-Making Behaviors: Drives, Emotions, and Personality" by Amine Chohra and Kurosh Madani, describes a framework for adaptive agent decision-making modeling of biological regulation and psychological mechanisms. For this purpose, the authors have worked out a perception–action cycle scheme for the agent–environment interactions and deduced framework for adaptive agent decision-making modeling. Next, they developed several motivation systems, drives, personality traits, emotions, and a neural architecture implementation of the framework. The results demonstrate how the personality and emotion of the agent can be used to regulate the intensity of the interaction, predicting a promising result in future, and the influences of the agent behavior, which could be very interesting for cooperative agents.

The fifth paper entitled "SWRL-Based Recommendation System for Provision of First Aid" by Martina Husáková proposes an SWRL-based prototype for decision-making if first aid is necessary. The author focuses on a conceptualization of the first aid application domain with the main aim of building a recommendation-based system providing advice in cases of emergency. The author shows that ontologies used in information and knowledge-based structures play an important role in modelling complex knowledge.

In the sixth paper, "Queue Lengths Management for Deterministic Queuing Systems," Martin Gavalec and Zuzana Němcová present a discussion on two methods for cost optimization of the deterministic queuing system based on the control of the queue lengths. The first method uses the evaluation of actual states in the particular service places according to their development. The second method is based on the simulation of the future states, and on this basis the appropriate time and type of modification of the system set-up are suggested.

The next paper, "W3SD: WordNet and Wiktionary-Based Approach for Word Sense Disambiguation" by Mohamed Ben Aouicha, Mohamed Ali Hadj Taieb, and Hania Ibn Marai, describes the proposed approach of W3SD, which is based on the words surrounding the polysemous word in a context. The meanings of words are represented by vectors composed of weighted nouns using WordNet and Wiktionary features through the taxonomic information content from WordNet and the glosses from Wiktionary. The main contribution of this paper is based on feature selection for disambiguation purpose.

In the eighth paper entitled "An MP/CP-Based Hybrid Approach for Optimization of the Resource-Constrained Scheduling Problems," Paweł Sitek and Jarosław Wikarek propose a hybrid approach based on integration mathematical programming and constraint logic programming to optimize resource-constrained scheduling problems. For evaluating the applicability and computational efficiency of the proposed approach and its implementation programming framework, the authors implement illustrative examples of optimization resource-constrained scheduling problems. A comparison of hybrid approach implementation environments is also presented.

In the ninth paper, "Analysis of the Structured Information for Subjectivity Detection in Twitter," Juan Sixto, Aitor Almeida, and Diego Lopez-de-Ipina analyze the opportunities of the structured information of social networks for subjectivity detection on Twitter micro-texts. They discuss the features of the structured information and their usefulness in the opinion mining sub-domain, especially in the subjectivity detection task. The authors present a novel classification of these features according to their origin.

The last paper entitled "An Efficient Parallel Method for Optimizing Concurrent Operations on Social Networks" by Hanh Phuong Du, Dang Hai Pham, and Hoa Ngoc Nguyen presents an approach to optimize the performance of both reading and writing concurrent operations on large-scale social networks. The authors focus on the directed and unweighted relationships among members in a social network. They work out an efficient parallel method based on utilizing an appropriate data structure, parallelizing the updating actions, and improving the performance of query processing.

April 2018 Ngoc Thanh Nguyen

Transactions on Computational Collective Intelligence

This Springer journal focuses on research in applications of the computer-based methods of computational collective intelligence (CCI) and their applications in a wide range of fields such as the Semantic Web, social networks, and multi-agent systems. It aims to provide a forum for the presentation of scientific research and technological achievements accomplished by the international community.

The topics addressed by this journal include all solutions to real-life problems for which it is necessary to use CCI technologies to achieve effective results. The emphasis of the papers published is on novel and original research and technological advancements. Special features on specific topics are welcome.

Contents

Fuzzy Logic and PD Control Strategies of a Three-Phase Electric Arc Furnace

Loredana Ghiormez[1(✉)], Octavian Prostean[1], Manuela Panoiu[2], and Caius Panoiu[2]

[1] Faculty of Automation and Computers, Politehnica University of Timisoara, Timisoara, Romania
{loredana.ghiormez, octavian.prostean}@upt.ro
[2] Faculty of Engineering Hunedoara, Politehnica University of Timisoara, Hunedoara, Romania
{manuela.panoiu, caius.panoiu}@upt.ro

Abstract. This paper presents a fuzzy control and a conventional proportional derivative control for the electrode positioning system of a three-phase electric arc furnace. Generally, it is necessary to maintain constant the arc lengths for these kinds of furnaces. The two control strategies proposed in this paper regulates the current of the electric arc because arc length depends by the electric arc current. In order to do this, a new model of the electric arc developed by the authors of this paper was used. This paper illustrates a comparison of the performance analysis of a conventional PD controller and a fuzzy based intelligent controller. The fuzzy intelligent based controller has two inputs and one output. The whole systems are simulated by using of Matlab/Simulink software. These systems are tested when applying a step disturbance in the process, when applying a sequence of step disturbances and when applying a second step disturbance before being compensate the effect of another disturbance. The responses of the closed-loop systems illustrates that the proposed fuzzy controller has better dynamic performance, rapidity and good robustness as compared to the proposed PD controller.

Keywords: Modeling · Fuzzy logic controller · PD controller
Current control

1 Introduction

Electric arc furnace (EAF) is the equipment that converts the electrical energy in thermal energy using graphite electrodes. Scrap is added in the furnace tank in order to be melted and to obtain liquid steel [1]. EAF taken into consideration in this paper is a three-phase one with direct action, so, the electric arcs appear between each of the three electrodes and the metal that will be melted. In order to melt the metals and to reduce the energy consumption it is necessary to optimize the delivered power to the furnace [2]. Arc power can be influenced in two ways: by modifying the position of the electrodes or by modifying the power delivered by the transformer to the EAF. A new position of the electrodes is performed at an EAF by using hydraulic actuators [3] so, obtaining different arc lengths.

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXIX, LNCS 10840, pp. 1–20, 2018.
https://doi.org/10.1007/978-3-319-90287-6_1

For a production cycle at an EAF can be found two main stages: melting and refining stages [4]. In the melting stage the metal is loaded in the furnace tank, it has solid shape and will be obtained the liquid state of this charging material. This stage is characterized by disturbances which influence the process because the scrap loaded in the furnace tank will change the position during the melting process and arc length will present different values. If there is not an automatic control of the position of the electrodes, these disturbances will be injected in the electrical power network or the electrode can break. In the refining stage, the metal is in the liquid state and the disturbances will be smaller, because arc length can be maintain constantly.

Electric current in the graphite electrodes will remain constant if the length of the electric arcs is constant [5]. The relationship between the arc length and the threshold arc voltage is presented in [6].

In the reference literature are illustrated several control strategies, these being executed through the electrode positioning system that consists of hydraulic actuators used to move the electrodes up or down [7–12].

In [7] is presented an arc furnace electrode control system based on a fuzzy self-tuning PID controller and the set-point parameter is a-phase current, but the process is approximated as a second-order object model. In [8] is proposed a method for power control of the EAFs which is based on the estimation of the instantaneous value of flexible cable inductance and the set-point parameter is the arc voltage. In [9] is presented a fuzzy PI controller tuned with genetic algorithms which is used to control the input energy to a three-phase EAF. In [10] is proposed a neural network control algorithm that can improve the control performance but can be applied to the relatively stable working-state because the data contains disturbances and neural network has the convergence problem. In [11] is proposed a fuzzy PID control strategy that can shorten the melting period and reduce the energy consumption, but the process in neglected in the control system. In [12] are presented adaptive control algorithms but these algorithms had not considered the input dead zone nonlinear characteristic, implying a steady-state error and possible instability of the system in EAF.

This paper is an extended version of our paper presented in [13]. In [13] was proposed a fuzzy control and a conventional proportional derivative control for the electrode positioning system of a three-phase EAF. A new model of the electric arc used to simulate the electric arc behavior was also proposed in order to model the process. Proposed control strategies were used to regulate the phase current of the EAF and a comparison of the performance analysis of both control strategies was also highlighted when was applied a single step disturbance in the process and the reference was kept constant. The comparison was analyzed taking into consideration the responses of the systems, which consists of an electric arc current variation, for a step disturbance. In the extended version reported in this paper are performed new tests in order to notice the performance of both proposed control strategies. For all the tests are illustrated the applied disturbance, the output of the execution element when disturbance is occurring, the controller output for both control strategies, the variations of the electric arc length, the response of the systems and the voltage-current characteristic of the electric arc for both of the proposed control strategies. In order to test the control systems are used three cases for a constant value of the reference parameter: (i) introducing of a step disturbance in the process at a moment of time of the system functioning (Sect. 4.1), (ii) introducing of a random sequence of positive or negative step disturbances in the process (Sect. 4.2), (iii) introducing of a second step disturbance in the process before being compensate the effect of another step disturbance in the case of multi-step disturbances, so, before the response of the systems reaching the stable state (Sect. 4.3). Also are extended the related paper conclusions.

In this paper is used the model of the electric arc proposed in [13] by the authors of this paper. The model represents more accurately the voltage-current characteristic (VIC) of the electric arc. It is necessary to model the behavior of the electric arc, because it is the main cause of the nonlinearity of the EAF [15]. This model is used than in the closed-loop system to regulate the electric arc current.

The two control strategies: fuzzy based intelligent control and PD control proposed in [13] are compared in order to notice which is better to use for the real installation.

2 Modeling of the EAF Electrical System

The electrical system of an electrode positioning system can be compound by two models. One model is for the modeling of the high current of the electric arc and one model for the modeling of the three-phase electrical supply system of the EAF.

2.1 Modeling of the Electric Arc

The VIC of the electric arc is the method used for the modeling of the statically and dynamically behavior of the electric arc for the three-phase EAF. In Fig. 1 is presented the VIC of the EAF that will be obtained with the proposed model. In this model, the VIC was divided in five sections in order to obtain a characteristic similar with the one from the real installation. Because the EAF is supplied with alternative voltage, in each period the voltage and the current are crossing two times by 0. Each time that voltage and current change its polarities, the arc will be reignite.

In Fig. 1 v_{ig} represents the ignition voltage, v_{ex} is the extinction voltage, v_m is the average between the ignition and the extinction voltage of the arc, v_{arc} is the voltage of the electric arc, i_1 is the current of the electric arc corresponding for the v_{ig}, i_2 is the current of the electric arc corresponding for the v_m, i_3 is the current of the electric arc corresponding for the v_{ex} and i_{arc} is the current of the electric arc [14]. R_1, R_2 and R_3 represents lines slope for the linear section of the VIC.

The first section of this characteristic is in the range $[-i_3, i_1]$ and electric current derivative is positive. The second section is in the range $[i_1, i_2]$ and electric current derivative is positive. The third section is in the range $[i_2, i_4]$ and electric current derivative is positive. The fourth section is greater than i_4 and electric current derivative is positive and the fifth section is greater than i_3 and electric current derivative is negative. These sections will be also for the negative semi-period of the electric parameters but the range will be different because the values of the electrical parameters are negative.

In Fig. 1 the arrows illustrate how will be designed the voltage-current characteristic of the electric arc. Red arrow illustrates the route when the electric arc current derivative is positive and black arrows illustrate the route when the electric arc current derivative is negative.

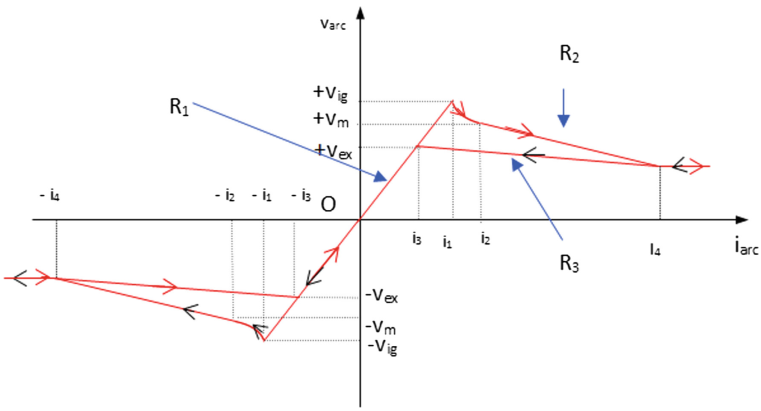


Fig. 1. Voltage-current characteristic of the electric arc divided by sections. (Color figure online)

In (6) is presented the model of the electric arc that is used in the closed-loop system. This model accurately represents the VIC of the electric arc. In the model from (6) were chosen values for the model parameters in order to accurately represent the VIC of the electric arc, $R_1 = 0.05$, $R_2 = -0.0007$ and $R_3 = -0.0003$. Limit i_1 is computed as presented in (1), limit i_2 is computed as presented in (2), limit i_3 is computed as presented in (3), limit i_4 is computed as presented in (4) and voltage v_m is computed as presented in (5).

$$i_1 = v_{ig}/R_1 \tag{1}$$

$$i_2 = 2 \cdot i_1 \tag{2}$$

$$i_3 = v_{ex}/R_1 \tag{3}$$

$$i_4 = (v_m - v_{ex} + i_3 \cdot R_3 - i_2 \cdot R_2)/(R_3 - R_2) \tag{4}$$

$$v_m = (v_{ig} + v_{ex})/2 \tag{5}$$

$$v_{arc} = \begin{cases} -v_{ex} + (i_3 - i_4) \cdot R_3, & i < -i_4 \text{ and } di/dt \geq 0 \\ & \text{or } i < -i_4 \text{ and } di/dt < 0 \\ -v_{ex} + (i + i_3) \cdot R_3, & i \in [-i_4, -i_3) \text{ and } di/dt \geq 0 \\ R_1 \cdot i, & i \in [-i_3, i_1) \text{ and } di/dt \geq 0 \\ & \text{or } i \in [-i_1, i_3) \text{ and } di/dt < 0 \\ v_{ex} + (v_{ig} - v_{ex}) \cdot e^{(i_1 - i)/i_2} & i \in [i_1, i_2) \text{ and } di/dt \geq 0 \\ v_m + (i - i_2) \cdot R_2 & i \in [i_2, i_4] \text{ and } di/dt \geq 0 \\ v_{ex} + (i_4 - i_3) \cdot R_3 & i > i_4 \text{ and } di/dt \geq 0 \\ & \text{or } i > i_4 \text{ and } di/dt < 0 \\ v_{ex} + (i - i_3) \cdot R_3 & i \in [i_3, i_4] \text{ and } di/dt < 0 \\ -v_{ex} + (v_{ex} - v_{ig}) \cdot e^{(i_1 + i)/i_2} & i \in [-i_2, -i_1] \text{ and } di/dt < 0 \\ -v_m + (i + i_2) \cdot R_2 & i \in [-i_4, -i_2] \text{ and } di/dt < 0 \end{cases} \tag{6}$$

2.2 Experimental Validation of the Model

The measured data was acquired from a three-phase EAF of 120 t using a data acquisition board that allows to simultaneously acquire data on 6 channels (3 channels for currents and 3 channels for voltages) for the low or medium voltage lines of the furnace transformer. The data acquisition was performed during the entire steel elaboration period. The acquisition frequency was 5 kHz and the data was acquired during 250 ms this implying that the signals were acquired during 12.5 periods.

In order to validate the model presented in (6), it is performed a comparison (see Fig. 2) for the simulated data and the measured data from the real installation taken into consideration to study.

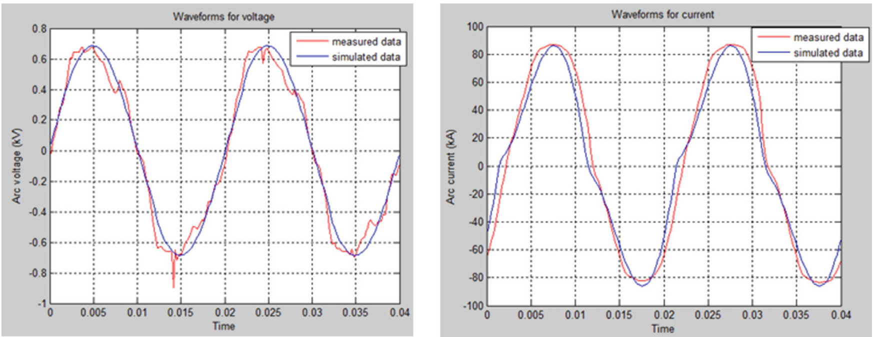


Fig. 2. Waveforms for voltage and current – measured and simulated data.

2.3 Modeling of the Three-Phase Electrical Supply System Circuit of the EAF

The power supply electric circuit of an EAF for a single phase can be considered as the one presented in Fig. 3a. This electric circuit is used in practical computations. The EAF has a resistive character, so, can be considered as a variable resistance. In Fig. 3a v_s represents the alternative voltage from the secondary side of the furnace transformer for a single phase, I is the electric current for a single phase in the secondary side of the furnace transformer, r is the total electrical resistance for a single phase of the secondary side of the furnace transformer, x is the total electrical reactance for a single phase of the secondary side of the furnace transformer and R_{arc} stands for the resistance of the electric arc, being reported at the voltage from the primary side of the furnace transformer.

In Fig. 3b is presented the block diagram for the implementation of the electric arc furnace in Matlab/Simulink. This block diagram was implemented by applying second Kirchhoff's theory in the electric circuit from Fig. 3a. In this study were chosen the following values of the circuit parameters: $r = 0.47$ mΩ, $x = 5.5$ mΩ. Voltage from the secondary side of the furnace transformer was set as considered that the transformer has plot set on tap *16* for which corresponds a line voltage of 894 V. These values are taken from a real industrial plant.

Total electrical resistance is obtained by adding the values of the source resistance and of the short network resistance. Total electrical reactance is obtained by adding the values of the source reactance and of the short network reactance.

Frequency considered of the circuit is 50 Hz. For the source sample time was set to 1/10000 s, meaning that are taken *10000 samples* in 1 s. *ARC* block contains the mathematical model presented in (6).

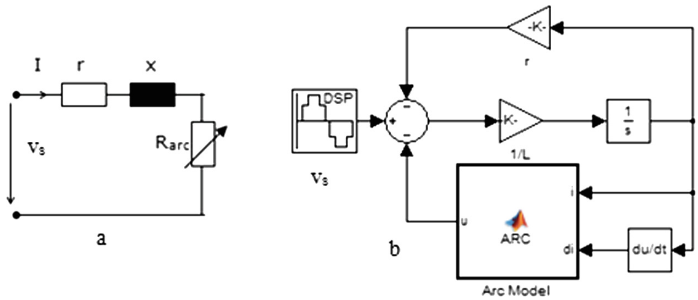


Fig. 3. (a) The power supply electric circuit of an EAF for a single phase and (b) Block diagram of the EAF system.

3 Control Strategies of the EAF

In this paper are proposed, designed and implemented two control strategies. One control strategy is based on the fuzzy intelligent controller and the other is based on the conventional PD controller. In both strategies, the control variable used in this paper is current of the electric arc and these control strategies are used to obtain constant arc length. In order to obtain the performance of the proposed control loops, the systems are tested when applying disturbances in the process in order to observe if the controller will reject these disturbances and which of the controllers has better performance.

In Fig. 4 is presented the block diagram of the proposed control closed-loop. In this block diagram I_{ref} is the reference arc current, e_I is the error computed as the difference between the reference value and obtained value of the arc current, d_{eI} is the derivative of the error, c is the command for the hydraulic actuator in order to move the electrode up or down with a corresponding speed, x is a value that will be added in the position of the arc in this way obtaining the value of the reference current, because arc length influence the electric arc current. I is the current of the electric arc obtained by the model of the electric arc. *Controller block* stands for the fuzzy logic controller (FLC) or the PD controller.

Electric arc block is the process in the control loop and the hydraulic actuator is the execution element. In both of the control strategies, the hydraulic actuator is considered as having a transfer function corresponding to a *PT1*. Chosen time constant for both of the control strategies is 0.75 s.

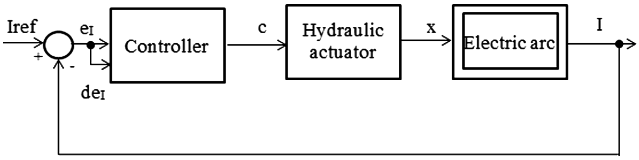


Fig. 4. Block diagram of the proposed control closed-loop.

3.1 Fuzzy Logic Controller of the EAF

FLC is suitable for systems that present nonlinearity and strong random disturbances [15]. In this paper was design a PD-like fuzzy controller that has two inputs and one output. The parameter of the EAF system under control is the current of the electric arc. The inputs of the FLC are the error and the derivative of the error, and the output of the controller is the command for the hydraulic actuator that will modify the position of the electrode, in this way, obtaining a new arc length, so, a new arc current. The design of the fuzzy inference structure for the FLC is illustrated in Fig. 5. The fuzzy inference method used is a *Mamdani-type*, so, the output membership functions used are distributed fuzzy sets. *Min-Max inference engine* is used for the inference and the *centroid* of a two-dimensional function is used for the defuzzification process. Membership functions for all the three variables of the fuzzy system are divided into seven fuzzy sets. These fuzzy sets are: *NL, NM, NS, Z, PS, PM* and *PL* which corresponds to *Negative Large, Negative Medium, Negative Small, Zero, Positive Small, Positive Medium* and *Positive Large*. Trapezoidal membership functions are used for *NL* and *PL* and triangular for the rest. The variables can be positive of negative depending on the obtained current.

Figure 6 presents the membership functions for the error of the system, this being an input variable. The universe of discourse for this variable is in the range [−1700; 1700] A. Figure 7 presents the membership functions for the derivative of the error, this being an input variable. The universe of discourse for this variable is in the range [−3e+06; 3e+06]. Figure 8 presents the membership functions for the command to the hydraulic actuator, this being an output variable. The universe of discourse for this variable is in the range [−40; 40].

The fuzzy system has two inputs and one output, each of the variables having seven fuzzy sets, so, are written *49 rules* which are describing the command for the hydraulic actuator taking into account the values of the inputs of the fuzzy system. These rules are presented in Table 1.

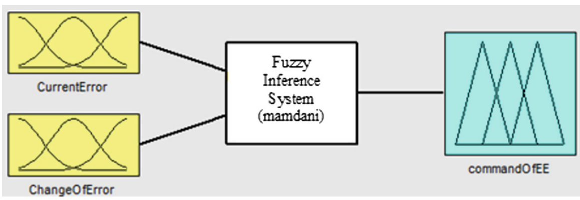


Fig. 5. Designing of the fuzzy inference system for the FLC.

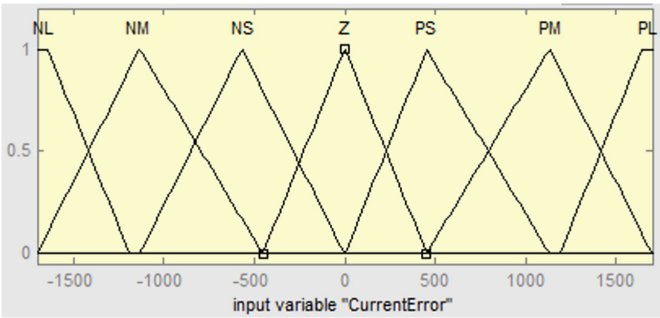


Fig. 6. Membership functions for the input variable "*CurrentError*".

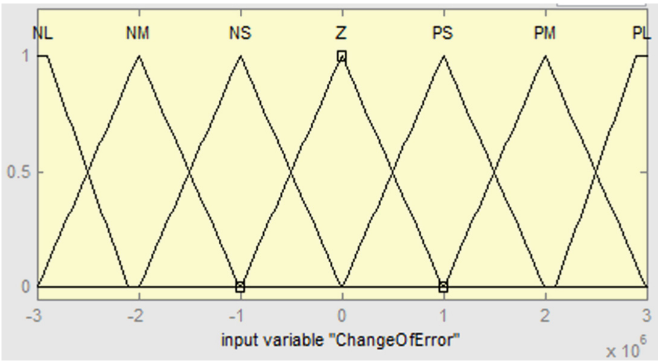


Fig. 7. Membership functions for the input variable "*ChangeOfError*".

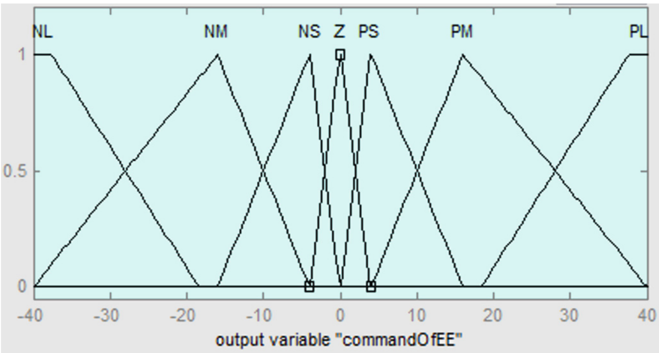


Fig. 8. Membership functions for the output variable "*CommandOfEE*".

In Fig. 9 is presented the Matlab/Simulink model for the FLC. In the model of the electric arc it is set an initial length of the arc of 31 cm, which will correspond to the value of the reference parameter. Value that will be obtained from the hydraulic actuator will be added to the initial arc length. Figure 9 illustrates the case when the disturbance is injected in the process. Value of the reference current is set to 56500 A.

Table 1. Fuzzy rules base for developing fuzzy inference system for the FLC.

c	d_{eI}						
e_I	NL	NM	NS	Z	PS	PM	PL
NL	PL	PL	PL	PL	PM	PS	Z
NM	PL	PL	PL	PM	PS	Z	NS
NS	PL	PL	PM	PS	Z	NS	NM
Z	PL	PM	PS	Z	NS	NM	NL
PS	PM	PS	Z	NS	NM	NL	NL
PM	PS	Z	NS	NM	NL	NL	NL
PL	Z	NS	NM	NL	NL	NL	NL

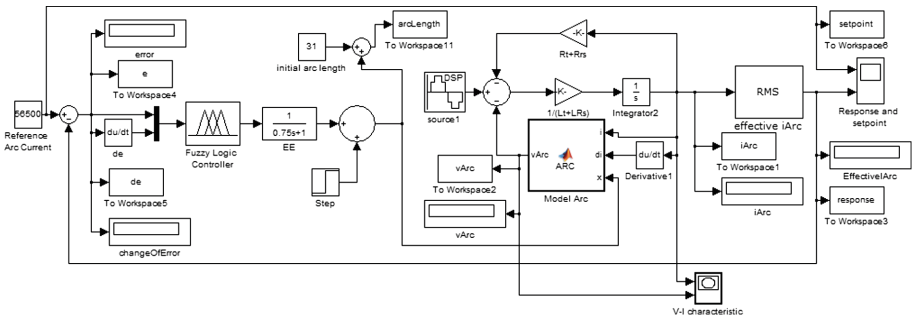


Fig. 9. Matlab/Simulink model for the FLC when is injected disturbance in the process.

3.2 PD Controller of the EAF

In this paper is designed a PD controller in order to make a comparison between the two proposed control strategies because the FLC is PD-like fuzzy controller. In Fig. 10 is presented the Matlab/Simulink block diagram for PD controller for the case when is injected a step disturbance in the process. So, PD controller is used for the rejection of this disturbance. This PD controller has two gains, one for the proportional component and one for the derivative component.

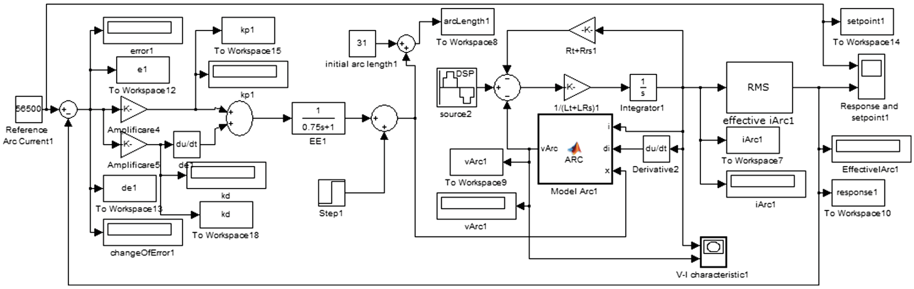


Fig. 10. Matlab/Simulink model for the PD controller when is injected disturbance in the process.

4 Comparisons of the Two Control Strategies for a Constant Value of the Reference

In order to test the control systems are used three cases for a constant value of the reference parameter. First case is when is injected a step disturbance in the process at a specific moment of time, the second case is when is injected a sequence of step disturbances in the process at randomly moments of time and the third case is when applying a second step disturbance before being compensate the effect caused by another step disturbance. In all the cases the sample time for the supply source was set to 0.0001 s, so, in one second will be taken *10000 samples*.

Disturbances which can occur in the real steel melting process are disturbances with permanent character, for example the suddenly movement of the solid charging material in the furnace tank during melting process or the consumption of the electrodes. Therefore, these kinds of disturbances can be simulated by means of step signals rather than impulse signals which do not correspond to a real physical phenomenon in the melting process.

4.1 Introduction of a Step Disturbance in the Process

In order to make a comparison between the responses obtained for the FLC and for the PD controller, after the first second of the system functioning was applied a disturbance of 2 cm (see Fig. 11a), which is corresponding for the real installation with the melting of some parts of the charging material and obtaining a new configuration of the scrap in the furnace's tank. So, arc length is increasing and the arc current is decreasing. The output of the execution element is illustrated in Fig. 11b and the controller output is represented in Fig. 12. One can notice that if the arc length is increased (caused by the step disturbance) the controller output is minimizing the error. The output of the execution element has negative value and will cause the electrode to be moved down in this way obtaining the initial value of the arc length, being rejected the disturbance and being kept constant the length of the electric arc.

In Fig. 13 is illustrated the corresponding variation of the arc length and in Fig. 14 is illustrated the response of the systems for the case when is injected a step disturbance in the process. Figure 15 illustrates the voltage-current characteristic of the electric arc obtained for the two considered controllers.

One can notice that both responses reach the stable state but the response of the system with FLC is faster and follow more precisely the value of the reference parameter as compared to the PD controller. Also, it can be observed that the response of the system with FLC has smaller overshot as compared to the response of the system with PD controller.

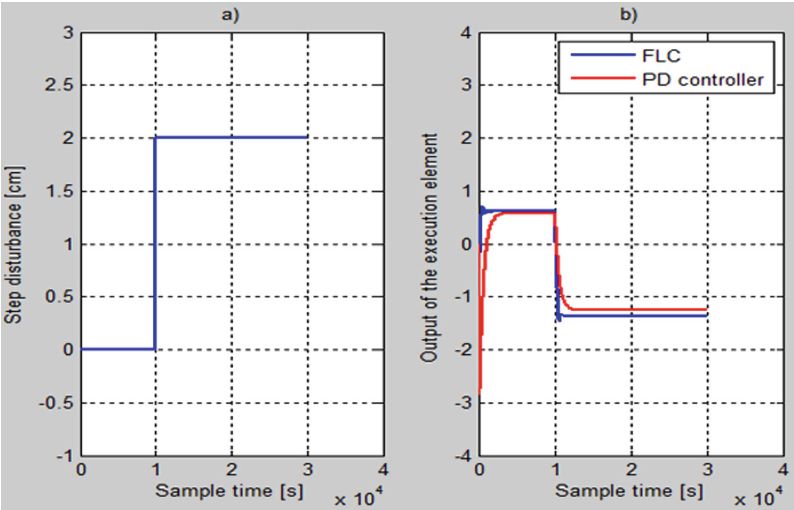


Fig. 11. (a) Step disturbance, (b) Output of the execution element.

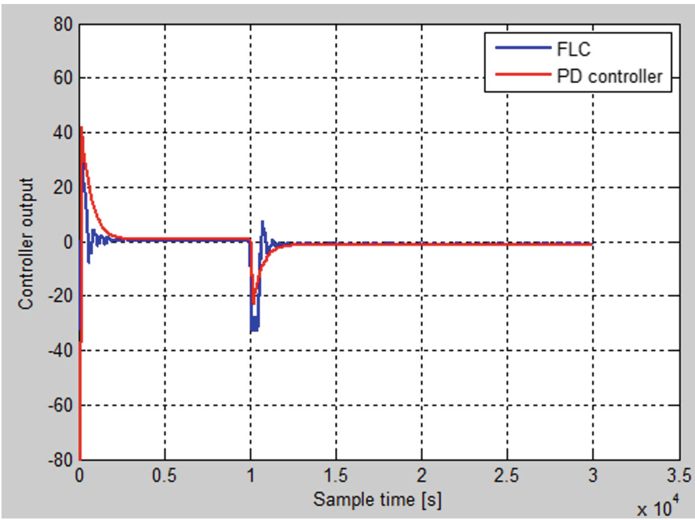


Fig. 12. Controller output for the FLC and for the PD controller.

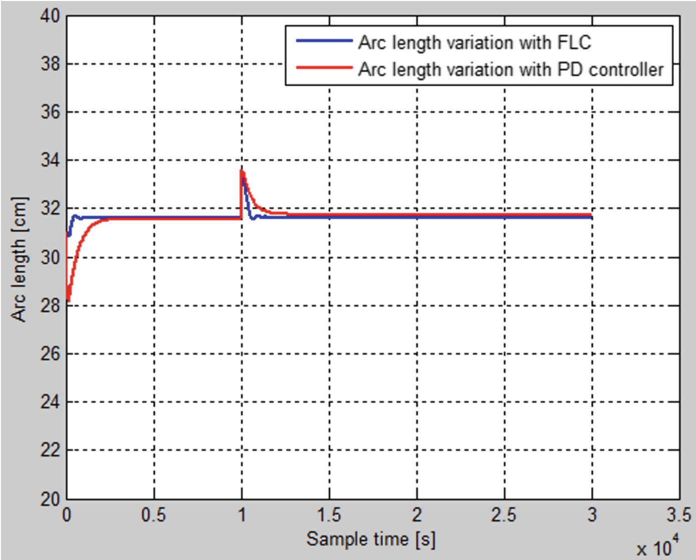


Fig. 13. Corresponding variation of the arc length for both of the controllers.

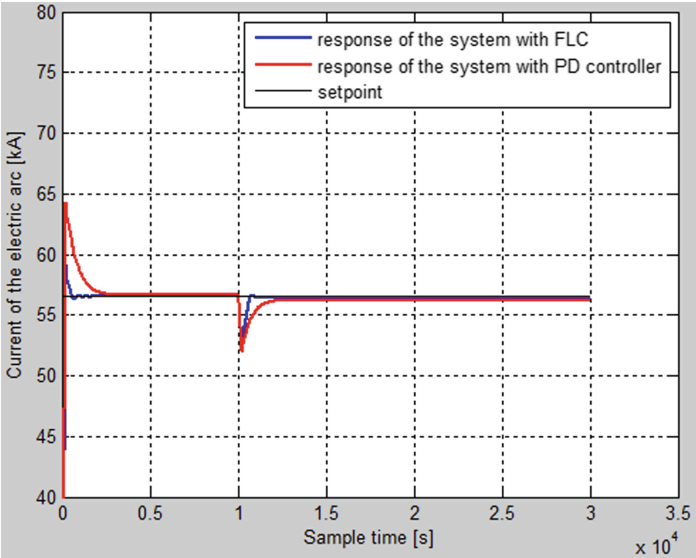


Fig. 14. Set-point and the response variation of the systems when is injected a step disturbance in the process.

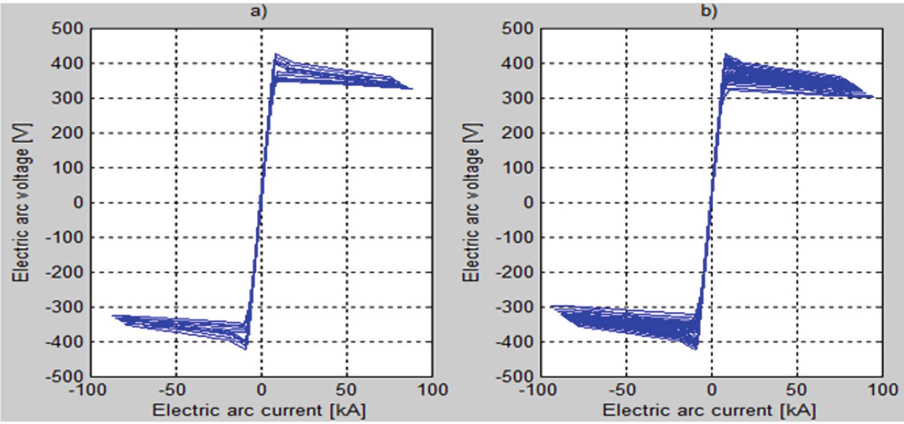


Fig. 15. Voltage-current characteristic of the electric arc obtained for: (a) FLC, (b) PD controller.

4.2 Introduction of a Sequence of Step Disturbances in the Process

The second test corresponds for injection of a sequence of five positive or negative step disturbances in the process (see Fig. 16a). The output of the execution element is illustrated in Fig. 16b and the controller output is represented in Fig. 17. The output of the execution element is proportional to the step disturbances. If the disturbance implies an increasing of the arc length the output of the execution element is having negative values and if disturbance implies a decreasing of the arc length the output of the execution element is having positive values. Therefore the electrode will be moved down or up in this way obtaining the initial value of the arc length, the disturbances are rejected and the length of the electric arc is kept constant.

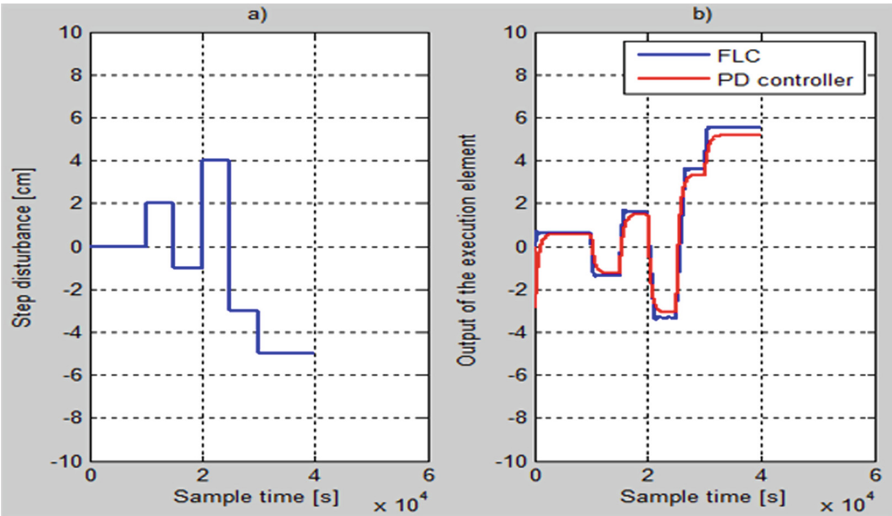


Fig. 16. (a) Sequence of step disturbances, (b) Output of the execution element.

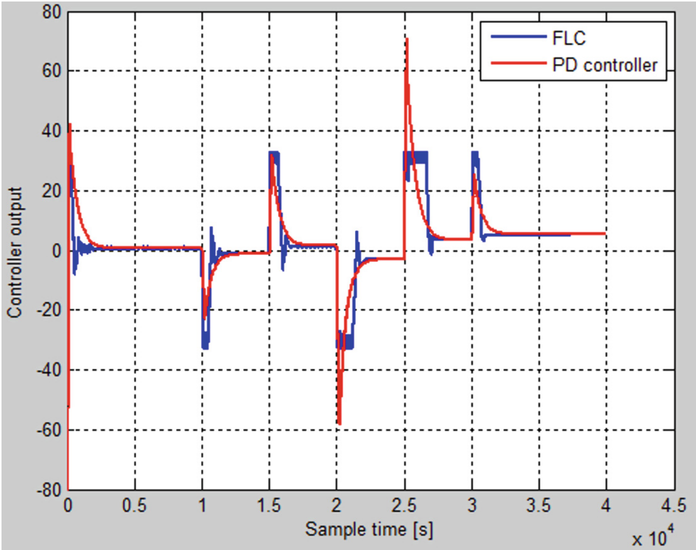


Fig. 17. Controller output for the FLC and for the PD controller.

In Fig. 18 is illustrated the corresponding variation of the arc length and in Fig. 19 is illustrated the response of the systems for the case when is injected a sequence of step disturbances in the process. Figure 20 illustrates the voltage-current characteristic of the electric arc obtained for the two considered controllers.

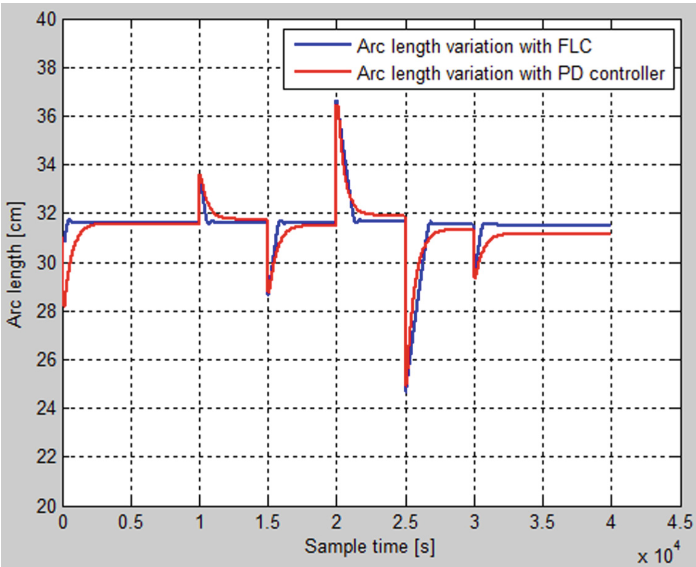


Fig. 18. Corresponding variation of the arc length for both of the controllers.

One can notice that even in this case both responses reach the stable state but the response of the system with FLC is faster and follow more precisely the value of the reference parameter as compared to the PD controller. Also, it can be observed that the response of the system with FLC has smaller overshot as compared to the response of the system with PD controller.

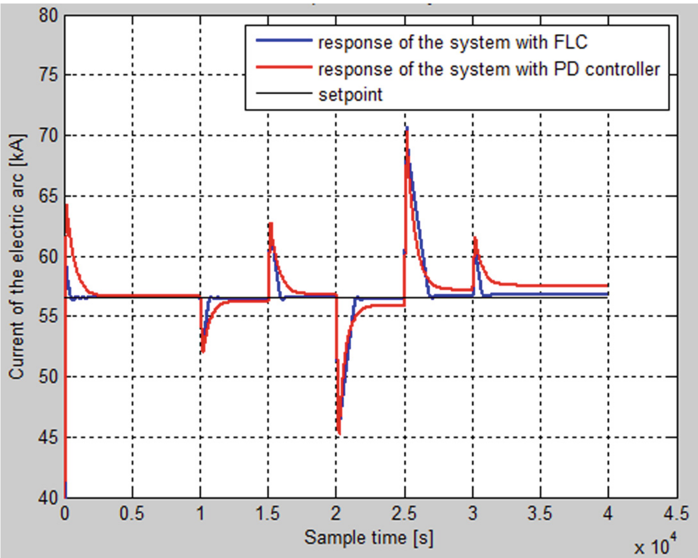


Fig. 19. Response of the systems when is injected a sequence of step disturbances in the process.

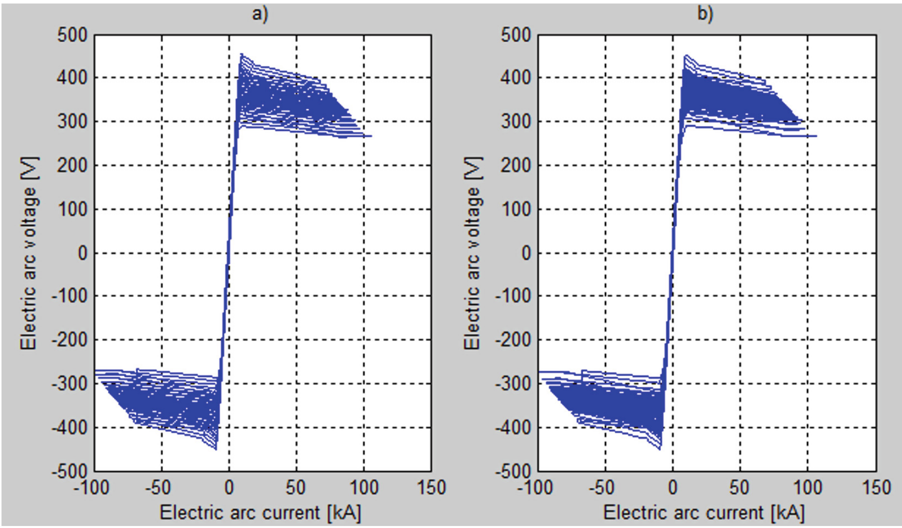


Fig. 20. Voltage-current characteristic of the electric arc obtained for: (a) FLC, (b) PD controller.

4.3 Introduction of a Step Disturbance in the Process Before Being Compensate the Effect of Another Disturbance

The third test is for injected of a second step disturbance in the process before being compensated the effect of another step disturbance, so, before the response of the systems reaching the stable state (see Fig. 21a). In this case are taken into consideration the following situations: applying of a positive step disturbance on another positive step disturbance, applying of a negative step disturbance on another negative step disturbance and applying of a positive step disturbance on another negative step disturbance.

The output of the execution element is illustrated in Fig. 21b and the controller output is represented in Fig. 22.

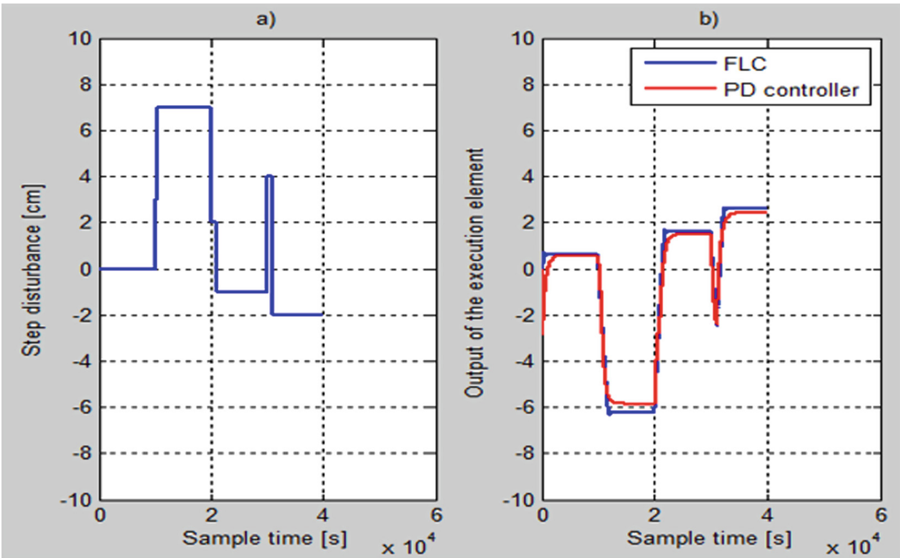


Fig. 21. (a) Injection of a step disturbances in the process before being compensate the effect of another disturbance, (b) Output of the execution element.

In Fig. 23 is illustrated the corresponding variation of the arc length and in Fig. 24 is illustrated the response of the systems for the presented case. Figure 25 illustrates the voltage-current characteristic of the electric arc obtained for the two considered controllers.

One can notice that also for this case both responses reach the stable state but the response of the system with FLC is faster and follow more precisely the reference parameter as compared to the PD controller. Also, it can be observed that the response of the system with FLC has smaller overshot as compared to the response of the system with PD controller.

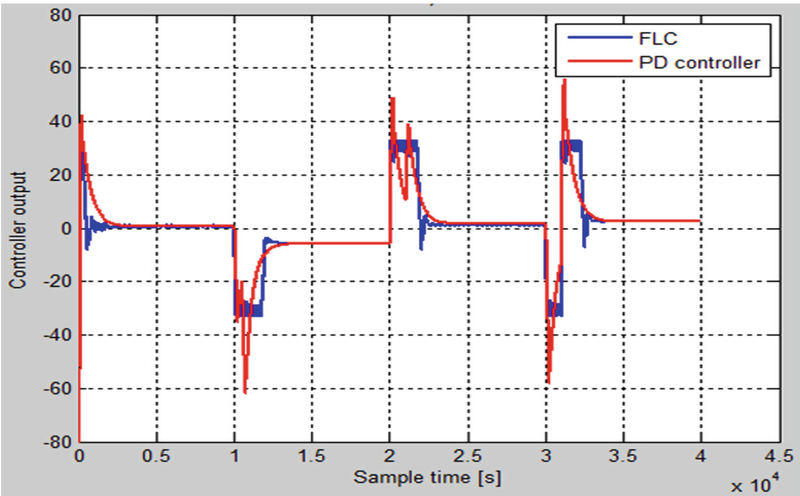


Fig. 22. Controller output for the FLC and for the PD controller.

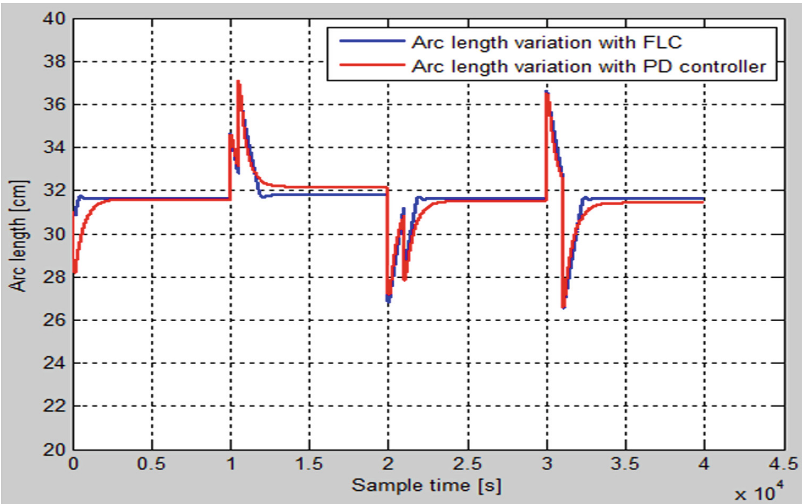


Fig. 23. Corresponding variation of the arc length for the FLC and for the PD controller.

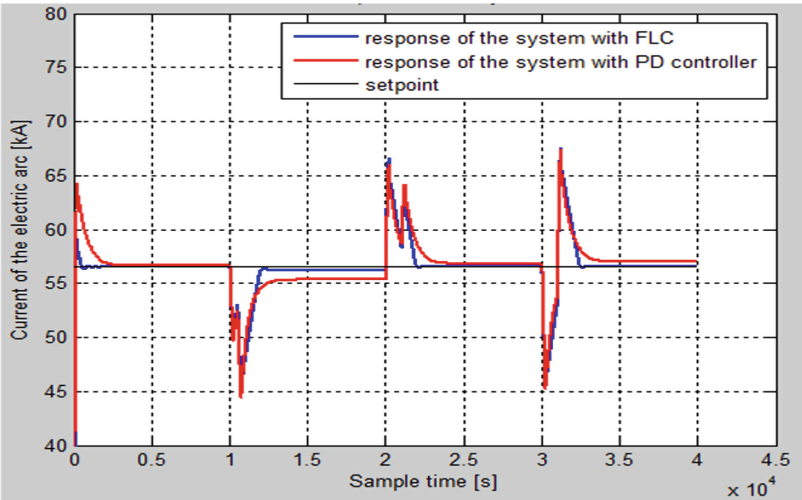


Fig. 24. Set-point and response variation of the systems when is applied a second step disturbance before being compensate the effect of another disturbance.

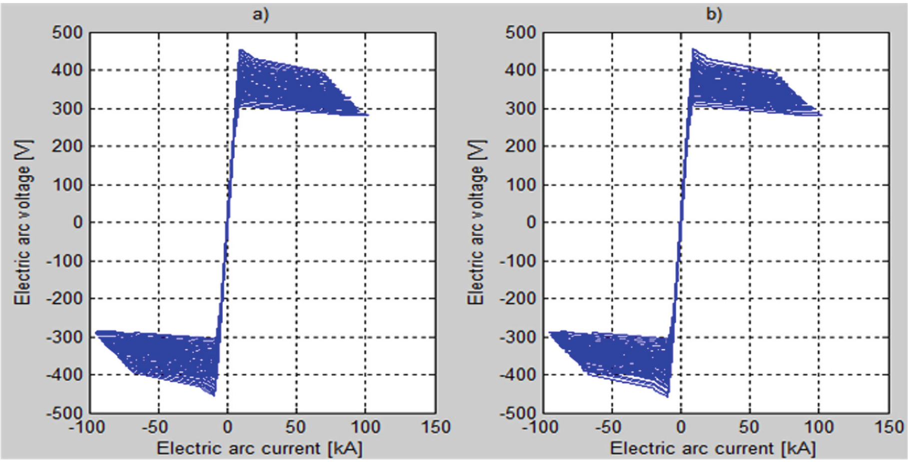


Fig. 25. Voltage-current characteristic of the electric arc obtained for: (a) FLC, (b) PD controller.

5 Conclusions

In this paper is used a new model of the electric arc for an electric arc furnace. The model is based on the voltage-current characteristic of the electric arc and it is an exponential model. In order to validate this model were made comparisons for both simulated data and acquired data from the real installation. One can noticed that the model of the electric arc accurately follows the electric parameters acquired from the real installation. This model is used in the implementation of the two control strategies used in the current control of the electric arc. The systems were tested in the closed-loop for both fuzzy logic and PD controllers. In order to make a comparison of the proposed control strategies were presented three cases. In the first case was considered that was injected a step disturbance in the process at a specific moment of time, in the second case was injected a sequence of step disturbances in the process at randomly moments of time and in the third case was injected a second step disturbance before being compensate the effect caused by another step disturbance.

It was demonstrated that system with fuzzy logic controller has better dynamic performance, rapidity and good robustness as compared to the PD controller. The response of the system is important because by using the electrode system regulation it is reduced the energy consumption and it is avoided the appearance of damages caused by the breakage of the electrodes, therefore, the productivity of steel will be increasing.

Starting from the proposed control strategies in the future works can be developed Matlab/Simulink models for both of the controllers in order to perform a comparison of the two control strategies for variable value of the reference signal.

Also, taking into consideration the presented control strategies can be design other control strategies, for example based on adaptive control such as using a FLC to tune a conventional controller.

References

1. Taslimian, M., Shabaninia, F., Vaziri, M., Vadhva, S.: Fuzzy type-2 electrode position controls for an electric arc furnace. In: 2012 IEEE 13th International Conference on Information Reuse and Integration (IRI), pp. 498–501. IEEE (2012)
2. Deaconu, S.I., Topor, M., Popa, G.N., Popa, I.: Comprehensive analysis for modernization of 100 t electric arc furnace for steel production. In: Industry Applications Society Annual Meeting, IAS 2009, pp. 1–6. IEEE (2009)
3. Hong, H., Mao, Z.: Controller design for electrode regulating system of electric arc furnace. In: 2015 27th Chinese Control and Decision Conference (CCDC), pp. 864–867. IEEE (2015)
4. Zheng, T., Makram, E.B.: An adaptive arc furnace model. IEEE Trans. Power Deliv. **15**(3), 931–939 (2000)
5. Brown, P., Langman, R.D.: Simulation of closed-loop energy control applied to arc furnaces. Iron Steel Inst. J. **205**(8), 837–847 (1967)
6. Montanari, G.C., Loggini, M., Cavallini, A., Pitti, L., Zaninelli, D.: Arc-furnace model for the study of flicker compensation in electrical networks. IEEE Trans. Power Deliv. **9**(4), 2026–2036 (1994)
7. Li, X., Lu, X., Wang, D.: Arc furnace electrode control system design. In: 2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE), vol. 4, pp. 364–366. IEEE (2010)
8. Khoshkhoo, H., Sadeghi, S.H.H., Moini, R., Talebi, H.A.: An efficient power control scheme for electric arc furnaces using online estimation of flexible cable inductance. Comput. Math. Appl. **62**(12), 4391–4401 (2011)

9. Moghadasian, M., Alenasser, E.: Modelling and artificial intelligence-based control of electrode system for an electric arc furnace. J. Electromagn. Anal. Appl. **3**, 47–55 (2011)
10. Zhang, X.: An embedded system to be applied to neural network predictive control in an electric arc furnace. In: ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Long Beach, p. 75. American Society of Mechanical Engineers (2005)
11. Babulu, K., Kumar, D.K.: Fuzzy self-adaptive PID controller design for electric heating furnace. Int. J. Eng. Invent. **1**(5), 10–21 (2012)
12. Liu, X., Zhang, D.: Research on discrete-time adaptive control of electrode regulator system of arc furnace. J. Syst. Simul. **16**(4), 634 (2004)
13. Ghiormez, L., Prostean, O., Panoiu, M., Panoiu, C.: Fuzzy logic and PD control strategies of a three-phase electric arc furnace. In: Nguyen, N.-T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds.) ICCCI 2016. LNCS (LNAI), vol. 9875, pp. 509–519. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45243-2_47
14. Hui, Z., Fa-Zheng, C., Zhuo-qun, Z.: Study about the methods of electrodes motion control in the EAF based on intelligent control. In: 2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE), vol. 4, pp. 68–71. IEEE (2010)
15. Yan, L.I., et al.: Model predictive control synthesis approach of electrode regulator system for electric arc furnace. J. Iron. Steel Res. Int. **18**(11), 20–25 (2011)

Bionic Hand Control in Real-Time Based on Electromyography Signal Analysis

Martin Tabakov[1(✉)], Krzysztof Fonal[1], Raed A. Abd-Alhameed[2], and Rami Qahwaji[2]

[1] Department of Computational Intelligence, Wroclaw University of Science and Technology, Wrocław, Poland
martin.tabakow@pwr.edu.pl, 168132@student.pwr.edu.pl
[2] School of Electrical and Computer Science, University of Bradford, Bradford, UK
{r.a.a.abd, r.qahwaji}@bradford.ac.uk

Abstract. In this paper fuzzy type-1 and type-2 models for control of bionic hand in real-time are proposed. The control process involves interpretation and analysis of surface electromyography signal (sEMG) acquired from patients with amputees. In the provided experiments, we assume the use of force sensing resistor to achieve better control of the bionic hand being modeled. The classical type-1 Mamdani control as well as the extended type-2 fuzzy models are considered for this application. The conducted experiments show comparable results with respect to applied assumptions that give the confidence to implement the proposed concept into real-time control process.

Keywords: Fuzzy control · Type-1 fuzzy control · Type-2 fuzzy control
Bionic hand · Electromyography · Signal analysis

1 Introduction

The theoretical background and hardware implementations of bionic limbs has widely been expanded in recent years in the field of medical applications [6, 17, 27] and robotics domain [15] as a result of the available electronic components with sufficient computing power with a very low cost. The process of such applications applied fuzzy logic in the design of robotic systems, as it is capable to handle a non-linear complex systems [2, 14, 24, 29]. Commonly, in the research concepts of bionic limbs engineering design, surface electromyography signals are the most often used control source for bionic limbs [6, 25, 37]. The research ideas have evolved so quickly that even neuro-controlled bionic arms, able to allow an amputee to move prosthetic arm as if it is a real limb directly by thinking, have been introduced [21, 26]. Studies were directed towards the introduction of non-invasive easy mountable solutions, able to intercept brain signals from many channels and later classify activities from specific channels revealing intention of activity [10]. Nevertheless, most of the research areas consider the problem of bionic limbs control, as a classification problem, i.e. if a certain input conditions are recognized (sensor information and sEMG values) then implemented mechanisms based on machine learning are applied [3, 5, 23]. Next, classifiers

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXIX, LNCS 10840, pp. 21–38, 2018.
https://doi.org/10.1007/978-3-319-90287-6_2

are used for functionality recognition. Unfortunately, from patient perspective, there is a problem with such a concept, namely muscle contraction related to certain functionality should be entirely executed. These concepts do not give the patient the sensing of real control over the bionic hand if muscles other than arm muscles are involved in the control process. Most of the research and existing solutions focus only on lower arm amputations which creates visible gap when it comes to upper arm amputation. Therefore, in our interpretation of real-time signal analysis, we focused on a proposal of sEMG signal analysis that is applicable in fuzzy control models, for real-time performance which should give the patient the sensing of control over the bionic hand irrelevant to the muscle used. In the model presented, we assume as a control target hand grasping, with respect to object resistance, i.e. the aim is to reduce the speed of the device used for hand grasping functionality with the increase of potential object resistance in real-time: during the sEMG signal acquisition.

The paper is organized as follows: Sect. 2 introduces background information about the fuzzy control models applied. Section 3 explains the basic steps of a sEMG processing. Section 4 presents the basic research idea proposed. Sections 5, 6 and 7 provide results and discussion and finally, Sect. 8 draws conclusion. Additionally, we provide comparison between type-1 and type-2 fuzzy control models applied, by corresponding visualization of the achieved results. The better performance of type-2 fuzzy sets is confirmed in related research fields as well [22].

This paper is an extended version of our previous research [28], published on the ICCCI'2016 international conference proceedings. We have extended significantly our work by adding research and experiments with type-2 fuzzy control model.

2 Basic Notions

In this section, the preliminaries of fuzzy sets [36] and fuzzy control systems of Mamdani type [16] are presented.

2.1 Type-1 Fuzzy Set

A type-1 fuzzy set A consists of a domain X of real numbers together with a function $\mu_A : X \rightarrow [0, 1]$, [36] i.e.:

$$A =_{df} \int_X \mu_A(x)/x, \quad x \in X \tag{1}$$

here the integral denotes the collection of all points $x \in X$ with associated membership grade $\mu_A(x) \in [0, 1]$. The function μ_A is also known as the membership function of the fuzzy set A, as its values represents the grade of membership of the elements of X to the fuzzy set A. The idea is to use membership functions as characteristic functions (any crisp set is defined by its characteristic function) to describe imprecise or vague information. This possibility along with the corresponding defined mathematical apparatus, initiated a number of applications. Assuming discrete domain, the basic set operations: union and intersection of two fuzzy sets A and B, are defined as follows: $\forall_{x\in X}\ \mu_{A\cup B}(x) =_{df} max\{\mu_A(x),\ \mu_B(x)\}$, $\forall_{x\in X}\ \mu_{A\cap B}(x) =_{df} min\{\mu_A(x),\ \mu_B(x)\}$.

2.2 Type-1 Fuzzy Logic Controller

Figure 1 shows the schematic diagram of a type-1 fuzzy controller. The main idea is that all input information are fuzzified and then processed with respect to the assumed knowledge base, inference method and the corresponding defuzzification method.

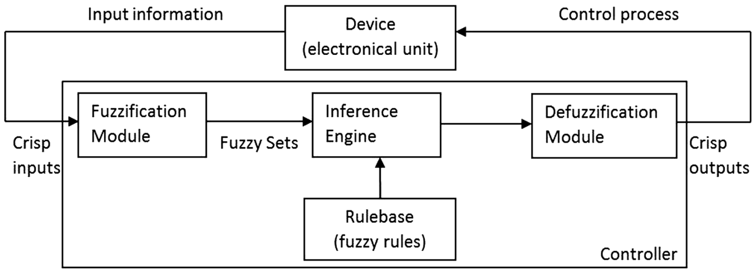


Fig. 1. Information flow within a typical type-1 fuzzy controller

Let consider the rule base of a fuzzy logic controller consisting of N rules which take the following form:

$$R^n :_{df} \text{ IF } (x_1 \text{ } is \text{ } X_1^n) \text{ o } \ldots \text{o } (x_i \text{ } is \, X_i^n) \text{ THEN } y \text{ } is \text{ } Y_n \tag{2}$$

where $X_i^n(\mathrm{i} = 1, \ldots, \mathrm{I};\ \mathrm{n} = 1, \ldots, \mathrm{N})$ are fuzzy sets defined over corresponding domains and Y_n is an output information, which in the Mamdani model [16] is assumed as fuzzy set as well, defined over some domain Y. The binary operator 'o' is the *t*- or *s-norm* ($\mathrm{o} \in \{\otimes, \oplus\}$; $\otimes, \oplus : [0,\ 1]^2 \rightarrow [0,\ 1]$) which have the commutative, associative and the monotonic properties, and have the constants 1 and 0 as unit elements, respectively. In fuzzy logic, the t-norm operator provides the characteristic of the AND operator, while the s-norm provides the characteristic of the OR operator [1].

Assuming an input vector $\bar{\mathbf{x}} = \{x_1^{'}, x_2^{'}, x_3^{'}, \ldots, x_i^{'}\}$, typical computations of a fuzzy system consist of the following steps:

(1) Compute the membership grades of $x_i^{'}$ on each X_i^n, $\mu_{X_i^n}(x_i^{'}), \mathrm{i} = 1, \ldots, \mathrm{I}$; $\mathrm{n} = 1, \ldots, \mathrm{N}$
(2) Compute the firing value of the n^{th} rule, $F^n(\bar{\mathbf{x}})$:

$$F^n(\bar{\mathbf{x}}) =_{df}\ \mu_{X_1^n}(x_1^{'}) \text{ o} \ldots \text{o } \mu_{X_i^n}(x_i^{'})\ \in [0,\ 1] \tag{3}$$

(3) Compute defuzzification output. The most common method is the centre of gravity (COG) method with assumed relation between the premise and the conclusion of the fuzzy rules as the *min* operator:

$$Y_{COG}(\bar{\mathbf{x}}) =_{df} \frac{\sum_{y \in Y} \mu_{\bigcup_n Y'_n}(y) \cdot y}{\sum_{y \in Y} \mu_{\bigcup_n Y'_n}(y)} \quad (4)$$

where

$$\forall_{y \in Y}\, \mu_{Y'_n}(y) =_{df} \; min\{f^n, \mu_{Y_n}(y)\}, \; \mathrm{n} = 1, \ldots, \mathrm{N} \quad (5)$$

The output value is directly related to the control process.

2.3 Type-2 Fuzzy Set

The possibility of introducing fuzzy sets and control systems along with the corresponding mathematical apparatus defined, initiated a number of applications. But still, researchers have shown some limitations of the introduced theory, which mainly reflects the assumption that membership grades are crisp values [7, 8, 19, 32, 35]. Therefore, the introduced theory has been extended, by the introducing of type-2 fuzzy set concepts, characterized by membership functions that are fuzzy themselves.

Generally, a type-2 fuzzy set, denoted below as $\tilde{A}$, is characterized by a type-2 membership function [19] $0 \leq \mu_{\tilde{A}}(x, u) \leq 1$, where $x \in X$ and $u \in J_x \subseteq [0, 1]$, i.e.:

$$\tilde{A} =_{df} \{((x, u), \mu_{\tilde{A}}(x, u)) |\, \forall x \in X, \forall u \in J_x \subseteq [0, 1]\} \quad (6)$$

$\tilde{A}$ might be also introduced as follows:

$$\tilde{A} =_{df} \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x, u)/(x, u) \quad J_x \subseteq [0, 1] \quad (7)$$

where $\int\int$ denotes union over all admissible x and u.

Currently, interval type-2 (IT2) fuzzy sets [19], a special case of type-2 fuzzy sets, are the most widely used because of their acceptable computational complexity and easy interpretation.

In this case, the amplitude of $\mu_{\tilde{A}}(x, u)$ equals 1 for $\forall x \in X$ and $\forall u \in J_x \subseteq [0, 1]$.

Uncertainty about $\tilde{A}$ is conveyed by the so called *footprint of uncertainty* (*FOU*) of $\tilde{A}$, defined as follows:

$$FOU(\tilde{A}) =_{df} \bigcup_{x \in X} J_x, \quad J_x \subseteq [0,\ 1] \quad (8)$$

The size of a *FOU* (the corresponding surface) is directly related to the uncertainty that is conveyed by an interval type-2 fuzzy set and what follows, a *FOU* with more area is more uncertain than one with less area.

The *upper membership function* and *lower membership function* of Ã are two type-1 membership functions $\underline{A}$ and $\overline{A}$ that bound the *FOU*, which might be used to describe J_x, i.e.:

$$J_x =_{df} \; [\mu_{\underline{A}}(x),\; \mu_{\overline{A}}(x)] \tag{9}$$

which obviously leads to the following as well:

$$FOU(\tilde{A}) =_{df} \bigcup_{x \in X} [\mu_{\underline{A}}(x),\; \mu_{\overline{A}}(x)] \tag{10}$$

An illustration of the above definitions, is given in Fig. 2.

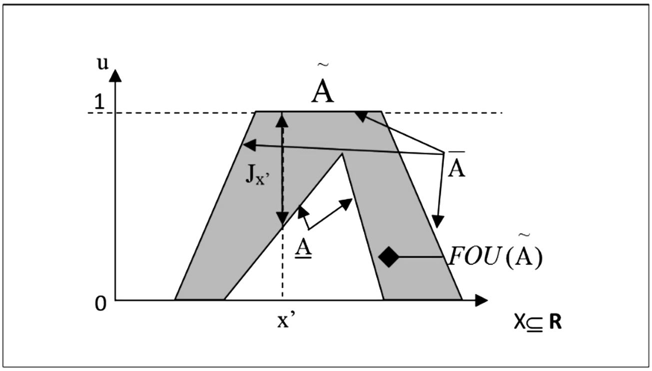


Fig. 2. An interval type-2 fuzzy set Ã.

2.4 Type-2 Fuzzy Logic Controller

Figure 3 shows the information flow within an IT2 fuzzy system (fuzzy controller as a classic example). It is very similar to its type-1 analogue. The major difference or specific is that because of the IT2 fuzzy sets used in the rulebase, the outputs of the inference engine are IT2 fuzzy sets, and a type reducer [11, 18] must be applied to convert them into a type-1 fuzzy sets in order to enable defuzzification procedure.

Below, we give a brief description of the basic steps of the computations in an IT2 fuzzy system [32]. Let consider the rulebase of an IT2 fuzzy system consisting of N rules taking the following form:

$$\tilde{R}^n :_{df} \text{ IF } (x_1 \; is \, \tilde{X}_1^n) \, o \ldots o \, (x_i \; is \, \tilde{X}_i^n) \text{ THEN } y \, is \; Y^n \tag{11}$$

where $\tilde{X}_i^n$(i = 1,…, I; n = 1,…, N) are IT2 fuzzy sets defined over corresponding domains, $Y^n = [\underline{y}^n, \overline{y}^n]$ is an interval which might be assumed as the centroid [12, 18] of the rule consequent and o ∈ {⊗, ⊕}.

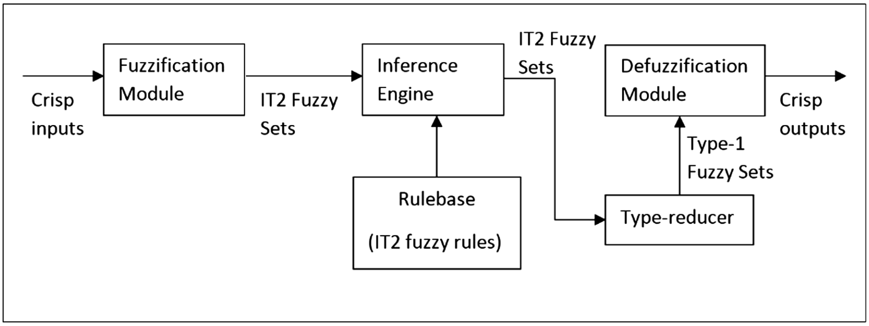


Fig. 3. Information flow within an IT2 fuzzy controller

Assuming an input vector $\bar{\mathbf{x}} = \{x_1^{'}, x_2^{'}, x_3^{'}, \ldots, x_i^{'}\}$, typical computations of an IT2 fuzzy system consist of the following steps:

1. Compute the membership intervals of $x_i^{'}$ on each $\tilde{X}_i^n$, $[\mu_{\underline{X}_i^n}(x_i^{'}),\ \mu_{\overline{X}_i^n}(x_i^{'})]$, i = 1, …, I; n = 1,…, N
2. Compute the firing interval of the n^{th} rule, $F^n(\bar{\mathbf{x}})$:

$$F^n(\bar{\mathbf{x}}) =_{df} \ [\mu_{\underline{X}_1^n}(x_1^{'})\, o \ldots o\, \mu_{\underline{X}_i^n}(x_i^{'}),\ \ \mu_{\overline{X}_1^n}(x_1^{'})\, o \ldots o\, \mu_{\overline{X}_i^n}(x_i^{'}) \equiv [\underline{f}^n, \overline{f}^n] \tag{12}$$

3. Apply type reduction to combine $F^n(\bar{\mathbf{x}})$ with corresponding rule consequents. There are some methods of type reduction [18, 33, 34], but the most commonly used one is the *center-of-sets* (COS) type reducer [18], in the basis of which lies the extension principle [36].

$$Y_{COS}(\bar{\mathbf{x}}) =_{df} \bigcup_{\substack{f^n \in F^n(\bar{\mathbf{x}}) \\ y^n \in Y^n}} \frac{\sum_{n=1}^{N} f^n y^n}{\sum_{n=1}^{N} f^n} = [y_l,\ y_r] \tag{13}$$

where [18, 20, 30]:

$$y_l =_{df} \min_{k \in [1, N-1]} \frac{\sum_{n=1}^{k} \overline{f}^n \underline{y}^n + \sum_{n=k+1}^{N} \underline{f}^n \underline{y}^n}{\sum_{n=1}^{k} \overline{f}^n + \sum_{n=k+1}^{N} \underline{f}^n} \equiv \frac{\sum_{n=1}^{L} \overline{f}^n \underline{y}^n + \sum_{n=L+1}^{N} \underline{f}^n \underline{y}^n}{\sum_{n=1}^{L} \overline{f}^n + \sum_{n=L+1}^{N} \underline{f}^n} \tag{14}$$

$$y_r =_{df} \max_{k \in [1, N-1]} \frac{\sum_{n=1}^{k} \underline{f}^n \overline{y}^n + \sum_{n=k+1}^{N} \overline{f}^n \overline{y}^n}{\sum_{n=1}^{k} \underline{f}^n + \sum_{n=k+1}^{N} \overline{f}^n} \equiv \frac{\sum_{n=1}^{R} \underline{f}^n \overline{y}^n + \sum_{n=R+1}^{N} \overline{f}^n \overline{y}^n}{\sum_{n=1}^{R} \underline{f}^n + \sum_{n=R+1}^{N} \overline{f}^n} \tag{15}$$

with L and R points determined by the dependencies:

$$\begin{aligned} \underline{y}^L \le y_l \le \underline{y}^{L+1} \\ \overline{y}^R \le y_r \le \overline{y}^{R+1} \end{aligned} \tag{16}$$

assuming $\{\underline{y}^n\}$ and $\{\overline{y}^n\}$ to be sorted in ascending order.

The y_l and y_r may be computed by the *Karnik-Mendel* algorithms [11, 18] or their variants [30, 31].

4. Compute the defuzzified output as:

$$y =_{df} \frac{y_l + y_r}{2} \tag{17}$$

3 Electromyography Signal Processing

Surface electromyography delivers a non-invasive method for the objective evaluation of the electrical activity of the skeletal muscles. It provides information about the functionality of the peripheral nerves and muscles [13]. In clinical research, this information has significant influence to the preparation of relevant procedures for patient rehabilitation. Myoelectric signals refer to the system of voluntary muscle contraction, which motor units are activated at different frequencies and their contributions to the signal are added asynchronously [9]. This signal presents harmonics with frequencies ranging from 15 Hz to about 500 Hz, and amplitude from approximately 50 µV to 5 mV [4].

In our research the sEMG signal is used as the basic information source for control; but, in order to make the raw signal interpretable, the corresponding sEMG signal processing is required. Delsys® Bagnoli™ EMG System for signal acquisition, along with dedicated tool for signal processing and analyses EMGWorks® Software were used. The Bagnoli acquisition device and EMG signal processing software completely cover all the steps necessary for the processing of any raw sEMG signal, as described below:

(a) Acquisition of the signal,
(b) Elimination of the energy power grid utility frequency – this is done by Band Stop filter, depending on the country (in Poland: 50 Hz),
(c) Proper identification of the sEMG signal bandwidth – this is done by Band Pass filter with parameters: 15 Hz–200 Hz (the signal is most intense in this bandwidth),
(d) RMS (Root Mean Square) filtering with time window, applying default options: window length 0.125 and window overlap 0.0625.

In Fig. 4, a sample raw sEMG signal (taken during repetitive rehabilitation exercises) and its processed form are shown, after applying steps (a)–(d).

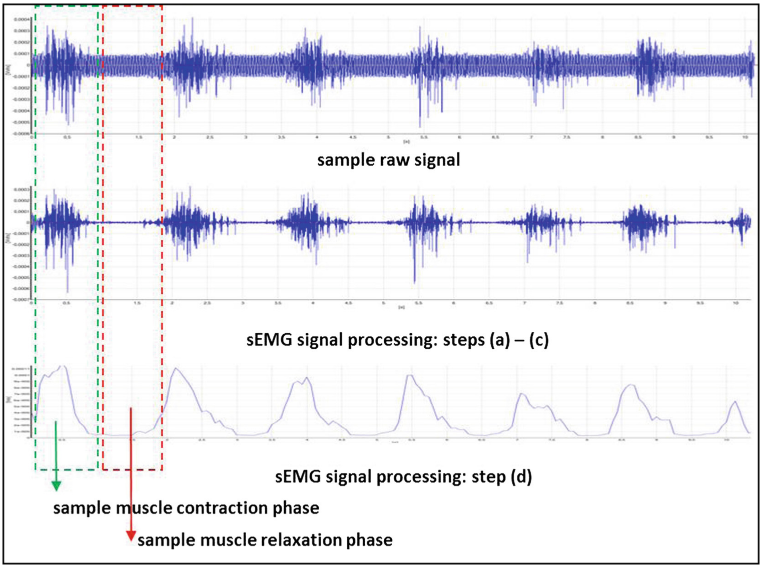


Fig. 4. Processing of a raw sEMG signal

It should be noted to process the raw signal in real-time, a corresponding time window which is slid among the time axis should be applied.

4 Basic Research Concept

The basic concept of the work presented, is to use the sEMG signal given by a human muscle to control a bionic hand. It is assumed the corresponding values from the signal, to be fuzzified and applied in a fuzzy Mamdani control model. In order to achieve this goal, we propose the following analysis of the sEMG signal:

(1) Apply polynomial regression of 3th degree, in order to generate smooth curve (for more details see Sect. 4.1) which represents the real-time sEMG signal, required for control (see Fig. 5). The polynomial, which is differentiable, is marked below with *f*.

(2) Calculate the discrete derivative in each time point *t*. We have applied the forward method of discrete derivative calculation:

$$\frac{df}{dt} =_{df} \frac{f(t + \Delta t) - f(t)}{\Delta t}, \tag{18}$$

where Δt = 0.0625 s. (referring to the sampling rate of the sEMG device used). The derivative is providing the information, about the rate of change of the function values that is related directly to the degree of muscle contraction in any time point.

(3) Fuzzify the derivative values, in order to provide information about the degree of function values change (see Fig. 6).

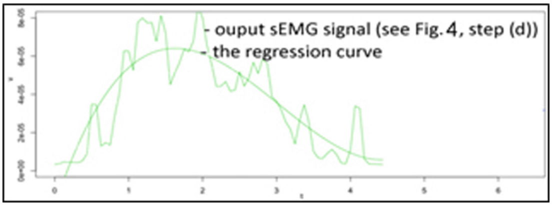


Fig. 5. Polynomial regression of a sample sEMG signal acquired during pectoralis major muscle contraction.

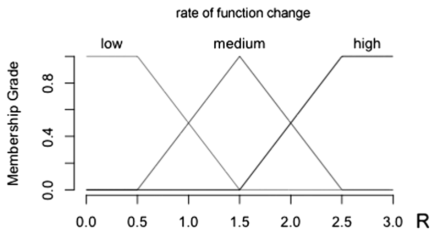


Fig. 6. 'Classical' (three membership functions, representing fuzzy sets: low, medium, high derivative values) fuzzification of the derivatives of *f*.

4.1 Real Time Signal Analysis

Obviously, in order to apply any regression, the whole set of measurements points given should be available. Therefore, the muscle contraction, assumed as an input for the proposed fuzzy control model, could be completed, which makes the real-time control process impossible under these assumptions.

That is why, in order to enable real-time control with regression, it is proposed the following solution of the above mentioned problem:

(1) For any patient with amputee, under a series of trainings (which provide sEMG signals of pectoralis major muscle controlled contractions), build a database of sEMG signals. For all the signals, a polynomial regression is applied which provides an averaged, personalised regression (see Fig. 7). This stage might be considered as a learning process, required for every patient.
(2) Next, in real-time, current regression of a sEMG signal might be constructed using data acquired in real-time in combination with the average regression curve (the remaining points of the curve, see Fig. 8(a)–(c). By mixing the partial sEMG signal with a priori given personalised control function (the average regression), we are able to predict a real time 'control function' (called below as current regression) in real-time.

Therefore, the idea of the real-time control is to use the regression curve which can be built during the contraction of the pectoralis major muscle. This muscle was chosen as it can be easily controlled by the patient.

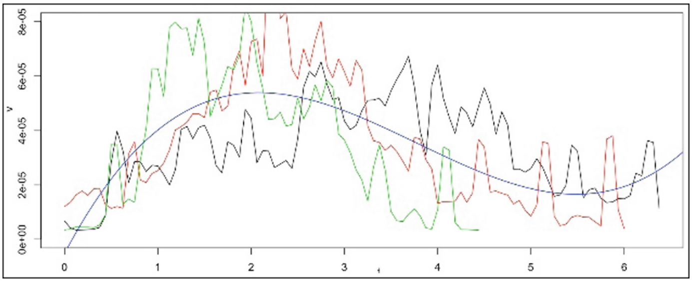
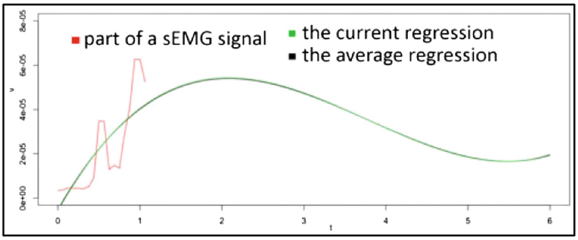
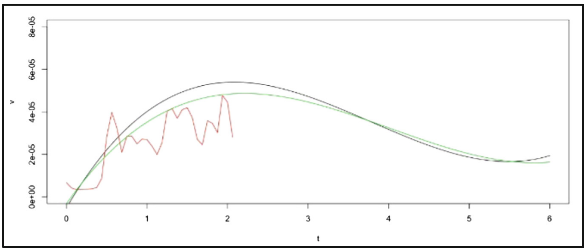


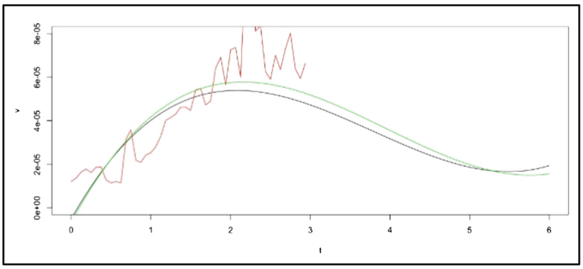
Fig. 7. The average, 'personalised' regression curve



(a) The current regression, which might be used for control, is calculated by combining sEMG signal measured in real-time with the average regression curve, personalised for each patient.



(b) Visualisation nr. 2 (example sEMG signal)



(c) Visualisation nr. 3 (example sEMG signal)

Fig. 8. **(a)–(c)** The process of building the control function (referred as 'current regression'), which is used for real-time control.

4.2 Force Sensing Resistors

Additionally, in order to improve the control process, we propose to use a force sensor resistor. The information given by the sensor is assumed as second information source for the fuzzy controller. The natural processes idea could be stated as follows:

- the relation between degree of muscle contractions and force sensors values should be reversed, i.e. when the patient is increasing the degree of contraction, the force sensor should not give any values (the patient does not grab any object)
- and with the increasing of the applied force on the sensor, the degree of muscle contraction should decrease (the patient is releasing the muscle contraction as object is being grabbed).

In this work, as it is not aimed to present real hardware implemented bionic hand, we have simulated the values of the force sensing resistor, according to the corresponding technical documentation (Interlink Electronics®, Force Sensing Resistor® (FSR). Below, a simple force-to-voltage conversion is presented, for different values of the measuring resistor RM (Fig. 9).

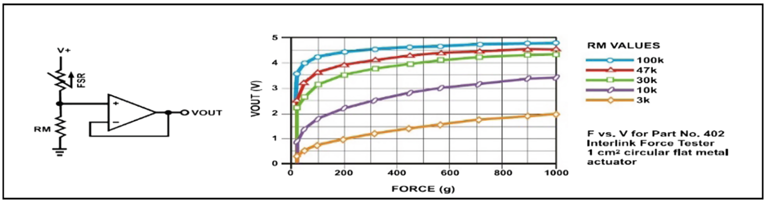


Fig. 9. FSR voltage divider; image source: http://interlinkelectronics.com/integ-guides.php

For the current experiments, a function is assumed for whose values are most closely related to the FSR voltage divider with RM = 10 kΩ. The proposed fuzzification of the FSR data is shown below (Fig. 10).

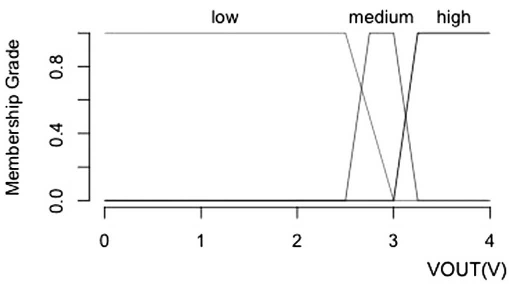


Fig. 10. Fuzzification of the output voltage.

4.3 Type-1 Fuzzy Controller Output

The idea of any control system is based on present of an electronical device, which can be controlled over system assumptions. The velocity of the device might be controlled with the fuzzy controller proposed. The degree of velocity of the device should be related to the degree of bionic hand reaction (degree of hand clump). There might be many such devices that could be designed for this purpose, for example a typical one, often applied in bionic limbs research is the servomechanism.

The proposed model of the device performance and the corresponding fuzzification are shown in Fig. 11, which can easily be related to the degree of performance of the control unit.

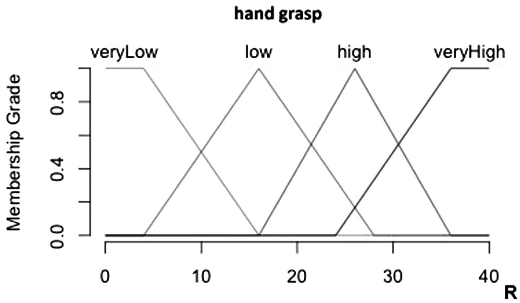
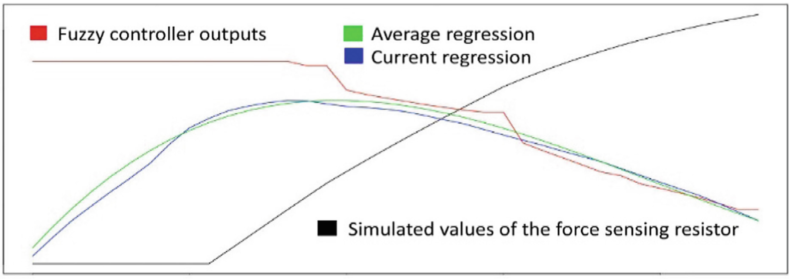


Fig. 11. Fuzzification over values of degree of control unit reaction (changes in the velocity of a hypothetical motor unit).

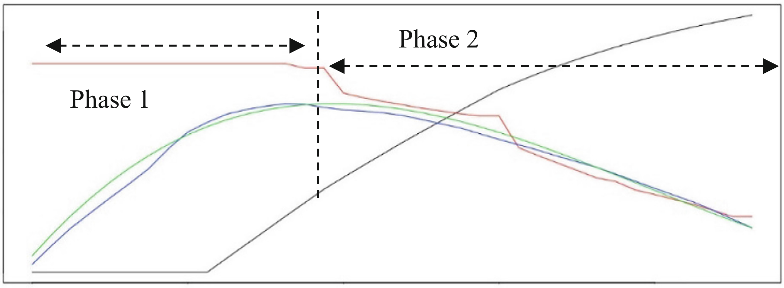
5 Type-1 Fuzzy Control - Experiments and Results

It should be noticed that the data of the force sensing resistor as well as the output device assumed to be controlled, are simulated. The result presents possible solution for further hardware implementation; however, on the other hand, the critical issue here is sEMG signal analysis concept for real-time control. The design of corresponding motor unit based on degree of performance (motor speed) that might be controlled, is trivial. In addition, the values applied for fuzzification are easily to change, in order to achieve optimal performance in real hardware implementation, which will not affect the main concept of the control process proposed.

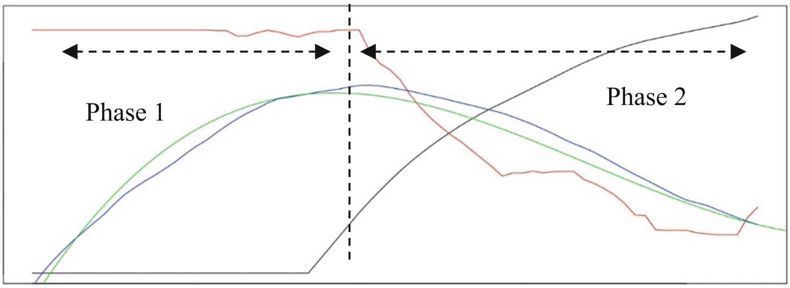
It is meant through these experiments to observe the expected performance on real sEMG signal data. Figure 12(a)–(c) presents the relations between the sEMG signal, acquired from patients with amputees, of the assumed force sensing curve and the outputs of the device proposed, meaning the degree of reaction of a simulated motor unit which controls the hand.



(a) Performance of system output with respect to the system inputs (assuming uniform domain and range of the functions used, for presentation purpose).



(b) Explanation of the presented data: Phase 1: Muscle contraction (increasing muscle tension) / Low values of the force sensing resistor or lack of resistance. Phase 2: Muscle contraction (muscle relaxation) / Rising values of the force sensing resistor.



(c) Visualization of another experiment

Fig. 12. **(a)–(c)** Values of fuzzy controller output: with the increase of the simulated force sensing resistor values along with the decreasing of the muscle tension, the outputs are decreasing which is related to reduction of controlled device reaction (for example, slowing down the controlled unit).

The above experiments, were executed with the following fuzzy rule base (Table 1):

Table 1. The fuzzy rule base used

Rules related to the muscle tension increasing phase
IF (RFCh is low) AND (V_{out} is high) THEN (grasp is low)
IF (RFCh is low) AND (V_{out} is medium) THEN (grasp is veryHigh)
IF (RFCh is low) AND (V_{out} is low) THEN (grasp is veryHigh)
IF (RFCh is high) AND (V_{out} is low) THEN (grasp is veryHigh)
IF (RFCh is high) AND (V_{out} is medium) THEN (grasp is low)
IF (RFCh is high) AND (V_{out} is high) THEN (grasp is veryLow)
IF (RFCh is medium) AND (V_{out} is low) THEN (grasp is veryHigh)
IF (RFCh is medium) AND (V_{out} is high) THEN (grasp is veryLow)
IF (RFCh is medium) AND (V_{out} is medium) THEN (grasp is high)
RFCh – rate of function change, refers to Fig. 6,
V_{out} – output voltage, refers to Fig. 10,
grasp - the degree of reaction of the hypothetical bionic hand, refers to Fig. 11.
Rules related to the muscle tension decreasing phase
IF (RFCh is low) AND (V_{out} is high) THEN (grasp is low)
IF (RFCh is low) AND (V_{out} is medium) THEN (grasp is high)
IF (RFCh is low) AND (V_{out} is low) THEN (grasp is veryHigh)
IF (RFCh is high) AND (V_{out} is low) THEN (grasp is veryHigh)
IF (RFCh is high) AND (V_{out} is medium) THEN (grasp is high)
IF (RFCh is high) AND (V_{out} is high) THEN (grasp is veryLow)
IF (RFCh is medium) AND (V_{out} is low) THEN (grasp is veryHigh)
IF (RFCh is medium) AND (V_{out} is high) THEN (grasp is veryLow)
IF (RFCh is medium) AND (V_{out} is medium) THEN (grasp is low)

6 Type-2 Fuzzy Control - Experiments and Results

Below, the fuzzy control concept proposed, was modified by applying interval type-2 fuzzy sets as a natural extension of the fuzzification used (see Fig. 13). The rule constituents were considered as a priori given intervals, determined on the basis of

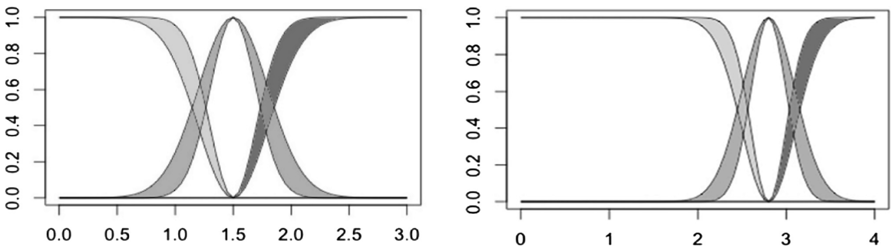
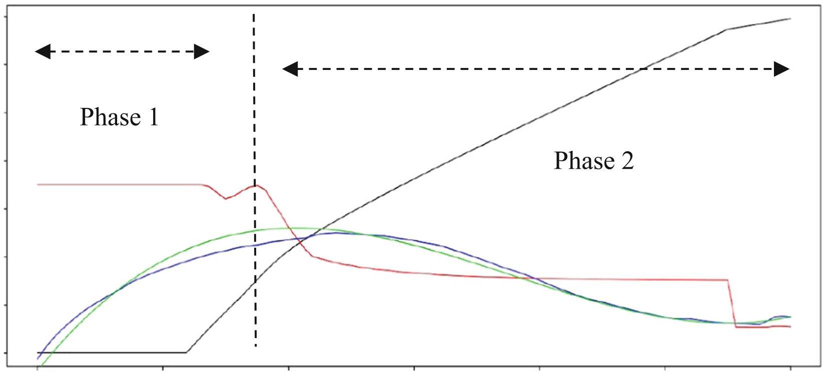
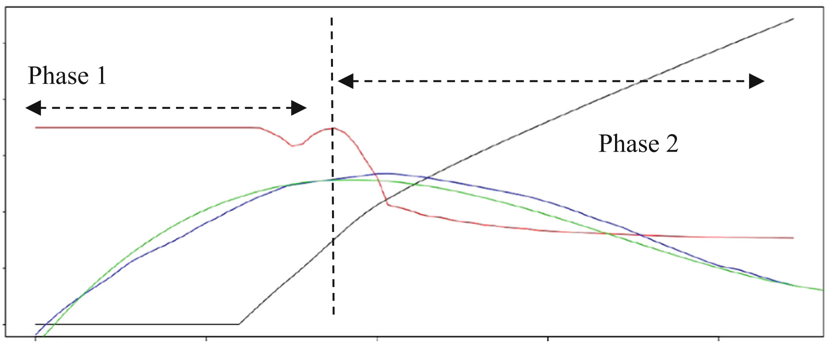


Fig. 13. Extended fuzzification by type-2 fuzzy sets, corresponding to fuzzification shown in Figs. 6 and 10, respectively

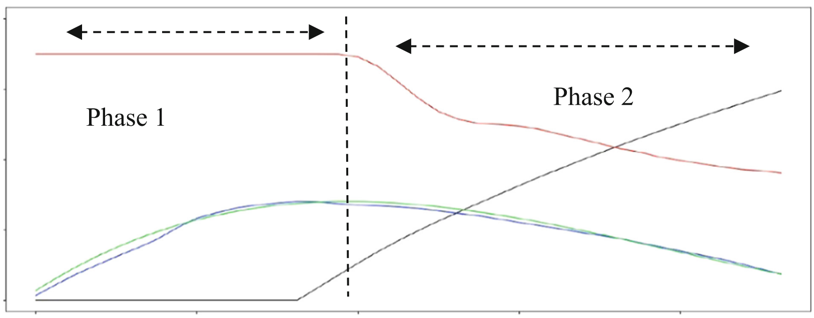
corresponding experiments. The extension proposed, showed significant improvement in terms of simulated fuzzy controller outputs as shown in Fig. 14(a)–(c).



(a) Experiments with type-2 fuzzy control (visualization 1)



(b) Experiments with type-2 fuzzy control (visualization 2)



(c) Experiments with type-2 fuzzy control (visualization 3)

Fig. 14. **(a)–(c)** Smoothening of the fuzzy controller outputs with interval type-2 fuzzy sets

7 Discussion

This work provides a proof of concept that sEMG can be analysed in real-time and gathered data can be applied in a fuzzy control model. The applied regression over sEMG data enables to get control values on ongoing basis. Together with other sensor data, such as force sensing resistors, a control mechanism for bionic hand can be introduced. We have acquired and analysed sEMG signals from two patients with amputees (male, aged 24 and female, aged 13) and we have achieved similar results which confirm the assumed concept. It should be noticed, a simulated model was proposed to proof the design concept and to establish the necessary tools for hardware implementation in terms the simulating information (the force sensing resistor data) as well as applying assumptions for the output values which model the control process. Nevertheless, this is not a critical disadvantage as we are proposing fuzzy model, which are known to be easily adjusted to real implementations. The major idea in this work is the proposal of real-time analysis of sEMG integrated in a fuzzy control model, which is easily implementable. We have tested both, type-1 and type-2 fuzzy control concepts and confirmed better performance, in terms of controller outputs, with the use of type-2 fuzzy sets. The IT2 fuzzy control applied, improved the control process (the smoothness of the curve defined by controller outputs), which will have positive affect on the performance of the device proposed. Our further work will be focused on the hardware implementation of the bionic hand concept proposed.

8 Conclusion

A fuzzy model control of bionic hand based on hand grasp functionality has been proposed and presented. Experiments were provided with type-1 as well as type-2 fuzzy control. The main contribution of this work was provided a method of sEMG signal analysis that able to be applied in real-time process. From patient perspective, it could be concluded, that the real-time sEMG signal analysis which is related to the corresponding actions of the bionic hand gives the feeling of natural hand control. The hardware implementation of the present proposal is in progress by the current authors.

Acknowledgement. This work was supported by the statutory funds of the Department of Computational Intelligence, Faculty of Computer Science and Management, Wroclaw University of Science and Technology and partially by EC under FP7, Coordination and Support Action, Grant Agreement Number 316097, ENGINE European Research Centre of Network Intelligence for Innovation Enhancement (http://engine.pwr.edu.pl/). All computer experiments were carried out using computer equipment sponsored by ENGINE project. Additionally, the authors would like to thank the Wroclaw City Council, for the opportunity to work with the Neuro-Rehabilitation Center for the Treatment of Spinal Cord Injuries 'Akson', under the 'Mozart' city programme 2014/2015.

References

1. Bronstein, I.N., Semendjajew, K.A., Musiol, G., Mühlig, H.: Taschenbuch der Mathematik, p. 1258. Verlag Harri Deutsch, Frankfurt (2001)
2. Chatterjee, A., Chatterjee, R., Matsuno, F., Endo, T.: Augmented stable fuzzy control for flexible robotic arm using LMI approach and neurofuzzy state space modeling. IEEE Trans. Ind. Electron. **55**(3), 1256–1270 (2008)
3. Crawford, B., Miller, K.J., Shenoy, P., Rao, R.P.N.: Real-time classification of electromyographic signals for robotic control. In: National Conference on Artificial Intelligence - AAAI, pp. 523–528 (2005)
4. Cram, J.R., Kasman, G.: Introduction to Surface Electromyography. Aspen Publishing, Gaterburg (1998)
5. Ferran, C., Dominguez, O., Wawrzyniak, Z.M.: Real-time processing of EMG signals for bionic arm purposes. In: Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments, Wilga, Poland (2016)
6. Gauthaam, M., Kumar, S.S.: EMG controlled bionic arm. In: Proceedings of the National Conference on Innovations in Emerging Technology-2011 Kongu Engineering College, Perundurai, Erode, Tamilnadu, India, 17–18 February, pp. 111–114 (2011)
7. Hagras, H.: A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots. IEEE Trans. Fuzzy Syst. **12**, 524–539 (2004)
8. Hagras, H.: Type-2 FLCs: a new generation of fuzzy controllers. IEEE Comput. Intell. Mag. **2**(1), 30–43 (2007)
9. Ielpo, N., Calabrese, B., Cannataro, M., Palumbo, A., Ciliberti, S., Grillo, C., Iocco, M.: EMG-miner: automatic acquisition and processing of electromyographic signals: first experimentation in a clinical context for gait disorders evaluation. In: IEEE 27th International Symposium on Computer-Based Medical Systems, pp. 441–446 (2014)
10. Issa, M.F., Shedeed, H.A.: Brain-EEG signal classification based on data normalization for controlling a robotic arm. Int. J. Tomogr. Simul. **29**(1), 72–85 (2016)
11. Karnik, N.N., Mendel, J.M., Liang, Q.: Type-2 fuzzy logic systems. IEEE Trans. Fuzzy Syst. **7**(6), 643–658 (1999)
12. Karnik, N.N., Mendel, J.M.: Centroid of a type-2 fuzzy set. Inf. Sci. **132**, 195–220 (2001)
13. Konrad, P.: The ABC of EMG: a practical introduction to kinesiological electromyography, Version 1.0. Noraxon Inc., US, April 2005
14. Lea, R.N., Jani, Y., Hoblit, J.: Fuzzy logic based robotic arm control. In: Proceedings of the Second IEEE ICFS, SF, CA, vol. 1, pp. 128–133 (1993)
15. Li, M., Jiang, Z., Wang, P., Sun, L., Ge, S.S.: Control of a quadruped robot with bionic springy legs in trotting gait. J. Bionic Eng. **11**(2), 188–198 (2014)
16. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. Int. J. Man Mach. Stud. **7**(1), 1–13 (1975)
17. Massa, B., Roccella, S., Carrozza, M.C., Dario, P.: Design and development of an underactuated prosthetic hand. In: International Conference on Robotics and Automation (ICRA), Washington DC, 11–15 May, pp. 3374–3379. IEEE (2002)
18. Mendel, J.M.: Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions. Prentice-Hall, Upper Saddle River (2001)
19. Mendel, J.M., John, R.I.: Type-2 fuzzy sets made simple. IEEE Trans. Fuzzy Syst. **10**(2), 117–127 (2002)
20. Mendel, J.M., Wu, D.: Perceptual Computing: Aiding People in Making Subjective Judgments. Wiley-IEEE Press, Hoboken (2010)

21. Pedreira, C., Martinez, J., Quiroga, R.Q.: Neural prostheses: linking brain signals to prosthetic devices. In: Proceedings on the ICROS-SICE International Joint Conference, Fukuoka, Japan, August 2009
22. Malik, O.A., Arosha Senanayake, S.M.N.: An interval type-2 fuzzy logic based classification model for testing single-leg balance performance of athletes after knee surgery. In: Chung, P., Soltoggio, A., Dawson, C.W., Meng, Q., Pain, M. (eds.) Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS). AISC, vol. 392, pp. 85–92. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-24560-7_11
23. Prahm, C., Schulz, A., Paaßen, B., Aszmann, O., Hammer, B., Dorffner, G.: Echo state networks as novel approach for low-cost myoelectric control. In: ten Teije, A., Popow, C., Holmes, J.H., Sacchi, L. (eds.) AIME 2017. LNCS (LNAI), vol. 10259, pp. 338–342. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59758-4_40
24. Precup, R.E., Teban, T.A., Oliveira, T., Petriu, E.M.: Evolving fuzzy models for myoelectric-based control of a prosthetic hand. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Vancouver, BC, Canada (2016)
25. Shenoy, P., Miller, K.J., Crawford, B., Rao, R.P.N.: Online electromyographic control of a robotic prosthesis. IEEE Trans. Biomed. Eng. **55**(3), 1128–1135 (2008)
26. Shekhar, H., Guha, R., Juliet, A.V., Sam, J., Kumar, J.: Mathematical modeling of neuro-controlled bionic. In: International Conference on Advances in Recent Technologies in Communication and Computing, pp. 576–578 (2009)
27. Spanias, J.A., Simon, A.M., Ingraham, K.A., Hargrove, L.J.: Effect of additional mechanical sensor data on an EMG-based pattern recognition system for a powered leg prosthesis. In: 7th International IEEE/EMBS Conference on Neural Engineering (NER), Montpellier, pp. 639–642 (2015)
28. Tabakov, M., Fonal, K., Abd-Alhameed, R.A., Qahwaji, R.: Fuzzy bionic hand control in real-time based on electromyography signal analysis. In: Nguyen, N.-T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds.) ICCCI 2016. LNCS (LNAI), vol. 9875, pp. 292–302. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45243-2_27
29. Tomas, S., Michal, K., Alena, K.: Fuzzy control of robotic arm implemented in PLC. In: IEEE 9th International Conference on Computational Cybernetics (ICCC), Tihany, pp. 45–49 (2013)
30. Wu, D., Mendel, J.M.: Enhanced Karnik-Mendel algorithms. IEEE Trans. Fuzzy Syst. **17**(4), 923–934 (2009)
31. Wu, D., Nie, M.: Comparison and practical implementation of type reduction algorithms for type-2 fuzzy sets and systems. In: Proceedings of IEEE International Conference on Fuzzy Systems, Taipei, Taiwan, pp. 2131–2138, June 2011
32. Wu, D.: On the fundamental differences between interval type-2 and type-1 fuzzy logic controllers. IEEE Trans. Fuzzy Syst. **20**(5), 832–848 (2012)
33. Wu, D.: An overview of alternative type-reduction approaches for reducing the computational cost of interval type-2 fuzzy logic controllers. In: Proceedings of IEEE World Congress on Computational Intelligence, Brisbane, Australia, June 2012
34. Wu, D.: Approaches for reducing the computational cost of interval type-2 fuzzy logic systems: overview and comparisons. IEEE Trans. Fuzzy Syst. **21**(1), 80–99 (2013)
35. Wu, H., Mendel, J.M.: Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems. IEEE Trans. Fuzzy Syst. **10**(5), 622–639 (2002)
36. Zadeh, L.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)
37. Zhuojun, X., Yantao, T., Yang, L.: sEMG pattern recognition of muscle force of upper arm for intelligent bionic limb control. J. Bionic Eng. **12**, 316–323 (2015)

Dynamic Decision Support in the Internet Marketing Management

Paweł Ziemba[1(✉)], Jarosław Jankowski[2], and Jarosław Wątróbski[3]

[1] The Jacob of Paradies University, Teatralna 25, 66-400 Gorzów Wielkopolski, Poland
pziemba@ajp.edu.pl

[2] West Pomeranian University of Technology, Żołnierska 49, 71-210 Szczecin, Poland
jjankowski@wi.zut.edu.pl

[3] University of Szczecin, Mickiewicza 64, 71-101 Szczecin, Poland
jaroslaw.watrobski@wneiz.pl

Abstract. The article deals with the problem of selecting an advertisement variant on the basis of dynamically-changing values of evaluation criteria. Therefore, a framework, used in an online environment, of a dynamic multi-criteria decision analysis (DMCDA) has been prepared. The framework was based on the PROMETHEE method which makes it possible to carry out a very detailed analysis of a decision process and obtained solutions. While applying the prepared framework, a number of ad variants were considered on the basis of the data collected during a subjective study and a field experiment. In the course of solving the decision problem, the advertiser's and website operator's perspectives as well as two aggregation strategies of dynamic data were considered. As a result, the following was obtained: partial rankings of variants, global rankings considering the advertiser's and publisher's points of view, GDSS rankings pointing to compromise solutions by merging the two points of view. The obtained solutions were verified by means of: examining correlation coefficients, a GAIA analysis and an analysis of ranking robustness to preference changes. The end result was that the most satisfying advertiser and publisher were determined.

Keywords: Internet marketing management · DMCDA · PROMETHEE GAIA

1 Introduction

The marketing science and commercial aspects are the base of marketing activity in both online and offline environments. However, especially for the management of online marketing, the technological background and supporting technologies play the key role. They are developed in several directions including campaign planning [1] or real time optimization towards better conversions [2]. New methods are implemented in the area of algorithms for computational advertising including adaptive approaches [3] or linear mathematical models [4] with their extensions [5]. Attempts to increase the effectiveness of online commercial activity often leads to negative side effects such as growing intrusiveness of online marketing content [6] and, as a result, physical or

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXIX, LNCS 10840, pp. 39–68, 2018.
https://doi.org/10.1007/978-3-319-90287-6_3

cognitive avoidance [7]. This is due to the fact that intensive online advertising can affect websites and makes information seeking more difficult and causes an increased cognitive load, frustration and other negative emotions [8]. Due to factors related to cognitive avoidance, the dropping performance of online advertising is observed over time [9]. Searching for compromise between content intrusiveness and its influence on user experience within a web system is one of directions of research in computational advertising area [10, 11].

Presented research is an extended and modified version of earlier work [12]. The approach proposed in this paper integrates data related to effectiveness of online marketing content and its cost together with the evaluation of marketing content intensity. In the proposed approach changes of online environments are taken into account and dynamic decision support with the use of multi-criteria decision analysis (MCDA) method is introduced. Direct application of the MCDA approach in this class of problems is hampered as the MCDA methodology is based on assumptions of stability of parameters forming part of the decision analysis, e.g. datasets, criteria, decision variants and evaluations. In online planning parameters may change dynamically and are conditioned on the changes in audience characteristics, variable efficiency of advertising message or competitors' activity [13]. Employing the classic MCDA approach without considering time evolution and past time periods is the way to oversimplify the problem [14].

In this context the aim of the paper is an adaptation of the MCDA approach for the needs of dynamic multi-criteria decision analysis (DMCDA) in the field of online marketing campaign management with focus on changing effectiveness and online conditions. The solution was verified on the basis of data from real advertising campaign with observed different effectiveness for each time period. In comparison with the paper [12], this paper is based on new data gathered during a field experiment in a real online environment and within a subjective study (surveys). Furthermore, in contrast with [12], the data gathered from the advertiser's and website operator's perspectives are considered in this paper. In fact, a two-level decision problem was considered here in order to find a compromise between effectiveness, intrusiveness, and cost of online marketing content as well as between the advertiser's and website operator's aims. To solve the decision problem, the PROMETHEE GDSS method was applied. Moreover, an analysis of the obtained solution from a descriptive perspective with the use of GAIA method was carried out. Both the applied methods and the considered multidimensional decision problem are a novelty in the paper compared to [12]. The paper is organized as follows: Sect. 2 includes the review of literature, Sect. 3 presents assumptions for the proposed approach. In the next section empirical results are presented followed by a summary in Sect. 5.

2 Literature Review

The development of electronic media and the growing role of online advertising made it possible to measure different results from online marketing campaigns and employ them in a decision process related to planning new campaigns. The main purpose is usually to increase efficiency of marketing activity in multiple dimensions. New metrics are employed in this area for media planning to evaluate online campaigns and are grounded on the direct response and interactions or longer-term influence on brand awareness [15, 16] with an indispensable quantitative approach to media planning [17]. The foundation for running an online campaign within a portal selling advertising space is the scheduling of the use of different creations at different slots. The analysis based on the planned behaviors and site pre-visit intentions are used while estimating potential audiences and the ability to screen proper advertising content [18].

At the operational level, problems deal with the real time campaign optimization [19] and searching for the best method of resource exploitation with the use of stochastic models [20] are discussed. Other areas include the use of available broadcasting resources [21], personalization of message [3] or choice of message content on the basis of context [22]. The basis for implementing marketing activity is planning and scheduling ad expositions with the participation of available broadcasting resources. Plans are implemented using advertising servers which carry out the selection of marketing content as a response to a request coming from the web browser [23]. The problem of the selection of advertisements was formulated as a task of linear programming with maximizing the number of interactions under given constraints which include the number of times an ad was displayed in a given period [2]. The basic model of linear optimization was developed towards a compromise between searching for and exploration of decision-making solutions [5] and a balanced distribution of broadcasts [4]. The approach proposed in [4] identified the probability of diversion from the emission of advertising in the data analysis. The optimum plan of emissions allows the parametrization of an advertising server to obtain the maximum number of interactions in a given period of time. Another approach on the operational level is based on the monitoring a user's activity and maximizing likelihood of interaction [24]. The approach employs user's session tracking as well as maximization of click probability with the solution based on Bayesian models and a generated ranking of advertisements with assigned probabilities. In other solutions the selection of advertising content is based on user profiles created during websites browsing [25]. The approach presented in [1] is founded on a multi-objective model whose main goal is to construct pricing strategies and maximize the revenue from the sale of advertising space and minimize advertisers' costs.

Advertising units used within campaigns frequently examine techniques based on colors, call-to-action messages, persuasion, animations and various layouts or alternating the structure of advertisements in the real time using data about consumer behavior [26–28]. The attempts to increase the performance of online marketing are linked to the growing intensity of online actions as well as negative feed-back from target users [6]. Internet users are more and more overloaded by different pieces of information. Because of their limited ability to process information, Internet users pay attention only to part of the content which is addressed to them [29]. Attempts to attract users' attention are related to the use of the high visibility components [30]. The user's experience is negatively affected by the intensive usage of audio, video and animations within the advertising content [31]. The overall evaluation results of the marketing campaign can be lowered because of the limited cognitive capacity caused by a negative affective response related to irritation and annoyance [32]. Information overload on the web leads to selective online perception and only a limited number of messages are processed while others are ignored. Researchers' and entrepreneurs' interests resulted in different experiments and ways of measuring the intrusiveness.

The literature review shows that majority of the available optimization systems and models is oriented towards increasing the number of interactions within a webpage and automatic selection of advertisements so that it is maximized. Even though maximizing broadcasts of invasive forms of advertising may increase the financial result, in a short while it may also result in the decrease of user experience and negatively affect the brand perception. The solution proposed in this paper integrates parameters related to effectiveness of message and its negative impact on the recipient resulting from the intensity of employed persuasive mechanisms. Taking into account several stages of decision process the solution gives the opportunity to reflect the changeability of preferences and measurement data. The basis of earlier-proposed linear models are Pareto-optimal solutions where bringing tasks down to one function makes it more difficult to take into account decision-makers' preferences and criteria. Therefore, we propose an approach based on the MCDA method. The next section presents assumptions for the proposed solution.

3 Materials and Methods

3.1 PROMETHEE Method

The MCDA methods are widely used in marketing, e-commerce, website evaluation, and online marketing campaigns planning [33–35]. In solving such problems MCDA methods can be applied since they are able to deal with complex decision processes, multiple and conflicting evaluation criteria, different scenarios and decision-makers preferences. A selection of a proper method for a given decision problem is essential, since different methods can produce different solutions of the same problem [36]. The MCDA method used in a decision making in the Internet marketing management, should cope with ranking problematics and should not use an incomparability relation, and consequently, the method can make it possible to achieve a total order in a ranking. Moreover, a selected MCDA method ought to apply indifference and preference thresholds. Consequently, this method would be characterized by non-absolute discriminating power of the criteria. Furthermore, it should be characterized by, at most, a partial degree of compensation of criteria, thus it would fulfil a sustainability concept to a great degree. Measuring data, on which the method will work can be called reliable, since these are objective measures expressed in quantitative scale. Furthermore, it is recommended that the method enables group decision support and makes it possible to broad analysis of the obtained solution. The analysis of individual MCDA methods [37] with relation to the discussed requirements points out that the PROMETHEE II with GAIA (Geometrical Analysis for Interactive Assistance) and GDSS (Group Decision Support System) extensions meet recommended characteristics.

The PROMETHEE II method allows the application of six types of preference functions *P*: a usual criterion, a U-shape criterion, a V-shape criterion, a level criterion, a V-shape with indifference threshold criterion, and a Gaussian criterion [38]. A global preference index of each variant is calculated according to the formula (1):

$$\pi(a,b) = \sum\nolimits_{j=1}^{n} P_j(a,b)w_j \tag{1}$$

where $P_j(a, b)$ means a value of preference function between a and b variants, n is a number of criteria and w_j is a importance assigned to the j-th criterion. $\Pi(a, b) \sim 0$ implies a weak and $\Pi(a, b) \sim 1$ implies a strong global preference of a over b. $P_j(a, b)$, $\Pi(a, b)$ and their opposites are real numbers without units, independent of the scales of the criteria. Their values belong to the range [0, 1]. Positive and negative outranking flows are calculated with the use of the formulas (2) and (3):

$$\Phi^+(a) = \frac{1}{m-1}\sum\nolimits_{i=1}^{m} \pi(a,b_i) \tag{2}$$

$$\Phi^-(a) = \frac{1}{m-1}\sum\nolimits_{i=1}^{m} \pi(b_i,a) \tag{3}$$

where m is a number of variants. The $\Phi^+(a)$ value represents how a variant a is outranking other variants, whereas the $\Phi^-(a)$ value shows how the variant a is outranked by the others. In the PROMETHEE II method in order to construct a total order of variants a net outranking flow described by the formula (4) should be calculated [39]:

$$\Phi_{net}(a) = \Phi^+(a) - \Phi^-(a) \tag{4}$$

The PROMETHEE GDSS method extends the Promethee II functionality with the concept of group decision support. The net outranking flows obtained by individual decision-makers are considered according to the PROMETHEE II method. In the preference table, the decision-makers substitute evaluation criteria. The PROMETHEE GDSS method recommends using a linear preference function with a preference threshold p = 2 [39]. The PROMETHEE method also enables a broad analysis of the results, including a sensitivity/robustness analysis [40], and also provides an analytical tool, GAIA [41].

3.2 Dynamic Multi-criteria Decision Analysis

In a typical decision problems related to Internet marketing management the dynamics and variability of the decision environment is observed. The user preference changes are reflected in changes of evaluation of decision variants. In that kind of decision problems a dynamic approach to multi-criteria decision analysis should be applicable. Dynamic multi-criteria decision analysis allows to expand the classical MCDA paradigm with additional components of the decision making process.

For example, in the paper [42] a comprehensive theory of dynamic multi-criteria decision analysis was introduced. The authors generalized the problem itself and extended the elements of the decision problem by introducing the notions of changeable spaces, a competence set analysis as well as habitual domains. Nonetheless, the majority of articles in the area are focused on developing the classic MCDA model only for chosen aspects of dynamic decision support such as changing sets of evaluation criteria or decision variants. What is more, the choice of the preference aggregation procedure for subsequent time periods is the most frequently discussed issue. The publication [13] discusses a dynamic MCDA model which deals with the problem of performance aggregation of variants in time. This model uses a historical set of variants and admits the changeability of variants in time as well as deleting and adding variants to a set of variants in subsequent examined periods of time. This approach was developed in the work [43] by taking into consideration many experts and considering risks related to a decision problem. The article [44] suggests that the dynamic decision model ought to take into account past, present and future data. Simple Additive Average is proposed to be used as an aggregation operator of the data, however, the authors emphasize that other operators are also permissible, e.g. min, max, Weighted Sum or Hammacher intersection. An approach suggested in [45] is also interesting, since a Data Envelopment Analysis method is proposed to be used as an aggregation operator of data from different periods of time. Additionally, three aggregation strategies were defined, i.e. Time Appreciated Aggregation, where older data have a higher priority over new ones; Time Depreciated Aggregation, in which the latest data are most important; Time Period Mostly Appreciated Aggregation, where data from a selected period of time have the highest weight and the date from preceding and following periods are less vital. The synthesis of other approaches can be found, among other things, in the following papers [14, 46, 47].

3.3 Dynamic MCDA Conceptual Framework

In the proposed approach data from multi-stage online campaign is integrated with measurements of intensity of advertising message and subsequently global and local objectives are employed in the evaluation of results from various perspectives. The proposed framework allows obtaining compromise solutions on the basis of measurements from a real environment and a decision-maker's preferences.

Due to the fact that the dynamic multi-criteria decision analysis is an extension of static approach [13], the conceptual framework is based on the five-stage model of a decision process proposed by Guitouni [48]. It is composed of: (I) the structuring of the decision problem, (II) the preference articulation and modelling, (III) the preference aggregation, (IV) the exploitation of aggregation, (V) the recommendation. The introduction here of dynamics of modelled decision problem requires expanding the classical MCDA approach with the changeability of partial evaluations of decision variants in time as well as the analysis of the influence of this changeability on the final score of the decision process. Additionally, due to the specific nature of online marketing-related issues the framework assumes a constant form of the set of decision variants as well as the family of criteria for their evaluation. It is in line with the specific nature of the considered decision problems.

While structuring the decision problem, decision variants (A) and the evaluation criteria (C) were determined. The variants were advertisements published on the websites. Next, by means of a field experiment in the online environment, data describing criterial performances of variants were gathered. The data were divided into time periods and for every period a performance table was constructed. In the next stage, a preference modelling was carried out, that is preference thresholds and importances of criteria were defined. Here, two perspectives were taken into account: an operator's, of a website on which advertisements were published, and an advertiser's. A preference aggregation was conducted with the use of, presented in Part 3.1 of this article, the PROMETHEE II method. The preference aggregation was conducted separately for every time period. The obtained variant performances and rankings were aggregated in a global evaluation by means of weighted average, according to the formula (5):

$$\Phi_{global}(a) = \frac{\sum \Phi_{net}^{k}(a) * r_k}{\sum r_k}, \quad k = 1 \ldots t \tag{5}$$

where: r_k – weight of a k-th period of time, $\Phi_{net}^{k-}(a)$ – net outranking flow of a variant a in k-th period, Φ_{global} (a) – overall outranking flow of a variant a on the basis of t periods of time. The research presented in the paper assumes a specified influence of each performance table on the final result of performance evaluation. In particular, the first strategy (Time Period Equal Aggregation, TPEA) assumed a balanced influence of data from each time period and in the second strategy (Time Period Depreciated Aggregation, TPDA) the latest data are more preferred than the older ones [37]. The final preference aggregation, consisting in combining performances obtained for the perspectives of the website operator and the advertiser, was carried out by means of the PROMETHEE GDSS method. In the next stage, an analysis of the obtained solution was conducted and a recommended advertising variant was indicated. The DMCDA framework formulated in this way is presented in Fig. 1.

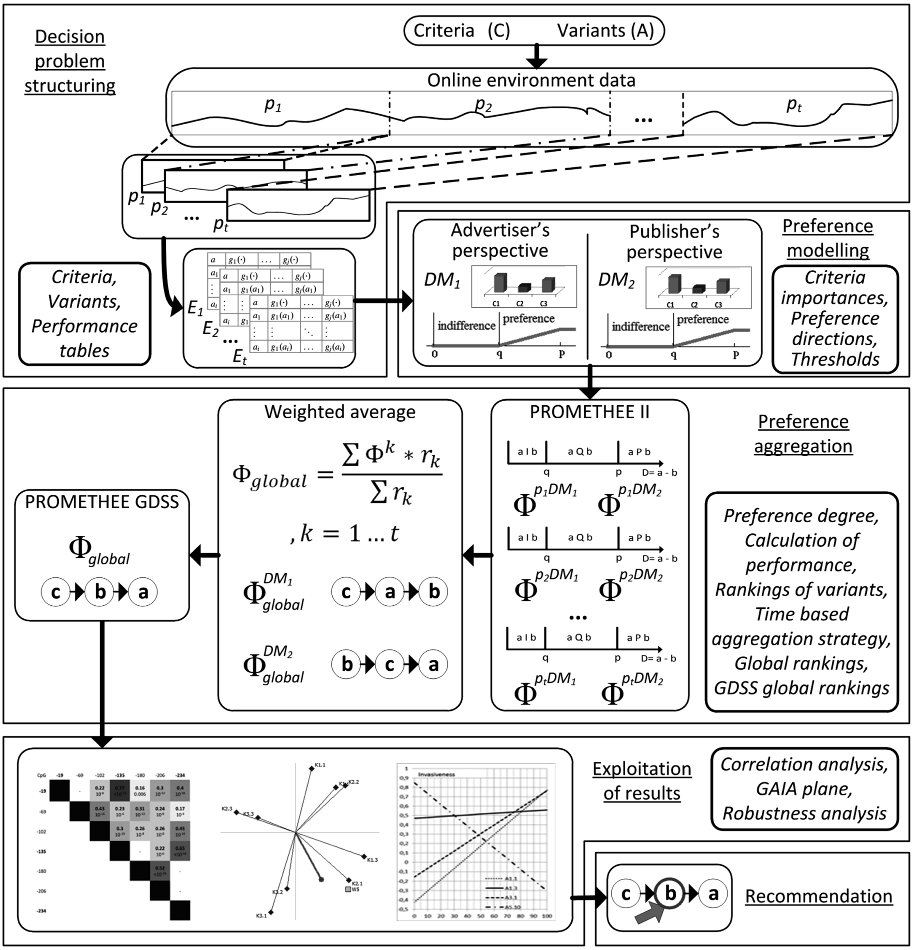


Fig. 1. The dynamic MCDA framework

4 Results

4.1 Structuring of the Decision Problem

The decision problem, dealt with in the research, considered selecting an advertising variant characterized by the highest total performance in a dynamically changing online environment. When structuring the decision problem, a set of decision variants (A) and criteria (C) was defined, with regard to which they were considered. The set of decision variants contained 50 ad units (A1.1–A5.10) published on a website. They were ads for 5 products (A1–A5), for each of them 10 alternative ad units were considered which differed in the level of intensity. The set of criteria consisted of three elements: C_1 - click-through rate (CTR), C_2 - advertiser's cost (AC)/publisher's profit (PP), C_3 - intensity of advertising content (IoA).

The values of variants for the criterion C_1 (click-through rate) were determined as a ratio of the number of clicks in an ad unit to the number of displays of the ad unit. The data were collected during the field experiment consisting in conducting an experimental advertising campaign in the real online environment. Individual advertisements were displayed on a popular website with a varied frequency. During every visit on the website, users saw, apart from its contents, one advertisement out of A1.1–A5.10. The data obtained during the research were of the quantitative nature and they concerned a summary number of projections of a given advertisement in a time period. These are the data to which the owner of the website and the advertiser have access. Most often, they have technical possibilities of collecting such data in a reliable way, since the conversion rate referring to the action of purchasing goods after entering the advertiser's website is based on the website's cookies which can be turned off by the user in his or her browser. Such data collection is 'transparent' to the user and it does not require him or her to take any additional action. Alternatively, one can collect declarative data from users via questionnaire research, however, such an approach is troublesome. First of all, such questionnaire research on a website is more invasive than click tracking. Furthermore, declarative data can significantly differ from users' real behaviour [49]. The automatic detection of users' behavior, used in the research and based on the analysis of real movement and interaction by click tracking, provides more realistic data, unlike questionnaires reflecting subjective and declarative views.

On the basis of the same data, the values of variants for the criterion C_2 (advertiser's cost) were also determined. Here, a cost per click (CPC) model, in which the advertiser pays for every click in an ad unit, was applied. An individual cost of clicking for given ad units of each of the products was different and started from $0.05 for the first (least invasive) ad unit (Ax.1) up to $0.5 for the tenth (most invasive) ad unit of a given product (Ax.10). The costs were determined on the basis of the data and the order of the Amazon company [50]. The total cost of each ad unit was calculated as a product of an individual cost and the number of clicks. However, it should be noted that an advertising cost incurred by the advertiser can be treated also as a profit of a website operator.

The values of individual ad units for the criterion C_3 (intensity of an ad unit) were determined in a subjective study on the basis of surveys. The subjective study was carried out before launching the experimental advertising campaign. Of course, as it has been emphasised earlier, the declarative data can be faulty. However, in the case of a criterion C_3 there is no other possibility of collecting data. In the perceptual experiment, visual advertisements were rated by users between the ages of 20 and 68. Fourteen respondents completed a benchmarking experiment based on the forced choice method, whose main task was to read the text and select which of the two adverts more negatively affected the task being performed and the dispersed attention. For additional reliability, all observers repeated the experiment three times, collecting 42 repetitions. During the online experiment, portal users were not investigated and their demographic characteristics were unknown.

It is important to note that the applied criteria are not the only potential categories of information useful in this type of research. However, most often the advertiser and the website owner have technical possibilities of collecting these kinds of information. Also, during an advertising campaign, they make decisions, first of all, on the basis the data reflected by the criteria [51]. Moreover, criteria C_1 and C_3 allow us partially to notice the user's point of view. The criterion C_1 indirectly provides information whether the advertisement or the advertised product is attractive for the user to enter the advertiser's website by clicking. On the other hand, the criterion C_3 gives a general view concerning the invasiveness of a given advertisement in the user's subjective opinion.

Owing to the fact that the research concentrated on presenting the dynamics of the online environment in time, the field experiment was carried out in four even time periods $p_1, \ldots, p_4$. For all the time periods, the values of decision variants for criteria C_1 and C_2 were obtained. In subsequent weekly time periods, individual advertisements were screened in total: t_1 – 365555,: t_2 – 229538,: t_3 – 445604,: t_4 – 153498 times. On the other hand, the total numbers of clicks amounted to: t_1 – 283,: t_2 – 498,: t_3 – 494,: t_4 – 546. As it had been noted, the values of variants for the criterion C_3 were determined on the basis of surveys before launching the experimental campaign. Next, the values of surveys were averaged and, in consequence, in every time period, the values of a given variant for the criterion $C\neg_3$ were the same. Thanks to the values of decision variants for individual criterion, a performance table for every time period was constructed. The values of decision variants for individual criteria in subsequent time periods are shown in Appendix 2.

4.2 Preference Modelling

In a preference modelling stage, for the PROMETHEE method, the following needed to be done: selecting criteria preference functions and defining values of thresholds for the criteria as well as defining the importances of the criteria and directions of preferences. The selection of preference functions and values of thresholds greatly affect the order of the variants in a ranking [52]. Moreover, the type of preference functions applied depends of the type of criteria. For quantitative criteria it is recommended that functions using the following are applied: a V-shape criterion, a V-shape with indifference threshold criterion or a Gaussian criterion [53]. The developed decision model applied the V-shape criterion. This function uses preference threshold p, whose value should fall within reliable min and max values taken by a given criterion [54]. In the developed decision model, the value of p threshold was assumed to be twice the standard deviation.

As far as the importances of criteria and preference directions are concerned, one needs to point out that in the research two approaches to the decision problem were considered: the advertiser's perspective and the website operator's (publisher's) perspective. From the advertiser's point of view, the most important criteria are the click-through rate (C_1) and advertiser's cost (C_2), since they directly determine the effectiveness of the ad. The intensity of advertising content (C_3) is a less significant criterion; however, high intensity can negatively influence the user's experience, therefore, his or her perception of the advertised product. From the publisher's point of view, their profit (C_2) as well as the intensity of advertising content (C_3) on such a level so as not to negatively influence the user's experience are the most important factors for them. Obviously, from the advertiser's point of view, the criterion C_2 should approach the minimum, whereas from the website operator's point of view, an expected preference direction for the criterion C_2 is the maximum. Moreover, for the two perspectives, the common preference direction for the criterion C_1 is the maximum and for C_3 is the minimum. The complete preference model taking into consideration subsequent time periods is presented in Table 1.

Table 1. The preference model considering time periods

Perspective	Features of the preference model	Criterion		
		C_1 - CTR	C_2 - AC/PP	C_3 - IoA
Advertiser	Preference direction	Max	Min	Min
	Importance of criterion	40%	40%	20%
Publisher	Preference direction	Max	Max	Min
	Importance of criterion	20%	40%	40%
Both	Preference function	V-shape	V-shape	V-shape
	Preference threshold (p) p_1	0.16%	2.84 $	0.5385
	Preference threshold (p) p_2	0.19%	3.32 $	0.5385
	Preference threshold (p) p_3	0.09%	3.21 $	0.5385
	Preference threshold (p) p_4	0.46%	3.65 $	0.5385

4.3 Performance Aggregation

The performance aggregation of variants was carried out separately for every time period. As a result, for every perspective (of the advertiser and the publisher), four rankings of variants along with the values of a net outranking flow attributed to the variants were obtained. The top 10 positions in these rankings are presented in Table 2 (the advertiser's perspective) and in Table 3 (the publisher's perspective). Complete rankings are presented in Appendix 3.

Table 2. Rankings of variants for subsequent time periods – the advertiser's perspective

Rank	p_1		p_2		p_3		p_4	
	Variant	Φ_{net}	Variant	Φ_{net}	Variant	Φ_{net}	Variant	Φ_{net}
1	A1.1	0.4232	A2.1	0.6042	A3.1	0.6352	A3.3	0.4595
2	A1.3	0.4182	A3.1	0.548	A1.3	0.5472	A3.1	0.4497
3	A2.2	0.3798	A3.2	0.4844	A3.2	0.4127	A5.3	0.4335
4	A2.1	0.3354	A2.3	0.4179	A2.3	0.4124	A3.2	0.425
5	A1.2	0.2458	A1.3	0.4166	A3.3	0.3937	A4.3	0.3997
6	A2.3	0.2315	A3.3	0.3992	A2.1	0.3552	A5.2	0.3791
7	A3.3	0.2202	A1.2	0.3735	A3.4	0.2807	A5.1	0.3641
8	A5.2	0.2178	A5.1	0.2431	A2.2	0.2768	A1.3	0.2811
9	A5.3	0.1704	A4.1	0.1922	A1.2	0.2715	A1.1	0.2225
10	A5.1	0.1476	A3.7	0.1552	A4.2	0.2431	A4.1	0.2213

Table 3. Rankings of variants for subsequent time periods – the publisher's perspective

Rank	p_1		p_2		p_3		p_4	
	Variant	Φ_{net}	Variant	Φ_{net}	Variant	Φ_{net}	Variant	Φ_{net}
1	A1.9	0.5807	A3.8	0.506	A1.9	0.375	A4.3	0.3619
2	A2.9	0.4701	A3.7	0.3881	A5.7	0.3536	A5.10	0.3515
3	A2.7	0.3804	A3.3	0.3647	A2.3	0.3438	A1.10	0.331
4	A1.3	0.3682	A4.7	0.3454	A3.9	0.2816	A4.10	0.2846
5	A1.10	0.3633	A5.7	0.2755	A1.3	0.246	A5.4	0.2713
6	A1.8	0.3439	A4.8	0.2622	A1.8	0.2429	A3.5	0.2443
7	A5.10	0.2769	A1.10	0.2535	A3.1	0.241	A4.7	0.235
8	A1.7	0.266	A4.9	0.2532	A4.7	0.2229	A3.3	0.2166
9	A2.10	0.2434	A3.10	0.2474	A3.8	0.2151	A1.3	0.2154
10	A1.1	0.2396	A5.9	0.2359	A2.7	0.2118	A2.9	0.2014

Another step was the aggregation of rankings from subsequent time periods into global rankings, separately for the advertiser's and publisher's perspectives. For every perspective two aggregation strategies were used, that is TPEA and TPDA. In the TPEA strategy, the values of weights $r_{k\neg}$ of subsequent time periods (see the formula (5)) amounted to $r_1 = r_2 = r_3 = r_4 = 0.25$. In the TPDA strategy, the weights $r_1 = 0.1$, $r_2 = 0.2$, $r_3 = 0.3$, $r_4 = 0.4$ were assumed. The global rankings, for both perspectives, obtained according to the presented strategies, are shown in Table 4 (10 top variants for each of them) as well as in Appendix 4 (complete rankings).

Table 4. Global rankings of variants from the advertiser's and publisher's perspectives

Rank	Advertiser's perspective				Publisher's perspective			
	TPEA		TPDA		TPEA		TPDA	
	Variant	Φ^{DM1}_{global}	Variant	Φ^{DM1}_{global}	Variant	Φ^{DM2}_{global}	Variant	Φ^{DM2}_{global}
1	A3.1	0.4448	A3.1	0.4947	A1.3	0.2612	A1.10	0.2432
2	A1.3	0.4158	A3.3	0.4038	A1.10	0.2567	A5.10	0.2406
3	A3.3	0.3682	A1.3	0.4017	A1.9	0.2469	A1.3	0.2398
4	A2.1	0.3595	A3.2	0.3972	A5.10	0.2365	A4.7	0.2163
5	A3.2	0.3469	A2.1	0.3182	A3.3	0.2246	A3.3	0.2117
6	A2.3	0.2616	A5.1	0.2778	A4.9	0.1931	A3.7	0.2029
7	A1.2	0.2604	A5.3	0.2515	A2.3	0.1850	A3.8	0.1989
8	A1.1	0.2527	A1.2	0.2410	A3.7	0.1822	A4.9	0.1927
9	A5.1	0.2460	A5.2	0.2382	A1.7	0.1724	A3.9	0.1726
10	A5.2	0.2159	A1.1	0.2281	A4.7	0.1668	A2.3	0.1711

The last step in the aggregation stage was to link the global rankings, obtained for the advertiser's and publisher's perspectives into a final PROMETHEE GDSS ranking considering both points of view. It was carried out separately for the results obtained earlier with the use of the TPEA and TPDA strategies. The final rankings for the two strategies are presented in Table 5 (10 top variants) and Appendix 1.

Table 5. PROMETHEE GDSS rankings joining the advertiser's and publisher's perspectives

Rank	TPEA		TPDA	
	Variant	Φ_{global}	Variant	Φ_{global}
1	A1.3	0.1727	A3.1	0.1691
2	A3.1	0.1515	A1.3	0.1636
3	A3.3	0.1512	A3.3	0.157
4	A2.1	0.1166	A3.2	0.1117
5	A2.3	0.1139	A2.3	0.1009
6	A3.2	0.0934	A2.1	0.0938
7	A5.1	0.0802	A5.1	0.0912
8	A5.3	0.0773	A5.3	0.0894
9	A1.1	0.0766	A4.3	0.0825
10	A3.7	0.0683	A3.7	0.0774

4.4 Exploitation of Results and Recommendation

The exploitation of the solution started with the analysis of correlation between individual rankings in the research. Table 6 contains Pearson correlation coefficients calculated on the basis of the values of net preference flows (Φ_{net}) variants in: (1) subsequent time periods (p_1–p_4), (2) global rankings (Φ_{global}) obtained according to TPEA and TPDA strategies. Significant $p < 0.05$ correlation coefficients were marked in Table 6. The values, presented in Table 6, of the correlation coefficients of 0.6–0.9 confirm that there are strong relationships between individual partial rankings (p_1–p_4) determined on the basis of the advertiser's preferences. Relationships between partial rankings (p_1–p_4) constructed on the basis of publisher's preferences are no that strong, what is confirmed by lower values of correlation coefficients. This phenomenon is influenced by differences in the advertiser's and publisher's preference models, i.e. different importances of criteria and different preference direction for a criterion C_2. Similar observations refer to correlations between partial rankings (p_1–p_4) and global rankings (TPEA, TPDA). These correlations are higher for rankings determined on the basis of the advertiser's preferences than for the rankings based on the publisher's preferences. Moreover, it is surprising that in the case of the TPDA aggregation strategy, the partial ranking p_4 is not always most significantly related to the global ranking, even though the weight of p_4 had the highest value ($r_1 = 0.1,\ r_2 = 0.2,\ r_3 = 0.3,\ r_4 = 0.4$). It can be confirmed by, for example, correlations between the rankings p_1-p_4 and TPDA in the advertiser's perspective. The correlation value of the ranking p_4 with the global ranking TPDA is lower than the correlation of the ranking p_2 with TPDA, and similarly, the correlation value of the ranking p_4 with the global ranking TPDA is lower than the correlation of the ranking p_3 with TPDA. A similar observation refers to the publisher's p_3 ranking. However, the use of the TPDA strategy, in comparison with TPEA, increases the strength of the relationship of the global ranking with the partial ranking p_4, at the same time, decreasing the relationship of the global ranking with the rankings p_1–p_3. As far as the correlations between the global rankings TPEA and TPDA are concerned, their values are at the level of 0.97–0.99, what indicates a strong relationship between the rankings.

Table 6. Correlations between partial and global rankings on the basis of the values of net preference flows

	Advertiser's perspective					Publisher's perspective				
	p_2	p_3	p_4	TPEA	TPDA	p_2	p_3	p_4	TPEA	TPDA
p_1	*0.653*	*0.748*	*0.598*	*0.828*	*0.769*	0.151	*0.439*	*0.394*	*0.661*	*0.526*
p_2		*0.889*	*0.661*	*0.917*	*0.898*		*0.697*	*0.614*	*0.794*	*0.832*
p_3			*0.679*	*0.945*	*0.929*			*0.438*	*0.832*	*0.816*
p_4				*0.836*	*0.89*				*0.789*	*0.857*
TPEA					*0.991*					*0.981*

Also, as part of exploiting the solution, the correlations of the PROMETHEE GDSS rankings with the global rankings (TPEA, TPDA) as well as the partial rankings (p_1–p_4) were examined. The results of the study are presented in Table 7.

Table 7. Correlations of the partial and global rankings with the GDSS rankings on the basis of net preference flows

	Advertiser's perspective						Publisher's perspective					
	p_1	p_2	p_3	p_4	TPEA	TPDA	p_1	p_2	p_3	p_4	TPEA	TPDA
GDSS TPEA	0.67	0.78	0.74	0.68	0.81	–	0.48	0.64	0.61	0.54	0.74	–
GDSS TPDA	0.59	0.76	0.72	0.74	–	0.81	0.36	0.68	0.59	0.60	–	0.73

The analysis of Table 7 indicates that the advertiser's perspective has a greater influence, compared to the publisher's perspective, on the compromise GDSS solution which merges the two points of view. It is confirmed by empirical observations based on the analysis of Tables 4 and 5 presented in Clause 4.3 of this article. When analyzing Table 4, one can notice that, from the advertiser's perspective, advertisement variants characterized by a lower unit cost of every interaction/click are usually better (Ax.1–Ax.5). On the other hand, variants with a higher unit cost of every interaction are more optimal for the publisher (Ax.6–Ax.10). In Table 5, which shows rankings taking into consideration both perspectives, there are definitely more variants with a lower unit cost of every interaction.

Another step within exploiting the solution was to conduct the GAIA analysis of TPEA and TPDA rankings obtained on the basis of the advertiser's and publisher's preferences. The planes of the GAIA rankings obtained on the basis of the advertiser's perspective are presented in Fig. 2. Figure 3 depicts the GAIA plane for the TPEA and TPDA rankings generated on the basis of the publisher's preferences. Comparing the GAIA planes for the TPEA and TPDA strategies presented in Figs. 2 and 3 allows us to find out that the GAIA planes for TPEA differ from their corresponding GAIA planes for TPDA only with the location of a vector Π showing the direction of searching for a compromise solution.

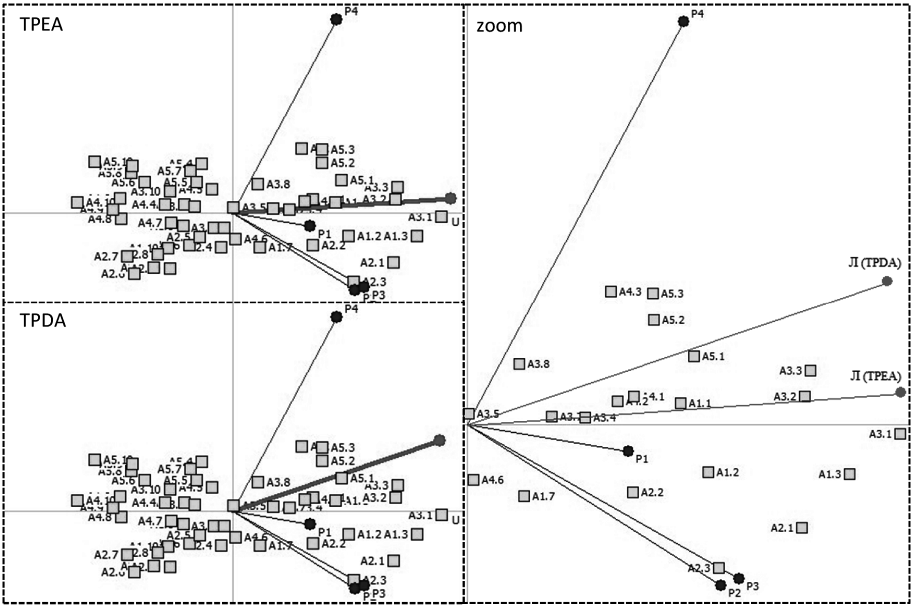


Fig. 2. GAIA planes of the TPEA and TPDA rankings obtained on the basis of the advertiser's preferences

When analysing Fig. 2, it should be noticed that the vector Л, both for the TPEA and TPDA strategies, indicates variants A3.1, A1.3, A3.3, A3.2, A2.1 as the most preferred ones. Basically, it is in accordance with the rankings for the advertiser's perspective, presented in Table 5, although there are shifts in the sequence, e.g. variants A3.2 and A2.1. There are more serious errors on the GAIA plane in further positions in the rankings. Vectors P_1–P_4, which symbolize proper partial rankings, do not allow precise reading of the sequence of top variants in the rankings. Furthermore, when analysing the directions of the vectors, one can deduce that the rankings p_2 as well as $p_{\neg 3}$ and partially p_1 are generally in conformity with each other. This observation is confirmed by the correlation analysis carried out earlier (see Tables 7 and 8). However, direction of a vector P_4, indicating that the ranking p_4 is in more conformity with the ranking p_1 than the rankings p_2 and p_3, does not coincide with the conclusions resulting from the correlation analysis. On the other hand, the lengths of vectors P_1–P_4 point out that the ranking p_4 influences most strongly and p_1 influences most weakly the global rankings TPEA and TPDA. Similarly, the observation is contradictory to the conclusions drawn from the correlation analysis.

The analysis of Fig. 3 points out that in the TPEA strategy the most preferred variants are A1.3, A1.10, A1.9, 5.10 and A3.3, whereas in the TPDA strategy there are A1.3, A1.10, A3.3, A5.10, A3.8. On the basis of Table 5 in can be noticed that the sequence resulting from the analysis of the vector *Л* for the TPDA strategy does not correspond precisely to the TPDA ranking, presented in Table 5, from the publisher's perspective. Moreover, the vectors P_1 and P_2 make it possible to correctly read the sequence of variants in proper partial rankings. According to the GAIA plane, in the

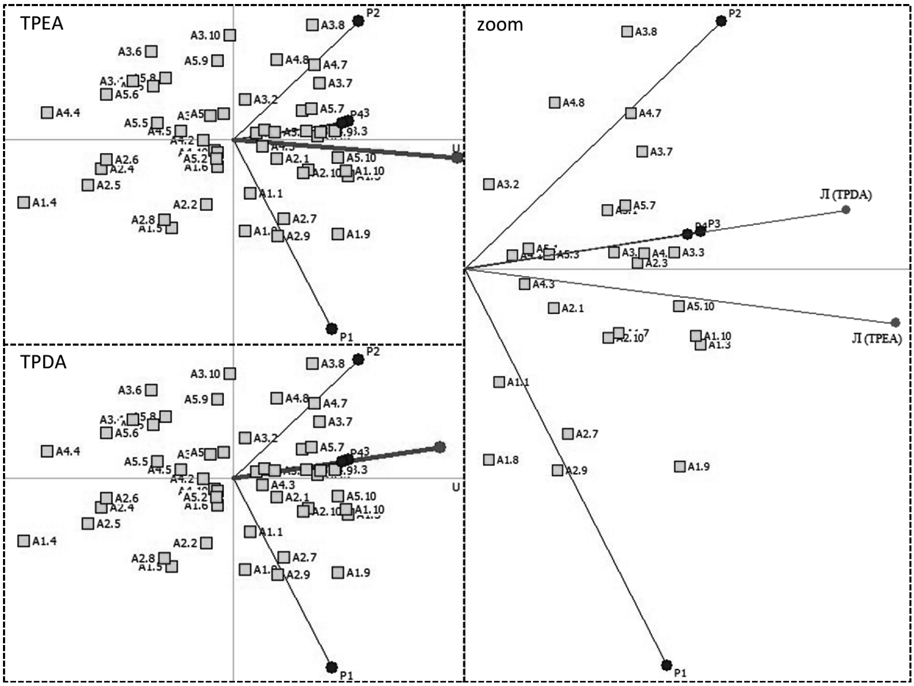


Fig. 3. GAIA planes of the rankings TPEA and TPDA obtained on the basis of the publisher's preference

ranking p_1, the top positions are taken by variants A1.9, A2.9 and A2.7, and in the ranking p_2 - A3.8, A4.7, A3.7. In the case of the vectors P_3 and P_4, the correct reading of the sequence of variants in the rankings corresponding to the vectors is no longer possible. On the basis of the directions of the vectors P_1–P_4, it can be concluded that the ranking p_1 is basically unrelated to the ranking p_2, what is in conformity with the results of the correlation analysis. Moreover, the directions of the vectors P_2–P_4 indicate that their corresponding rankings are strongly related to each other. There is such a relationship indeed, however, the GAIA plane wrongly suggests that with the ranking p_4, the ranking p_3 is strongly related, not p_2. Nevertheless, the correlation analysis pointed out that the ranking p_4 is strongly related to the ranking p_2. What is more, on the GAIA plane, it was not reflected that there is the strongest relationship between the rankings p_2 and p_3 among the publisher's partial rankings. As for the influence of the partial rankings on the global rankings TPEA and TPDA, the lengths of the vectors P_1–P_4 suggest that the rankings p_1 and p_2 have the greatest influence and the influence of the rankings p_3 and p_4 is less significant. It is obviously contradictory to the conclusions drawn from the correlation analysis carried out earlier. It ought to be noted that the inconsistencies between the GAIA analysis and the correlation analysis, indicated when analysing Figs. 2 and 3, result from the errors on the GAIA plane. The reason for the errors is the loss of some information as a result of the multidimensional projection of the decision problem (in this case, five dimensions reflecting the partial rankings and the global ranking) on the two-dimensional GAIA plane.

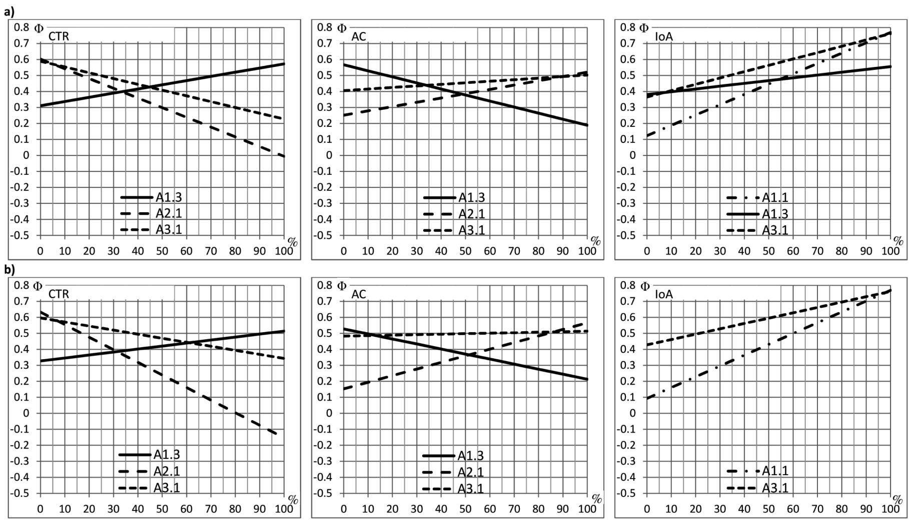


Fig. 4. Charts depicting the robustness of the solutions to the changes of criterial importances: (a) TPEA – advertiser's perspective, (b) TPDA - advertiser's perspective

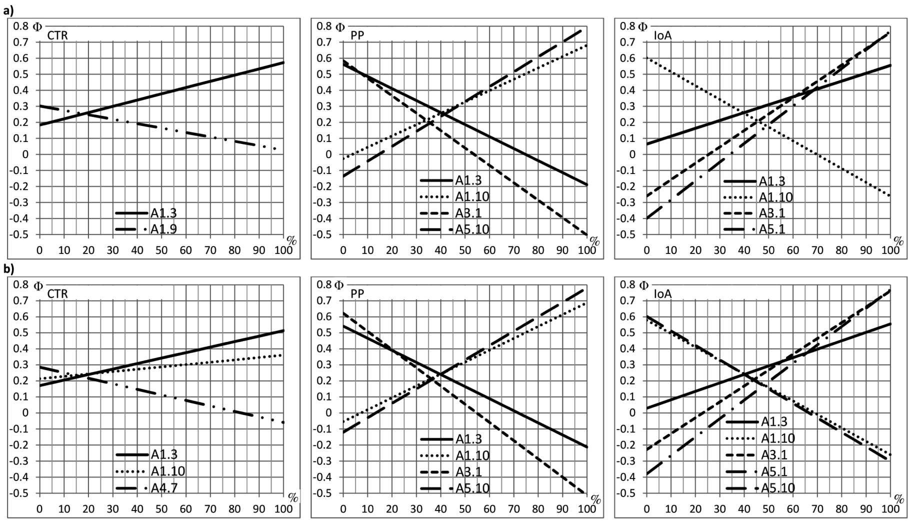


Fig. 5. Charts depicting the robustness of the solutions to the changes of criterial importances: (a) TPEA – publisher's perspective, (b) TPDA – publisher's perspective

The last step taken when exploiting the solution was the analysis of the robustness of the global rankings TPEA and TPDA, which were obtained respectively for the advertiser's and publisher's perspectives, to the changes of importances of criteria in decision subproblems p_1–p_4. Figures 4 and 5 present the results of the robustness analysis for the best variants.

Comparing the charts, depicted in Fig. 4, generated for the advertiser's perspective, one can notice their similarities. However, it should be noted that in the case of the TPDA strategy, the solution to the decision problem (variant A3.1) is more robust to the changes of criterial importances than the solution obtained with the use of the TPEA strategy. It is caused by the fact that the variant A3.1 takes the top position in the latest partial rankings (p_1–p_4), which have higher weights in the TPDA strategy. Therefore, the application of TPDA additionally enhances the variant A3.1.

The comparison of the global rankings TPEA and TPDA for the publisher's perspective (see Fig. 5) indicates that the robustness of both rankings to the changes of criterial importances is very similar, even though in this case, the TPDA ranking is slightly less robust than TPEA. Moreover, the publisher's rankings are significantly less robust than the advertiser's rankings.

Additionally, the analysis of the solution's robustness to weight changes of individual time periods was conducted. The results are presented in Fig. 6. This analysis was carried out with the assumption that the input weights of time periods are equal (TPEA), what means that when, for example, the weight which is being analyzed equals 0, then every other equals 0.33. The analysis of Fig. 6 indicates that, similarly to

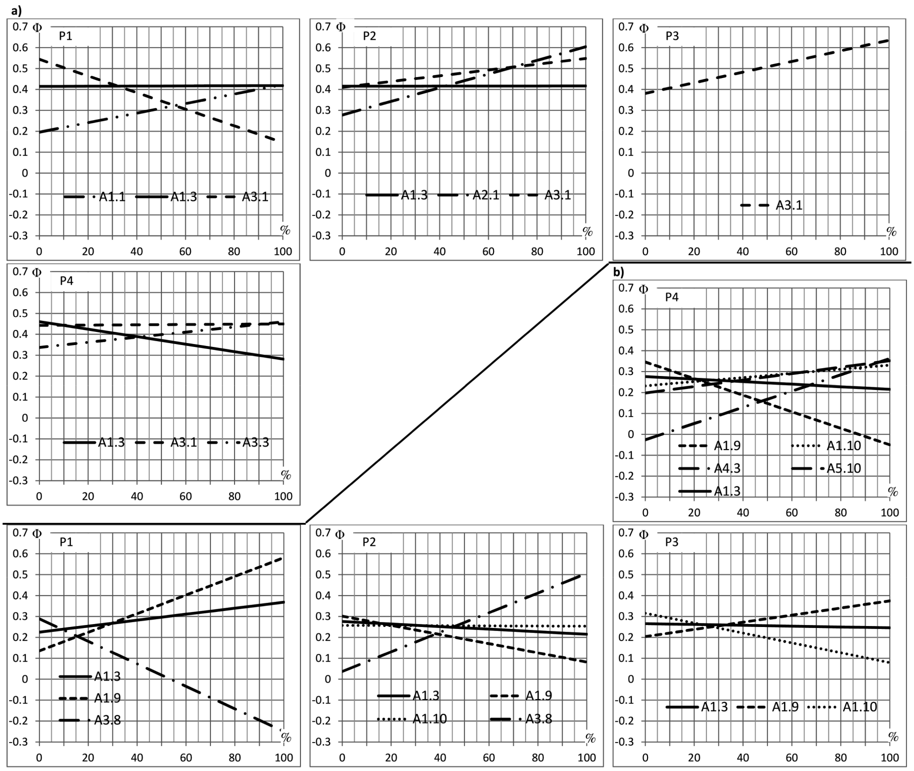


Fig. 6. Charts of the solutions' robustness to changes of weights of time periods: (a) advertiser's perspective, (b) publisher's perspective

the analysis concerning changes of the criterial importances, the ranking obtained for the advertiser's perspective is more stable and robust. The position of the variant A3.1 in this ranking is so strong that it takes the first position in the ranking irrespective of the weights of a time period p_3. As for the publisher's ranking, it is robust to changes of weights of time periods only to a very limited extent. Moreover, even a minimal change of weights of individual time periods may cause changes in the first position of the publisher's ranking. The above-mentioned conclusions also fully refer to the results, which are almost the same, of an analogous robustness analysis for the TPDA strategy. For that reason, the results will not be presented here separately.

Based on the presented rankings and analyses, the advertisement variant A3.1 should be indicated as a recommended solution to the decision problem. The variant takes the first position in the advertiser's global rankings and the top positions in the partial rankings for subsequent time periods. Additionally, it is also a good solution for the publisher.

5 Conclusions

The presented results confirm the ability of application of the proposed procedure for obtaining decision solutions given changeability of measurement data and the presence of multiple criteria. The results indicate significant dynamics of obtained solutions in the field of online marketing campaign management and its effectiveness. Therefore, it is necessary to continuously evaluate the effectiveness of advertising in relation to its other aspects, such as invasiveness and profits for the owner's website or costs incurred by the advertiser.

In the methodological aspect, the use of the dynamic multi-criteria decision analysis approach based on the PROMETHEE method has revealed broader possibilities of the decision solution analysis as well as the interpretation of the obtained results. Obviously, the proposed approach makes it possible to carry out analyses referring to a greater extent to criteria or source data and not only to time periods and their aggregation strategies. However, the analyses presented in the articles made it possible to show strengths and weaknesses of individual rankings and obtained decision solutions. Moreover, the view of different perspectives in the GDSS rankings proved that, although there seem to be contradictory points of view, it is feasible to indicate solutions attractive both to the advertiser and the publisher. These are the decision variants A1.3, A3.3 and A3.1, in particular.

While summarizing the research carried out, it should be noted that the contribution included the following highlights:

- the adaptation of the PROMETHEE method and the GAIA analysis for the dynamic MCDA in the area of online marketing campaign management by means of considering changeability in time of performance of decision variants and considering different perspectives of an attitude to a marketing campaign,
- solving a multilevel decision problem including finding a compromise between effectiveness, invasiveness for the user and the cost of an online advertisement as well as a compromise between the advertiser and the website owner,

- the wide analysis of the obtained solution with the use of research into the correlation between time periods, the GAIA analysis in the dimension of time as well as research into the stability of decision solutions in the dimension of weights of criteria and time periods.

The conducted research determines further directions of development of the proposed dynamic approach to MCDA, that is, applying different preferences and criteria in subsequent time periods, the possibility of extending a set of variants, taking into consideration forecasts concerning future time periods in a solution. An interesting issue is also adding the user's perspective and considering it similarly to, for example, [55–57], directly to the proposed dynamic approach. Although it should be noted that the clickability criterion (C_1) and in the invasiveness criterion (C_3) indirectly also comprise this perspective in the considered dynamic decision problem.

Appendix 1

See Table 8.

Table 8. PROMETHEE GDSS rankings

Decision variant	TPEA		TPDA		Decision variant	TPEA		TPDA	
	Φ_{global}	Rank	Φ_{global}	Rank		Φ_{global}	Rank	Φ_{global}	Rank
A1.1	0.0766	9	0.0633	12	A3.6	−0.0596	39	−0.0498	37
A1.2	0.0561	14	0.0446	16	A3.7	0.0683	10	0.0774	10
A1.3	0.1727	1	0.1636	2	A3.8	0.0527	15	0.0731	11
A1.4	−0.1371	49	−0.1562	50	A3.9	0.0162	21	0.0275	18
A1.5	−0.0389	34	−0.0572	39	A3.10	−0.0406	36	−0.0197	28
A1.6	−0.0328	33	−0.0481	36	A4.1	0.0594	12	0.0628	13
A1.7	0.0588	13	0.0493	15	A4.2	0.0236	19	0.0309	17
A1.8	−0.0165	28	−0.0371	34	A4.3	0.0597	11	0.0825	9
A1.9	0.0041	23	−0.0237	30	A4.4	−0.1450	50	−0.1458	49
A1.10	0.0279	18	0.0153	20	A4.5	−0.0415	37	−0.0325	31
A2.1	0.1166	4	0.0938	6	A4.6	−0.0509	38	−0.0387	35
A2.2	0.0371	17	0.0125	22	A4.7	0.0095	22	0.0147	21
A2.3	0.1139	5	0.1009	5	A4.8	−0.0402	35	−0.0333	32
A2.4	−0.0799	43	−0.0839	43	A4.9	−0.0221	30	−0.0206	29
A2.5	−0.0989	46	−0.1231	47	A4.10	−0.0974	45	−0.0801	42
A2.6	−0.1282	48	−0.1353	48	A5.1	0.0802	7	0.0912	7
A2.7	−0.0275	31	−0.0509	38	A5.2	0.0467	16	0.0498	14
A2.8	−0.0784	42	−0.0897	45	A5.3	0.0773	8	0.0894	8
A2.9	−0.0160	26	−0.0369	33	A5.4	−0.0193	29	0.0031	24

(*continued*)

Table 8. (*continued*)

Decision variant	TPEA		TPDA		Decision variant	TPEA		TPDA	
	Φ_{global}	Rank	Φ_{global}	Rank		Φ_{global}	Rank	Φ_{global}	Rank
A2.10	0.0025	24	−0.0053	25	A5.5	−0.0624	40	−0.0572	40
A3.1	0.1515	2	0.1691	1	A5.6	−0.1222	47	−0.1215	46
A3.2	0.0934	6	0.1117	4	A5.7	0.0217	20	0.0245	19
A3.3	0.1512	3	0.1570	3	A5.8	−0.0947	44	−0.0866	44
A3.4	−0.0318	32	−0.0163	27	A5.9	−0.0653	41	−0.0615	41
A3.5	−0.0147	25	0.0117	23	A5.10	−0.0160	27	−0.0088	26

Appendix 2

See Table 9.

Table 9. Values of variants for individual criteria

Decision variant	CTR [%]				AC/PP [$]				IoA
	p_1	p_2	p_3	p_4	p_1	p_2	p_3	p_4	p_1–p_4
A1.1	0.15	0.08	0.08	0.27	0.50	0.25	0.45	0.80	0.0026
A1.2	0.10	0.26	0.11	0.24	0.40	0.50	0.70	0.60	0.2584
A1.3	0.24	0.33	0.20	0.56	1.50	1.65	1.50	2.25	0.1843
A1.4	0.18	0.06	0.04	0.28	1.40	0.40	0.60	1.40	0.8474
A1.5	0.27	0.13	0.14	0.34	2.75	1.25	2.75	2.50	0.7028
A1.6	0.21	0.26	0.21	0.32	2.70	3.30	4.50	3.00	0.8386
A1.7	0.17	0.28	0.15	0.44	2.45	2.80	2.80	3.85	0.4392
A1.8	0.21	0.17	0.14	0.22	3.60	1.60	4.80	2.40	0.5785
A1.9	0.33	0.14	0.12	0.23	6.30	4.05	6.75	3.15	0.5
A1.10	0.22	0.38	0.12	0.60	4.50	4.50	4.00	7.50	0.6481
A2.1	0.12	0.34	0.11	0.11	0.60	0.40	0.50	0.35	0.0053
A2.2	0.22	0.15	0.12	0.17	1.20	0.60	0.80	1.30	0.3413
A2.3	0.11	0.33	0.19	0.16	0.90	1.65	2.40	1.95	0.1799
A2.4	0.13	0.17	0.13	0.38	1.40	0.80	1.60	3.00	0.7989
A2.5	0.15	0.21	0.06	0.15	2.00	1.50	0.75	1.50	0.7434
A2.6	0.11	0.14	0.13	0.23	1.80	1.80	2.70	4.20	0.9259
A2.7	0.19	0.14	0.12	0.12	3.85	2.80	4.55	4.20	0.5317
A2.8	0.13	0.08	0.08	0.25	2.80	0.80	2.00	4.00	0.6138
A2.9	0.16	0.11	0.06	0.18	4.05	2.25	1.80	4.95	0.3889
A2.10	0.11	0.22	0.13	0.28	3.00	3.00	3.50	5.00	0.4709
A3.1	0.01	0.33	0.19	0.53	0.05	0.80	0.80	0.70	0.0106
A3.2	0.01	0.36	0.15	0.62	0.10	1.20	0.90	1.00	0.291

(*continued*)

Table 9. (*continued*)

Decision variant	CTR [%]				AC/PP [$]				IoA
	p_1	p_2	p_3	p_4	p_1	p_2	p_3	p_4	p_1–p_4
A3.3	0.12	0.38	0.14	0.70	1.20	2.55	1.05	1.65	0.1693
A3.4	0.03	0.32	0.18	0.54	0.40	2.60	1.60	1.40	0.7566
A3.5	0.07	0.22	0.16	0.90	1.25	2.50	2.50	5.00	0.7063
A3.6	0.01	0.36	0.26	0.35	0.30	3.30	4.50	1.50	0.8836
A3.7	0.04	0.36	0.13	0.54	1.05	3.85	2.80	3.15	0.3519
A3.8	0.01	0.47	0.16	0.70	0.40	6.00	3.60	3.20	0.5053
A3.9	0.09	0.27	0.18	0.58	2.70	4.05	4.50	4.05	0.5952
A3.10	0.01	0.34	0.10	0.77	0.50	5.50	3.00	5.00	0.7302
A4.1	0.01	0.17	0.08	0.28	0.10	1.10	1.10	0.90	0.0053
A4.2	0.03	0.16	0.12	0.34	0.40	0.90	1.30	0.80	0.2937
A4.3	0.03	0.13	0.05	0.77	0.60	1.05	0.75	2.70	0.1455
A4.4	0.02	0.17	0.07	0.23	0.60	2.20	1.40	1.20	0.8201
A4.5	0.05	0.24	0.03	0.62	1.50	2.50	0.75	3.50	0.5608
A4.6	0.02	0.34	0.16	0.49	0.90	2.40	2.70	3.00	0.7989
A4.7	0.02	0.29	0.10	0.41	0.70	4.55	4.20	4.55	0.418
A4.8	0.02	0.23	0.11	0.38	0.80	5.60	4.40	4.80	0.5556
A4.9	0.06	0.34	0.09	0.49	3.60	4.50	5.85	4.95	0.6032
A4.10	0.05	0.17	0.07	0.69	3.00	3.00	4.00	8.00	0.7989
A5.1	0.01	0.19	0.09	0.44	0.05	1.00	0.95	0.80	0.0026
A5.2	0.11	0.19	0.08	0.56	0.70	1.10	0.90	0.90	0.3042
A5.3	0.07	0.21	0.08	0.65	0.60	2.10	1.20	1.50	0.1852
A5.4	0.10	0.21	0.11	1.22	1.20	3.20	3.00	4.80	0.6905
A5.5	0.07	0.18	0.12	0.42	1.00	2.50	3.25	1.75	0.672
A5.6	0.05	0.22	0.08	0.37	0.90	3.60	2.70	1.80	0.8571
A5.7	0.07	0.19	0.13	0.36	1.40	4.55	4.55	2.10	0.373
A5.8	0.02	0.17	0.07	0.42	0.40	4.00	3.20	2.80	0.672
A5.9	0.02	0.20	0.08	0.34	0.45	5.85	3.60	2.25	0.5714
A5.10	0.15	0.23	0.09	0.83	4.50	6.50	5.00	6.50	0.672

Appendix 3

See Tables 10 and 11.

Table 10. Rankings of variants for subsequent time periods – advertiser's perspective

Decision variant	p_1		p_2		p_3		p_4	
	Φ_{net}	Rank	Φ_{net}	Rank	Φ_{net}	Rank	Φ_{net}	Rank
A1.1	0.4232	1	0.1272	13	0.2379	11	0.2225	9
A1.2	0.2458	5	0.3735	7	0.2715	9	0.1507	14

(*continued*)

Table 10. (*continued*)

Decision variant	p_1		p_2		p_3		p_4	
	Φ_{net}	Rank	Φ_{net}	Rank	Φ_{net}	Rank	Φ_{net}	Rank
A1.3	0.4182	2	0.4166	5	0.5472	2	0.2811	8
A1.4	0.1104	13	−0.182	38	−0.1859	38	−0.093	33
A1.5	0.0975	14	−0.1183	31	0.0067	24	−0.116	35
A1.6	−0.0157	25	−0.1371	34	−0.0064	25	−0.2268	42
A1.7	0.0794	16	0.0987	17	0.136	14	−0.0818	31
A1.8	−0.0405	27	−0.0439	27	−0.1786	37	−0.1593	36
A1.9	−0.0098	21	−0.3561	48	−0.3546	47	−0.2029	40
A1.10	−0.1081	37	0.0076	22	−0.2152	40	−0.2735	44
A2.1	0.3354	4	0.6042	1	0.3552	6	0.1432	15
A2.2	0.3798	3	0.1103	16	0.2768	8	0.0073	22
A2.3	0.2315	6	0.4179	4	0.4124	4	−0.0153	23
A2.4	0.0025	20	−0.0312	25	0.0747	19	−0.165	37
A2.5	−0.0137	24	−0.0054	24	−0.1029	30	−0.1708	38
A2.6	−0.1371	40	−0.2276	42	−0.094	29	−0.4509	50
A2.7	−0.0789	31	−0.219	41	−0.2274	41	−0.4069	49
A2.8	−0.1334	39	−0.1184	32	−0.1234	31	−0.3193	46
A2.9	−0.105	36	−0.1473	35	−0.0917	28	−0.3838	48
A2.10	−0.1591	43	−0.0615	28	−0.0476	27	−0.3426	47
A3.1	0.1461	11	0.548	2	0.6352	1	0.4497	2
A3.2	0.0655	18	0.4844	3	0.4127	3	0.425	4
A3.3	0.2202	7	0.3992	6	0.3937	5	0.4595	1
A3.4	−0.094	33	0.092	19	0.2807	7	0.1544	13
A3.5	−0.0963	34	−0.0847	30	0.118	16	0.031	21
A3.6	−0.1607	45	0.0414	20	0.0394	23	−0.056	27
A3.7	−0.0109	22	0.1552	10	0.0844	18	0.1131	17
A3.8	−0.0509	28	0.0312	21	0.051	22	0.1784	12
A3.9	−0.2119	47	−0.1273	33	−0.0085	26	−0.0392	24
A3.10	−0.1407	41	−0.161	37	−0.2055	39	−0.04	25
A4.1	0.1412	12	0.1922	9	0.1744	13	0.2213	10
A4.2	0.0702	17	0.1168	15	0.2431	10	0.2024	11
A4.3	0.0907	15	0.0949	18	0.068	20	0.3997	5
A4.4	−0.1592	44	−0.1885	40	−0.1567	36	−0.1052	34
A4.5	−0.1275	38	0.0069	23	−0.1353	34	0.0651	18
A4.6	−0.1908	46	0.139	12	0.065	21	−0.0704	30
A4.7	−0.034	26	−0.0717	29	−0.2385	42	−0.1734	39
A4.8	−0.0983	35	−0.3285	46	−0.2663	43	−0.2738	45
A4.9	−0.3953	49	−0.0389	26	−0.4756	50	−0.209	41
A4.10	−0.4148	50	−0.2764	45	−0.4683	49	−0.266	43
A5.1	0.1476	10	0.2431	8	0.2291	12	0.3641	7

(*continued*)

Table 10. (*continued*)

Decision variant	p_1		p_2		p_3		p_4	
	Φ_{net}	Rank	Φ_{net}	Rank	Φ_{net}	Rank	Φ_{net}	Rank
A5.2	0.2178	8	0.1523	11	0.1143	17	0.3791	6
A5.3	0.1704	9	0.1225	14	0.1217	15	0.4335	3
A5.4	−0.0113	23	−0.1844	39	−0.15	35	0.1228	16
A5.5	−0.0516	29	−0.1565	36	−0.1314	33	0.0408	20
A5.6	−0.1422	42	−0.2628	44	−0.2855	44	−0.0637	28
A5.7	0.0034	19	−0.2623	43	−0.1239	32	0.0614	19
A5.8	−0.0883	32	−0.3559	47	−0.3301	46	−0.0696	29
A5.9	−0.0602	30	−0.4132	49	−0.3068	45	−0.0451	26
A5.10	−0.2566	48	−0.4153	50	−0.4389	48	−0.0869	32

Table 11. Rankings of variants for subsequent time periods – publisher's perspective

Decision variant	p_1		p_2		p_3		p_4	
	Φ_{net}	Rank	Φ_{net}	Rank	Φ_{net}	Rank	Φ_{net}	Rank
A1.1	0.2396	10	−0.0775	31	−0.005	27	0.0332	25
A1.2	0.0344	24	−0.0183	25	−0.0508	30	−0.1268	34
A1.3	0.3682	4	0.2151	13	0.246	5	0.2154	9
A1.4	−0.1442	35	−0.614	50	−0.5979	50	−0.4438	49
A1.5	0.1978	14	−0.3915	46	−0.0691	32	−0.2164	42
A1.6	0.0688	19	−0.109	36	0.166	15	−0.2542	44
A1.7	0.266	8	0.1206	19	0.1454	17	0.1576	16
A1.8	0.3439	6	−0.2328	43	0.2429	6	−0.1904	39
A1.9	0.5807	1	0.0822	21	0.375	1	−0.0503	31
A1.10	0.3633	5	0.2535	7	0.0791	20	0.331	3
A2.1	0.2135	12	0.1806	15	0.06	22	−0.0646	32
A2.2	0.2112	13	−0.1792	40	−0.0771	33	−0.1477	36
A2.3	0.1569	15	0.2176	12	0.3438	3	0.0218	26
A2.4	−0.1771	37	−0.461	49	−0.2898	45	−0.2059	40
A2.5	−0.0354	29	−0.3127	44	−0.489	49	−0.4206	47
A2.6	−0.2167	39	−0.4513	48	−0.2253	43	−0.2073	41
A2.7	0.3804	3	−0.0909	34	0.2118	10	−0.0007	28
A2.8	0.1382	17	−0.4149	47	−0.2266	44	−0.0318	30
A2.9	0.4701	2	−0.0705	30	−0.1235	39	0.2014	10
A2.10	0.2434	9	0.0585	23	0.1664	14	0.1819	13
A3.1	0.0164	27	0.2076	14	0.241	7	0.1315	19
A3.2	−0.1272	33	0.1243	18	0.0325	24	0.0467	23
A3.3	0.2149	11	0.3647	3	0.1022	18	0.2166	8
A3.4	−0.3995	48	−0.0844	33	−0.1676	42	−0.2798	45

(*continued*)

Table 11. (*continued*)

Decision variant	p_1		p_2		p_3		p_4	
	Φ_{net}	Rank	Φ_{net}	Rank	Φ_{net}	Rank	Φ_{net}	Rank
A3.5	−0.2142	38	−0.167	39	−0.0615	31	0.2443	6
A3.6	−0.5065	50	−0.0382	27	0.1704	12	−0.4243	48
A3.7	−0.0198	28	0.3881	2	0.1689	13	0.1915	12
A3.8	−0.2498	42	0.506	1	0.2151	9	0.1454	17
A3.9	0.0873	18	0.1454	16	0.2816	4	0.1258	20
A3.10	−0.3923	47	0.2474	9	−0.1417	41	0.1975	11
A4.1	0.0242	25	0.0765	22	0.0571	23	0.045	24
A4.2	−0.0711	31	−0.1064	35	0.0075	25	−0.0925	33
A4.3	0.044	23	−0.0269	26	−0.0954	36	0.3619	1
A4.4	−0.4237	49	−0.3226	45	−0.45	48	−0.4676	50
A4.5	−0.105	32	−0.047	28	−0.4136	47	0.1065	21
A4.6	−0.3732	45	−0.1151	37	−0.0937	34	−0.1586	38
A4.7	−0.1361	34	0.3454	4	0.2229	8	0.235	7
A4.8	−0.2269	41	0.2622	6	0.1595	16	0.142	18
A4.9	0.1539	16	0.2532	8	0.195	11	0.1703	15
A4.10	−0.0514	30	−0.2158	42	−0.1196	37	0.2846	4
A5.1	0.0193	26	0.0876	20	0.0624	21	0.104	22
A5.2	0.0529	21	−0.0638	29	−0.1235	38	0.0035	27
A5.3	0.0672	20	0.1423	17	−0.0177	28	0.173	14
A5.4	−0.1739	36	−0.0831	32	−0.0946	35	0.2713	5
A5.5	−0.2243	40	−0.1856	41	−0.0292	29	−0.2407	43
A5.6	−0.374	46	−0.1242	38	−0.294	46	−0.37	46
A5.7	0.0458	22	0.2755	5	0.3536	2	−0.0127	29
A5.8	−0.356	44	−0.0165	24	−0.1379	40	−0.1303	35
A5.9	−0.2811	43	0.2359	10	0.0002	26	−0.1533	37
A5.10	0.2769	7	0.2298	11	0.0879	19	0.3515	2

Appendix 4

See Table 12.

Table 12. Global rankings of variants –TPEA and TPDA strategies

Decision variant	Advertiser's perspective				Publisher's perspective			
	TPEA		TPDA		TPEA		TPDA	
	Φ^{DM1}_{global}	Rank	Φ^{DM1}_{global}	Rank	Φ^{DM2}_{global}	Rank	Φ^{DM2}_{global}	Rank
A1.1	0.2527	8	0.2281	10	0.0476	24	0.0202	29
A1.2	0.2604	7	0.2410	8	−0.0404	32	−0.0662	34
A1.3	0.4158	2	0.4017	3	0.2612	1	0.2398	3

(*continued*)

Table 12. (*continued*)

Decision variant	Advertiser's perspective				Publisher's perspective			
	TPEA		TPDA		TPEA		TPDA	
	Φ^{DM1}_{global}	Rank	Φ^{DM1}_{global}	Rank	Φ^{DM2}_{global}	Rank	Φ^{DM2}_{global}	Rank
A1.4	−0.0876	30	−0.1183	31	−0.4500	50	−0.4941	50
A1.5	−0.0325	23	−0.0583	26	−0.1198	38	−0.1658	42
A1.6	−0.0965	31	−0.1216	32	−0.0321	30	−0.0668	35
A1.7	0.0581	18	0.0358	19	0.1724	9	0.1574	15
A1.8	−0.1056	33	−0.1301	34	0.0409	25	−0.0155	30
A1.9	−0.2309	45	−0.2597	44	0.2469	3	0.1669	12
A1.10	−0.1473	36	−0.1833	39	0.2567	2	0.2432	1
A2.1	0.3595	4	0.3182	5	0.0974	18	0.0496	23
A2.2	0.1936	12	0.1460	16	−0.0482	33	−0.0969	36
A2.3	0.2616	6	0.2243	11	0.1850	7	0.1711	10
A2.4	−0.0298	22	−0.0496	25	−0.2835	46	−0.2792	46
A2.5	−0.0732	27	−0.1016	30	−0.3144	48	−0.3810	48
A2.6	−0.2274	44	−0.2678	46	−0.2752	45	−0.2624	45
A2.7	−0.2331	46	−0.2827	49	0.1252	16	0.0831	19
A2.8	−0.1736	39	−0.2018	41	−0.1338	39	−0.1499	39
A2.9	−0.1820	40	−0.2210	43	0.1194	17	0.0764	21
A2.10	−0.1527	38	−0.1795	38	0.1626	12	0.1587	14
A3.1	0.4448	1	0.4947	1	0.1491	15	0.1681	11
A3.2	0.3469	5	0.3972	4	0.0191	26	0.0406	26
A3.3	0.3682	3	0.4038	2	0.2246	5	0.2117	5
A3.4	0.1083	16	0.1550	15	−0.2328	44	−0.2190	44
A3.5	−0.0080	20	0.0212	20	−0.0496	35	0.0245	28
A3.6	−0.0340	24	−0.0184	22	−0.1997	43	−0.1769	43
A3.7	0.0855	17	0.1005	17	0.1822	8	0.2029	6
A3.8	0.0524	19	0.0878	18	0.1542	14	0.1989	7
A3.9	−0.0967	32	−0.0649	29	0.1600	13	0.1726	9
A3.10	−0.1368	35	−0.1239	33	−0.0223	28	0.0467	24
A4.1	0.1823	13	0.1934	13	0.0507	23	0.0529	22
A4.2	0.1581	15	0.1843	14	−0.0656	36	−0.0631	33
A4.3	0.1633	14	0.2083	12	0.0709	21	0.1152	17
A4.4	−0.1524	37	−0.1427	35	−0.4160	49	−0.4289	49
A4.5	−0.0477	25	−0.0259	23	−0.1148	37	−0.1014	37
A4.6	−0.0143	21	0.0001	21	−0.1852	42	−0.1519	40
A4.7	−0.1294	34	−0.1587	36	0.1668	10	0.2163	4
A4.8	−0.2417	47	−0.2649	45	0.0842	20	0.1344	16
A4.9	−0.2797	48	−0.2736	47	0.1931	6	0.1927	8
A4.10	−0.3564	50	−0.3437	50	−0.0256	29	0.0297	27
A5.1	0.2460	9	0.2778	6	0.0683	22	0.0798	20

(*continued*)

Table 12. (*continued*)

Decision variant	Advertiser's perspective				Publisher's perspective			
	TPEA		TPDA		TPEA		TPDA	
	Φ^{DM1}_{global}	Rank	Φ^{DM1}_{global}	Rank	Φ^{DM2}_{global}	Rank	Φ^{DM2}_{global}	Rank
A5.2	0.2159	10	0.2382	9	−0.0327	31	−0.0431	32
A5.3	0.2120	11	0.2515	7	0.0912	19	0.0991	18
A5.4	−0.0557	26	−0.0339	24	−0.0201	27	0.0461	25
A5.5	−0.0747	28	−0.0596	27	−0.1700	41	−0.1646	41
A5.6	−0.1886	41	−0.1779	37	−0.2906	47	−0.2984	47
A5.7	−0.0804	29	−0.0647	28	0.1656	11	0.1607	13
A5.8	−0.2110	43	−0.2069	42	−0.1602	40	−0.1324	38
A5.9	−0.2063	42	−0.1987	40	−0.0496	34	−0.0422	31
A5.10	−0.2994	49	−0.2752	48	0.2365	4	0.2406	2

References

1. Du, H., Xu, Y.: Research on multi-objective optimization decision model of web advertising - takes recruitment advertisement as an example. Int. J. Adv. Comput. Technol. **4**(10), 329–336 (2012)
2. Langheinrich, M., Nakamura, A., Abe, N., Kamba, T., Koseki, Y.: Unintrusive customization techniques for web advertising. Comput. Netw. **31**(11–16), 1259–1272 (1999)
3. Kazienko, P., Adamski, M.: AdROSA - adaptive personalization of web advertising. Inf. Sci. **177**(11), 2269–2295 (2007)
4. Chickering, D.M., Heckerman, D.: Targeted advertising with inventory management. In: Proceedings of 2nd ACM Conference on Electronic Commerce, pp. 145–149 (2000)
5. Nakamura, A., Abe, N.: Improvements to the linear programming based scheduling of web advertisements. Electron. Commer. Res. **5**, 75–98 (2005)
6. Zha, W., Wu, H.D.: The impact of online disruptive ads on users' comprehension, evaluation of site credibility, and sentiment of intrusiveness. Am. Commun. J. **16**(2), 15–28 (2014)
7. Goldstein, D.G., McAfee, R.P., Suri, S.: The cost of annoying ads. In: Proceedings of 22nd International Conference on World Wide Web, pp. 459–470 (2013)
8. Brajnik, G., Gabrielli, S.: A review of online advertising effects on the user experience. Int. J. Hum.-Comput. Interact. **26**(10), 971–997 (2010)
9. Kelly, L., Kerr, G., Drennan, J.: Avoidance of advertising in social networking sites. The teenage perspective. J. Interact. Advert. **10**(2), 16–27 (2010)
10. Jankowski, J., Ziemba, P., Wątróbski, J., Kazienko, P.: Towards the tradeoff between online marketing resources exploitation and the user experience with the use of eye tracking. In: Nguyen, N.T., Trawiński, B., Fujita, H., Hong, T.-P. (eds.) ACIIDS 2016. LNCS (LNAI), vol. 9621, pp. 330–343. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49381-6_32
11. Jankowski, J., Wątróbski, J., Ziemba, P.: Modeling the impact of visual components on verbal communication in online advertising. In: Núñez, M., Nguyen, N.T., Camacho, D., Trawiński, B. (eds.) ICCCI 2015. LNCS (LNAI), vol. 9330, pp. 44–53. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24306-1_5

12. Jankowski, J., Wątróbski, J., Ziemba, P.: Dynamic MCDA approach to multilevel decision support in online environment. In: Nguyen, N.-T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds.) ICCCI 2016. LNCS (LNAI), vol. 9875, pp. 553–564. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45243-2_51
13. Campanella, G., Ribeiro, R.A.: A framework for dynamic multiple-criteria decision making. Decis. Support Syst. **52**(1), 52–60 (2011)
14. Hashemkhani Zolfani, S., Maknoon, R., Zavadskas, E.K.: An introduction to Prospective Multiple Attribute Decision Making (PMADM). Technolog. Econ. Dev. Econ. **22**(2), 309–326 (2016)
15. Novak, T.P., Hoffman, D.L.: New metrics for new media: toward the development of web measurement standards. World Wide Web J. **2**(1), 213–246 (1997)
16. Pavlou, P.A., Stewart, D.W.: Measuring the effects and effectiveness of interactive advertising: a research agenda. J. Interact. Advert. **1**(1), 61–77 (2000)
17. Hoffman, D.L., Novak, T.P.: When exposure-based advertising stops making sense (And what CDNOW did about it). http://yaffecenter.org/downloads/Complete_List_of_Working_Papers/cdnow.pdf. Accessed 31 May 2017
18. Wu, G.: Applying the Rossiter-Percy grid to online advertising planning. The role of product/brand type in previsit intentions. J. Interact. Advert. **8**(1), 15–22 (2007)
19. Kim, C., Park, S., Kwon, K., Chang, W.: How to select search keywords for online advertising depending on consumer involvement: an empirical investigation. Expert Syst. Appl. **39**(1), 594–610 (2012)
20. Chakrabarti, D., Kumar, R., Radlinski, F., Upfal, E.: Mortal multi-armed bandits. In: Koller, D., Schuurmans, D., Bengio, Y. (eds.) Advances in Neural Information Processing Systems, vol. 21, pp. 273–280 (2008)
21. Chakrabarti, D., Agarwal, D., Josifovski, V.: Contextual advertising by combining relevance with click feedback. In: Proceeding of 17th International Conference on World Wide Web, pp. 417–426 (2008)
22. Fan, T.K., Chang, C.H.: Blogger-centric contextual advertising. Expert Syst. Appl. **38**(3), 1777–1788 (2011)
23. Amiri, A., Menon, S.: Efficient scheduling of internet banner advertisements. ACM Trans. Internet Technol. **3**(4), 334–346 (2003)
24. Gupta, N., Khurana, U., Lee, T., Nawathe, S.: Optimizing display advertisements based on historic user trails. In: Proceedings of ACM SIGIR, Workshop: Internet Advertising (2011)
25. Giuffrida, G., Reforgiato, D., Tribulato, G., Zarba, C.: A banner recommendation system based on web navigation history. In: Proceedings of IEEE Symposium on Computational Intelligence and Data Mining (CIDM), pp. 291–296 (2011)
26. Urban, G.L., Liberali, G., MacDonald, E., Bordley, R., Hauser, J.R.: Morphing banner advertising. Mark. Sci. **33**(1), 27–46 (2014)
27. Zorn, S., Olaru, D., Veheim, T., Zhao, S., Murphy, J.: Impact of animation and language on banner click-through rates. J. Electron. Commer. Res. **13**(2), 173–183 (2012)
28. Yoo, C.Y., Kim, K.: Processing of animation in online banner advertising: the roles of cognitive and emotional responses. J. Interact. Mark. **19**(4), 18–34 (2005)
29. Lang, A.: The limited capacity model of mediated message processing. J. Commun. **50**(1), 46–70 (2000)
30. Turatto, M., Galfano, G.: Color, form and luminance capture attention in visual search. Vis. Res. **40**(13), 1639–1643 (2000)
31. Rosenkrans, G.: The creativeness and effectiveness of online interactive rich media advertising. J. Interact. Advert. **9**(2), 18–31 (2009)
32. Yoo, C.Y., Kim, K., Stout, P.A.: Assessing the effects of animation in online banner advertising. hierarchy of effects model. J. Interact. Advert. **4**(2), 49–60 (2004)

33. Ziemba, P., Jankowski, J., Wątróbski, J., Piwowarski, M.: Web projects evaluation using the method of significant website assessment criteria detection. In: Nguyen, N.T., Kowalczyk, R. (eds.) Transactions on Computational Collective Intelligence XXII. LNCS, vol. 9655, pp. 167–188. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49619-0_9
34. Ziemba, P., Wątróbski, J., Jankowski, J., Wolski, W.: Construction and restructuring of the knowledge repository of website evaluation methods. In: Ziemba, E. (ed.) Information Technology for Management. LNBIP, vol. 243, pp. 29–52. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30528-8_3
35. Tsafarakis, S., Lakiotaki, K., Matsatsinis, N.: Applications of MCDA in marketing and e-Commerce. In: Zopounidis, C., Pardalos, P.M. (eds.) Handbook of Multicriteria Analysis, pp. 425–448. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-540-92828-7_15
36. Celik, M., Deha Er, I.: Fuzzy axiomatic design extension for managing model selection paradigm in decision science. Expert Syst. Appl. **36**(3), 6477–6484 (2009)
37. Wątróbski, J., Jankowski, J., Ziemba, P.: Multistage performance modelling in digital marketing management. Econ. Sociol. **9**(2), 101–125 (2016)
38. Wątróbski, J., Ziemba, P., Jankowski, J., Zioło, M.: Green energy for a green city - a multi-perspective model approach. Sustainability **8**(8), 702 (2016)
39. Brans, J.P., Mareschal, B.: PROMETHEE methods. In: Figueira, J., Greco, S., Ehrgott, M. (eds.) Multiple Criteria Decision Analysis: State of the Art Surveys. International Series in Operations Research & Management Science, vol. 78, pp. 163–195. Springer, Boston (2005). https://doi.org/10.1007/0-387-23081-5_5
40. Saltelli, A., Tarantola, S., Chan, K.: A role for sensitivity analysis in presenting the results from MCDA studies to decision makers. J. Multi-Criter. Decis. Anal. **8**, 139–145 (1999)
41. Mareschal, B., Brans, J.P.: Geometrical representations for MCDA. Eur. J. Oper. Res. **34**, 69–77 (1988)
42. Yu, P.L., Chen, Y.C.: Dynamic MCDM, habitual domains and competence set analysis for effective decision making in changeable spaces. In: Ehrgott, M., Figueira, J.R., Greco, S. (eds.) Trends in Multiple Criteria Decision Analysis. International Series in Operations Research & Management Science, vol. 142, pp. 1–35. Springer, New York (2010). https://doi.org/10.1007/978-1-4419-5904-1_1
43. Zulueta, Y., Martell, V., Martínez, J., Martínez, L.: A dynamic multi-expert multi-criteria decision making model for risk analysis. In: Castro, F., Gelbukh, A., González, M. (eds.) MICAI 2013. LNCS (LNAI), vol. 8265, pp. 132–143. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-45114-0_11
44. Jassbi, J.J., Ribeiro, R.A., Varela, L.R.: Dynamic MCDM with future knowledge for supplier selection. J. Decis. Syst. **23**(3), 232–248 (2014)
45. Chen, Y., Li, K.W., He, S.: Dynamic multiple criteria decision analysis with application in emergency management assessment. In: IEEE International Conference on Systems Man and Cybernetics (SMC), pp. 3513–3517 (2010)
46. Agrell, P.J., Wikner, J.: An MCDM framework for dynamic systems. Int. J. Prod. Econ. **45** (1–3), 279–292 (1996)
47. Kornbluth, J.S.H.: Dynamic multi-criteria decision making. J. Multi-Criter. Decis. Anal. **1** (2), 81–92 (1992)
48. Guitouni, A., Martel, J.M., Vincke, P.: A framework to choose a discrete multicriterion aggregation procedure. Defence Research Establishment Valcatier (1998)
49. Zenebe, A., Zhou, L., Norcio, F.: User preferences discovery using fuzzy models. Fuzzy Sets Syst. **161**(23), 3044–3063 (2010)
50. Johnson, T.: The 2016 Amazon Average CPC by Category Report. http://www.cpcstrategy.com/blog/2016/09/2016-amazon-sponsored-products-average-winning-bids-by-category/. Accessed 21 May 2017

51. Jankowski, J., Kazienko, P., Wątróbski, J., Lewandowska, A., Ziemba, P., Zioło, M.: Fuzzy multi-objective modeling of effectiveness and user experience in online advertising. Expert Syst. Appl. **65**, 315–331 (2016)
52. Podvezko, V., Podviezko, A.: Dependence of multi-criteria evaluation result on choice of preference functions and their parameters. Technolog. Econ. Dev. Econ. **16**(1), 143–158 (2010)
53. Deshmukh, S.C.: Preference ranking organization method of enrichment evaluation (PROMETHEE). Int. J. Eng. Sci. Invent. **2**(11), 28–34 (2013)
54. Roy, B.: The outranking approach and the foundations of electre methods. In: Bana e Costa, C.A. (ed.) Readings in Multiple Criteria Decision Aid, pp. 155–183. Springer, Heidelberg (1990). https://doi.org/10.1007/978-3-642-75935-2_8
55. Wątróbski, J., Ziemba, P., Jankowski, J., Wolski, W.: PEQUAL – e-Commerce websites quality evaluation methodology. In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds.) Proceedings of 2016 Federated Conference on Computer Science and Information Systems. ACSIS, vol. 8, pp. 1317–1327 (2016). https://doi.org/10.15439/2016f46
56. Chmielarz, W., Zborowski, M.: Aspects of mobility in e-Marketing from the perspective of a customer. In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds.) Proceedings of 2016 Federated Conference on Computer Science and Information Systems. ACSIS, vol. 8, pp. 1329–1333 (2016). https://doi.org/10.15439/2016f46
57. Ziemba, P., Wątróbski, J., Karczmarczyk, A., Jankowski, J., Wolski, W.: Integrated approach to e-commerce websites evaluation with the use of surveys and eye tracking based experiments. In: Proceedings of 2017 Federated Conference on Computer Science and Information Systems. ACSIS, vol. 11, pp. 1019–1030 (2017). https://doi.org/10.15439/2017f320

Biological Regulation and Psychological Mechanisms Models of Adaptive Decision-Making Behaviors: Drives, Emotions, and Personality

Amine Chohra(✉) and Kurosh Madani

Images, Signals, and Intelligent Systems Laboratory (LISSI/EA 3956),
Paris-East University (UPEC), Senart Institute of Technology,
Avenue Pierre Point, 77127 Lieusaint, France
{chohra,madani}@u-pec.fr

Abstract. The aim of this paper is to suggest a framework for adaptive agent decision-making modeling of biological regulation and psychological mechanisms. For this purpose, first, a perception-action cycle scheme for the agent-environment interactions and the deduced framework for adaptive agent decision-making modeling are developed. Second, motivation systems: drives (homeostatic regulation), personality traits (five-factor model), and emotions (basic emotions) are developed. Third, a neural architecture implementation of the framework is suggested. Then, first tests related to a stimulation-drive (from a moving object), for two different agent personalities, and the activation level of emotions are presented and analyzed. Finally, a discussion is given in order to highlight important problems related to the adaptive decision-making behavior, the common currency that should have each system in the suggested framework, and the neural architecture, as well as to detail the ways they are solved. The obtained results demonstrate how the personality and emotion of the agent can be used to regulate the intensity of the interaction; predicting a promising result in future: to demonstrate how the nature of the interaction (stimulation-drive, social-drive, …) influences the agent behavior which could be very interesting for cooperative agents.

Keywords: Complex systems · Decision-making
Agent-environment interactions · Perception-action cycle
Adaptive goal-directed behavior

1 Introduction

The motivation mechanisms inspired from drives, personality traits, and emotions modulate the cognitive system of an agent to make it function better in a complex, unpredictable environment than it could with its cognitive system alone, i.e., to allow the agent to make better decisions, to learn more effectively, to interact more appropriately with others [1, 2].

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXIX, LNCS 10840, pp. 69–83, 2018.
https://doi.org/10.1007/978-3-319-90287-6_4

The cognitive science implies the cognition (computation, “information processing psychology”, manipulation of data structures stored in memory, formal operations carried out on symbol structures), perception, and action [3–5]. Its basic aim is identifying the functional architecture of cognition, in terms of rules and representations as well as a form that is more analog and more biologically plausible, that mediate thought.

Researchers from *artificial intelligence*, *computer science*, *brain and cognitive science*, and *psychology* have been oriented, by the end of 1980s, towards a new field to build intelligent machines called *embodied cognitive science* or *new artificial intelligence* or *behavior-based artificial intelligence* [6].

The brain does not run ‘programs’: it does something entirely different, i.e., it does not do mathematical proofs, but *controls behavior*, to ensure our survival [6]. The researchers from these various disciplines agreed that *intelligence* always manifests itself in *behavior* and consequently that we must understand the behavior [6]. In fact, a particular attention must be given on thinking and high-level cognition focusing on the interaction with the real world. This interaction is always mediated by a body, i.e., the intelligence needs to be ‘*embodied*’. This, has rapidly changed the research disciplines of artificial intelligence and cognitive science towards a new research field which is exerting more and more its influence on psychology, neurobiology, and ethology, as well as engineering science.

By another way, throughout recorded human intellectual history, there has been active debate about the nature of the role of emotions or “passions” in human behavior [7], with the dominant view being that passions are a negative force in human behavior [8]. By contrast, some of the latest research has been characterized by a new appreciation of the positive functions served by emotions [9]. An appreciation for the positive functions is not entirely new in behavioral science. Darwin, in 1872, was one of the first to hypothesize the adaptive mechanisms through which emotion might guide human behavior [10].

The great interest and large investigations in decision-making research associated with the emergence of behavioral decision theory, then, largely ignored the role played by the irrationality part (related to affect, in general) in decision-making [7]. However, with the research developments particularly psychology-related fields from 1990s, a great interest have been oriented towards the role of the irrationality part (related to affect) in decision-making.

The aim of this paper is to suggest a framework for adaptive agent decision-making modeling of biological regulation and psychological mechanisms. It integrates drives, personality traits, and emotions in order to:

- use this framework as a test bed to test, analyze, and compare different pertinent models of drives developed from biological regulation and survival of social organisms, personality traits developed from the field of personality psychology, and emotions, central aspects of biological regulation,
- analyze the impacts (effects), to assess the variation consequences of different personality and emotion aspects on the decision agents make,
- emphasize the adaptive behaviors emerging from agent-environment interactions.

This paper is an extended version of the research works, suggesting such a framework for adaptive agent decision-making modeling of biological regulation and psychological mechanisms, developed in [11]. In this extended version more details and numerous references are added throughout all the paper. In addition, in order to highlight important problems of the suggested framework and to detail the ways they are solved, a discussion is added. This discussion is mainly related to the adaptive decision-making behavior, the common currency that should have each system, and the suggested neural architecture.

Thus, first, a perception-action cycle scheme for the agent-environment interactions and the deduced framework for adaptive agent decision-making modeling are developed in Sect. 2. Second, motivation systems: drives, personality traits, and emotions are developed in Sect. 3. Third, a neural architecture implementation of the framework is suggested in Sect. 4. Then, in Sect. 5 first tests related to a stimulation-drive, for two different agent personalities, and the activation level of emotions are presented and analyzed. Finally, a discussion is given, in Sect. 6, in order to highlight important problems related to the adaptive decision-making behavior, the common currency that should have each system in the suggested framework, and the neural architecture, as well as to detail the ways they are solved.

2 Agent-Environment Interactions (Perception-Action Cycle)

The perception-action cycle scheme for the agent-environment interactions, suggested in Fig. 1, is inspired from [12].

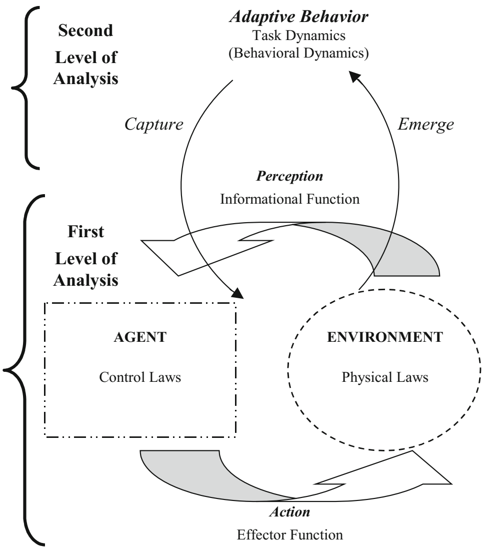


Fig. 1. Agent-environment interactions (perception-action cycle).

The agent and environment are coupled in two ways:

- an informational function (which maps properties of the agent-environment system into informational variables, in accordance with laws of ecological perception-action approach to the control of behavior):

$$\mathbf{i} = \lambda(\mathbf{e}), \tag{1}$$

where **i** is a vector of informational variables, **e** is a vector of environmental variables;

- an effector function (which transforms the vector of action variables into muscle activation patterns that produce forces in the environment, action is thus characterized as a relation defined over the agent, causal forces, and the environment):

$$\mathbf{f} = \beta(\mathbf{a}), \tag{2}$$

where **f** is a vector of external forces, and **a** is a vector of agent state variables (which describes the current state of the action system).

Thus, the adaptive, goal-directed behavior emerges from these local interactions between an agent governed by the control laws Ψ and an environment governed by the physical laws Φ such as:

$$\begin{aligned} \dot{\mathbf{a}} &= \Psi(\mathbf{a}, \mathbf{i}), \\ \dot{\mathbf{e}} &= \Phi(\mathbf{e}, \mathbf{f}). \end{aligned} \tag{3}$$

Indeed, adaptive behavior emerges from task dynamics or behavioral dynamics (information-based approach to perception, dynamical systems approach to action).

From this, the deduced framework for agent decision-making modeling in behavior-based systems is then suggested in Fig. 2.

This framework integrates particularly the motivation systems: drives (homeostatic regulation), personality traits (five-factors model), and emotions (basic emotions).

3 Motivation Systems

In this Section, the motivation systems consists of:

- an homeostatic regulation system implementing the drives of the agent,
- an emotion system implementing the emotions of the agent,
- a personality system implementing the personality traits of the agent.

3.1 Drives (Homeostatic Regulation)

The design of homeostatic regulation system is inspired by ethological views of the analogous process in animals [13, 14]. However, it is a simplified and idealized model of those discovered in living systems. The drive features are: its temporally cyclic behavior with three regimes (under-stimulated, homeostatic, and overwhelming), i.e., if

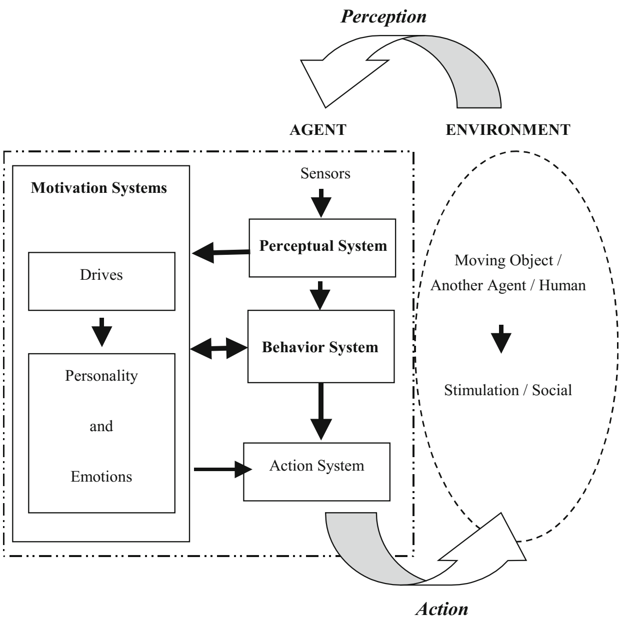


Fig. 2. A framework for agent decision-making modeling in behavior-based systems.

no stimulation, a drive will tend to increase in intensity unless it is satiated. This is analogous to an animal's degree of hunger or level of fatigue, both following a cyclical pattern [9, 15, 16].

3.2 Personality (Five-Factor Model)

Social psychologists believe that human behavior is determined by both a person's characteristics and the social situation. They also believe that the social situation is frequently a stronger influence on behavior than are a person's characteristics, [17]. From this purpose, it is very important to integrate the personality traits in the suggested framework.

The model in personality which appears to represent a major conceptual and empirical advance in the field of personality psychology is the five-factor model in personality (Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to experience) developed in [18].

3.3 Emotions (Basic Emotions)

Emotions are not a luxury, they play a role in communicating meanings to others, and they may also play the cognitive guidance role [9].

In fact, emotions are another important motivation system for complex organisms. They seem to be centrally involved in determining the behavioral reaction to environmental (often social) and internal events of major significance for the needs and goals of a creature [19, 20].

4 Neural Architecture Implementation

In the design and achievement of the architecture of the suggested framework for agent decision-making modeling, first, carry out the aspects of computation and cognition (issues in the foundations of cognitive science) related to cognitive functional architecture, respecting the biological and psychological nature under the cognitive impenetrability condition (non influence by purely cognitive factors as goals, beliefs, inferences, tacit knowledge, …) is of great importance [3, 4]. This allows the fixed capacities of mind (called its functional architecture) avoiding the particular representations and algorithms used on specific occasions. This in turn requires that the fixed architectural function and the algorithms be independently validated in order to examine the fundamental distinction between a behavior governed by rules and representations, and a behavior that is merely the result of the causal structure of the underlying biological system.

Second importance is related to the neural aspect of functional architecture which is highly distributed network of interacting neurons [5, 16, 21].

The basic computational process, implemented as a value based system (influences graded in intensity, instead of simply being "on" or "off"), is modeled by its activation level A_i which is computed by Eq. (4).

$$A_i = \left(\sum_{j=1}^{n} \omega_{ji}\, i_j\right) + b, \qquad (4)$$

where the inputs i_j are integer values, the weights ω_{ji}, and the bias b, over the number of inputs n. The process is active when the activation level A_i exceeds a threshold T.

The weights can be either positive or negative; a positive weight corresponds to an excitatory connection and a negative weight corresponds to an inhibitory connection.

Each perceptual unit, drive, emotion, personality, behavior, and motor process is modeled as a different type that is specifically tailored for its role in the overall architecture.

4.1 Perceptual Units

The antecedent conditions come through the perceptual system where they are assessed with respect to the agent's "well being" and active goals. Thus, the perceptual system, which is inspired from [14], is built of perceptual units:

– related to stimulus (from moving object, another agent, human) as shown in Fig. 3(a), i.e., there is a set of perceptual units defined for each stimulus, that indicate its presence (time: short, medium, long), absence (time: short, medium, long), nature (moving object, another agent, or human), quality (intensity of stimulus: too low, just right, too high), desirability (desired or not desired),

- related to drives (stimulation drive, social drive, …) as shown in Fig. 3(b), i.e., there is a perceptual unit defined for each regime (under-stimulated, homeostatic, overwhelming) of each drive, to represent how well each drive is being satiated,
- related to behaviors as shown in Fig. 4(a), i.e., there is a set of perceptual units defined for each behavior, that indicate whether its goal has been achieved or not, and if not, then for how long.

Note that the perceptual unit implementation is designed for each regime k (under-stimulated, homeostatic, overwhelming) of each drive i (stimulation-drive, social-drive, …).

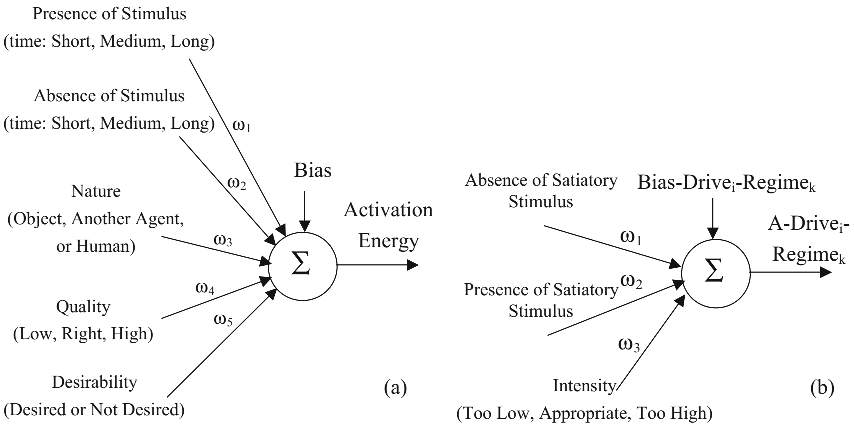


Fig. 3. (a) Perceptual unit implementation related to stimulus. (b) Perceptual unit implementation related to drives.

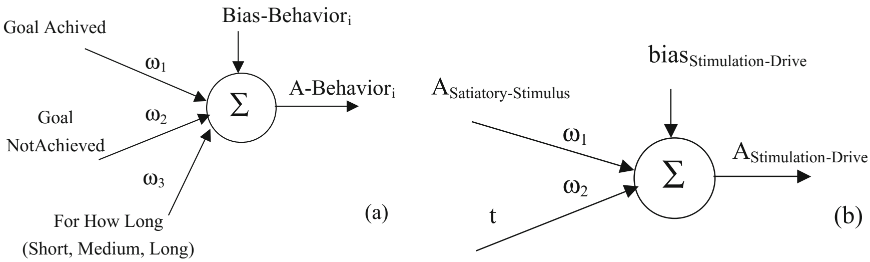


Fig. 4. (a) Perceptual unit implementation related to behaviors. (b) Drive implementation ($A_{\text{Stimulation-Drive}}$).

4.2 Drives

The drive model given in Fig. 4(b) concerns the Stimulation-Drive leading to a cyclic behavior of a drive, where t is a temporal input; given no stimulation, a drive will tend to increase in intensity ($A_{Stimulation\text{-}Drive}$) unless it is satiated (homeostatic regime). Note that a similar drive model will be implemented in future concerning Social-Drive.

4.3 Personality

The five-factor model in personality, suggested in Fig. 5, is inspired from [18, 22, 23] implying five broad dimensions which are used to describe human personality: Openness to experience (Op), Agreeableness (Ag), Conscientiousness (Co), Extraversion (Ex), and Neuroticism (Ne).

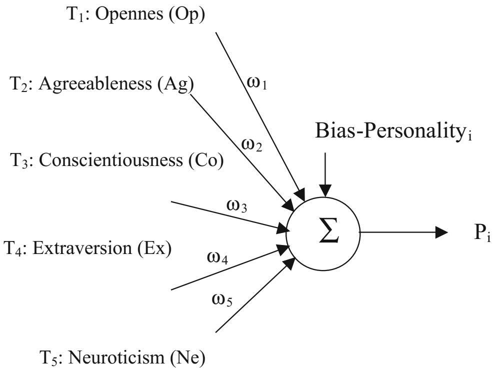


Fig. 5. Personality model.

Thus, combined personality value P_i affecting a behavior is defined in Eq. (5) as:

$$P_i = (\sum_{j=1}^{5} \omega_{ji}\ T_j) + b, \tag{5}$$

where T_j denotes the intensity of each personality parameter, and ω_{ji} the influence (inverse influence–1, no influence 0, direct influence +1) on a particular behavior.

4.4 Emotions

The relations between emotions and behavioural responses, i.e., under what conditions certain “emotions” and behavioural responses arise, are given in Table 1. This table is derived from the evolutionary, cross-species, and social functions hypothesized by [10, 19, 20]. Then, the perceptual behavioral or motivational information is tagged (arousal, valence, stance) with affective information [9]. Note that the stance is related to how approachable the percept is to the agent. Moreover, each regime of a drive biases arousal and valence differently, contributing to the activation of different emotions.

Table 1. Relations between emotions and behaviors under antecedent conditions.

Antecedent conditions	Emotion	Behavior
Difficulty in achieving goal	Anger	Complain
Presence of an undesired stimulus	Disgust	Withdraw
Presence of a threatening (overwhelming)	Fear	Escape
Prolonged presence of a desired stimulus	Calm	Engage
Success in achieving goal of active behavior	Joy	Display pleasure
Prolonged absence of a desired stimulus	Sadness	Display sorrow
A sudden, close stimulus	Surprise	Startle response
Appearance of a desired stimulus	Interest	Orient
Need of an absent and desired stimulus	Boredom	Seek

Table 2. Influence and intensity of personality parameters, Openness to experience (Op), Agreeableness (Ag), Conscientiousness (Co), Extraversion (Ex), and Neuroticism (Ne).

Agent	Op	Ag	Co	Ex	Ne
Personality1: Influence Intensity	0	**−1** **2**	0	0	**+1** **4**
Personality2: Influence Intensity	0	**−1** **3**	0	0	**+1** **7**

5 Tests, and Results

In this Section, first tests related to a Stimulation-Drive, from a Stimulus of a moving object, presented in Fig. 6, for two different agent personalities, and the activation level of emotions are presented and analyzed. Influence and intensity parameters presented in Table 2 concern two different personalities (since personality trait of Neuroticism has been found to influence avoidance behavior [24], Escape in Table 1 in our concern, and Agreeableness to inversely influence it):

In Fig. 8, the activation of emotions: Sadness (corresponding to prolonged absence of a desired stimulus, as shown in Table 1), followed by Interest (appearance of a desired stimulus), then Fear (presence of a threatening, corresponding to overwhelming stimulus, as shown in Fig. 7), followed by Interest (appearance of a desired stimulus), and finally Sadness (prolonged absence of a desired stimulus).

Note that In Fig. 8, Fear appears and crosses Interest the first time, around t = 65 s, for Agent-Personality1.

In Fig. 9, Fear appears and crosses Interest the first time, around t = 63 s, for Agent-Personality2. Note that in this case Fear appears and crosses Interest earlier than in the case of the Agent-Personality1; moreover, Fear in the case of the Agent-Personality2 reaches more intensity than in the case of the Agent-Personality1.

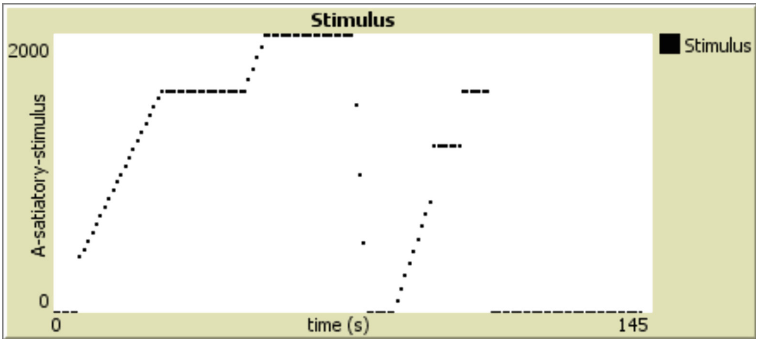


Fig. 6. The used Stimulus (A-satiatory-stimulus).

This can be explained by the following:

– in Personality 1, from Eq. (5):

$$P_i = (-1) * (2) + (+1) * 4 + b = 2 + b ;$$

– in Personality 2, from Eq. (5):

$$P_i = (-1) * (3) + (+1) * 7 + b = 4 + b.$$

This means that Personality 2 will demonstrate the avoidance behavior, Escape, (earlier and with great intensity) than in the case of Personality 1.

Note that, as personality traits of one agent, remain invariant throughout execution, the corresponding behavioral P_i is computed only once at the beginning of execution.

Thus, the personality effect is indirect, i.e., it influences emotion generation rather than the behaviors themselves.

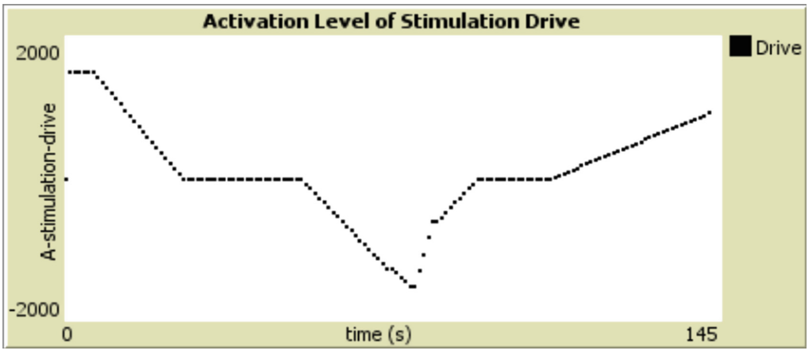


Fig. 7. Stimulation-drive results (from the stimulus, Fig. 6).

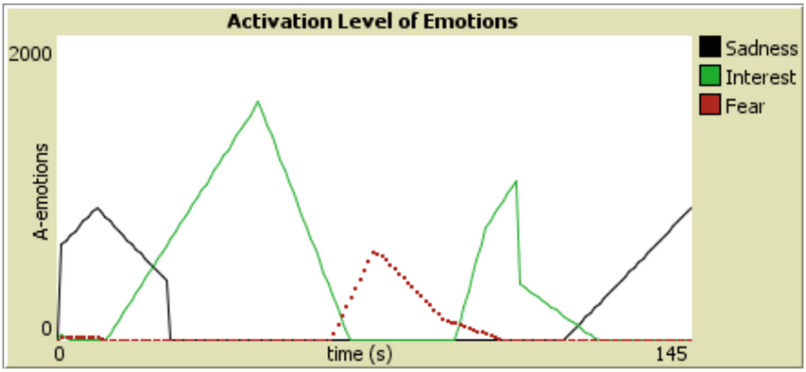


Fig. 8. Activation of emotions in case of Agent-Personality1 (from stimulation-drive Fig. 7).

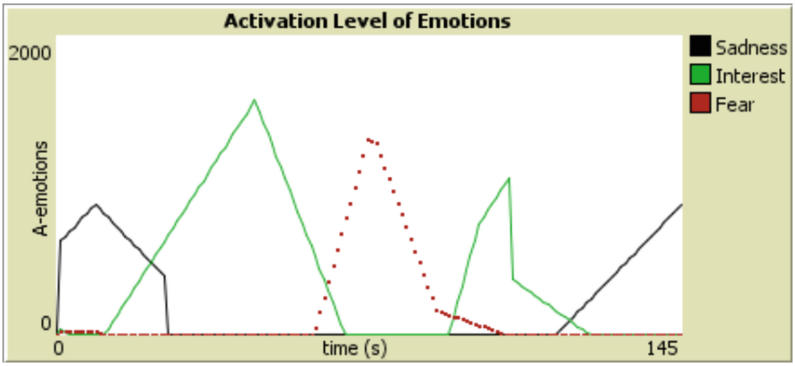


Fig. 9. Activation of emotions in case of Agent-Personality2 (from the same stimulus, Fig. 6).

6 Discussion

In this Section, a discussion is given in order to highlight important problems related to the adaptive decision-making behavior, the common currency that should have each system in the suggested framework, and the neural architecture, as well as to detail the ways they are solved.

6.1 Adaptive Decision-Making Behavior

In a changing, unpredictable, and more or less threatening environment, the behavior of an animal is adaptive so long as the behavior allows the animal to survive. Under the same conditions, the behavior of an agent (robot) is considered to be adaptive so long as the agent (robot) can continue to perform the functions for which it was built [25]. Under these circumstances, it is obvious that one can associate with an animat a certain number of state variables upon which its survival or successful operation depends, and that each of these state variables has a specific range within which the animat's continued survival or operation is preserved. Ashby referred to such variables long ago as essential variables [16, 26].

Indeed, Ashby proposes the definition that: a form of behavior is ‘adaptive’ if it maintains the essential variables within physiological limits. The thesis that ‘adaptation’ means the maintenance of essential variables within physiological limits is thus seen to hold not only over the simpler activities of primitive animals but over the more complex activities of the ‘higher’ organisms.

The ranges of such essential variables describe a zone of viability inside the given state space, allowing the animat to be referenced at any instant by a point within this zone [25]. Under the influence of environmental or behavioral variations affecting the animat, the corresponding reference point moves and may at times approach the limits of the viability zone. In this case, the animat’s behavior can be called adaptive so long as it avoids transgressing the boundary of viability [26, 27].

Such behavior can be generated by means of several different or complementary abilities and architectures [25]. For example, the laws governing the animat’s operation may rely upon various homeostatic mechanisms thanks to which, if the reference point alluded to earlier moves away from an adapted point of equilibrium (adapted because it is suitably located within the viability zone), this process tends to return it to its original position, thereby decreasing the risk that it will pass outside the limits of the zone. Other ways in which to lower this risk involve the use of high-quality sensory organs or motor apparatus that allows the animat to detect as early as possible that it is approaching these limits and/or to move away from them quickly and effectively. In this line of reasoning, it is obvious that the equivalent of a nervous system is mandatory in order to connect the animat’s perceptions with its actions and that reflex circuits activated as quickly as possible increase the adaptive nature of its behavior.

6.2 Common Currency

Ethology, comparative psychology, and neuroscience have shown that observable behavior is influenced by internal factors (i.e., motivations, past experience, …) and by external factors (i.e., perception) [2]. This demands that different types of systems be able to communicate and influence each other despite their different functions and modes of computation.

This has led ethologists such as McFarland and Bosser [13] and Lorenz [28] to suggest that there must be a *common currency*, shared by the perception, motivation, and behavior systems. In this scheme, the perception system generates values based on environmental stimuli, and the motivation system generates values based on internal factors. Both sets of values are passed to the behavior system, where competing behaviors use them to compute their relevance and then compete for expression (action system) based on this value. Within different systems, each can operate on their own currencies.

Furthermore, as the system becomes more complex, it is possible that some agents may conflict with others (such as when competing for shared resources). If each agent computes its relevance in terms of a shared currency, conflicting agents can compete based on this value.

An interesting theory is proposed in [29] according to which pleasure is such a common currency needed in order to achieve the ranking of motivations in case of conflict between two or more of them. In fact, the perception of pleasure, as measured operationally and quantitatively by choice behavior (in the case of animals), or by the rating of the intensity of pleasure or displeasure (in the case of humans) can serve as such a common currency. The tradeoffs between various motivations would thus be accomplished by simple maximization of pleasure.

6.3 Neural Architecture

It is common for biologically inspired architectures [2] to be constructed from a network of interacting elements such as subsumption architecture [30], neural networks [31], or agent architectures [32].

Inspired by models of intelligence in natural systems, the design of the neural architecture features both a cognitive system and a motivation system including drives, personality traits, and emotions. Both operate in parallel and are deeply intertwined to foster appropriately adaptive functioning of the agent in the environment as it interacts with people.

Thus, the suggested neural architecture is inspired from [2] and implemented as an agent architecture where each computational element is conceptualized as a specialist [32]. Hence, each drive, behavior, perception unit, emotion process is modeled as a different type of specialist that is specifically tailored for its role in the overall system architecture.

Hence, although the specialists differ in function, they all follow a basic activation scheme. In fact, units are connected to form networks of interacting processes that allow for more complex computation. This involves connecting the outputs of one unit to the inputs of other units. When a unit is active, besides passing messages to the units connected to it, it can also pass some of its activation energy. This is called spreading activation and is a mechanism (originally conceptualized by Lorenz [28]) by which units can influence the activation or suppression of other units. Such an activation mechanism, in the suggested neural architecture, is inspired by ethological models similar to that developed in [33].

7 Conclusion

In this paper, a framework for adaptive agent decision-making modeling, deduced from a perception-action cycle scheme, has been suggested. Then, motivation systems: drives (homeostatic regulation), personality traits (five-factor model), and emotions (basic emotions) have been developed. Afterwards, a neural architecture implementation of the framework has been suggested. Finally, a discussion has been given in order to highlight important problems related to the adaptive decision-making behavior, the common currency that should have each system in the suggested framework, and the neural architecture, as well as to detail the ways they are solved.

The first tests related to a stimulation-drive (from a moving object), for two different agent personalities, and the activation level of emotions are presented and analyzed. The obtained results demonstrate how the personality and emotion of the agent can be used to regulate the intensity of the interaction; predicting a promising result in future: to demonstrate how the nature of the interaction (stimulation-drive, social-drive, …) influences the agent behavior which could be very interesting for cooperative agents.

After investigating the behavior system, the action system, and the social-drive to test the interactions agent-agent and human-agent, it is interesting to investigate different cooperative strategies with different emotion regulation strategies [34], and the learning from interaction [6, 35].

References

1. Graham, S., Weiner, B.: Theories and principles of motivation, Chap. 4. In: Handbook of Educational Psychology, pp. 63–84. Macmillian, New York (1996)
2. Breazeal, C., Brooks, R.A.: Robot emotion: a functional perspective. In: Fellous, J-M., Arbib, M. (Eds.) Who Needs Emotions: The Brain Meets the Robot, pp. 271–310. MIT Press (2005)
3. Pylyshyn, Z.W.: Computation and cognition: issues in the foundations of cognitive science. Behav. Brain Sci. **3**, 111–169 (1980)
4. Pylyshyn, Z.W.: Computation and Cognition: Towards a Foundation for Cognitive Science. The MIT Press, Cambridge (1984)
5. Dawson, M.R.W.: From embodied cognitive science to synthetic psychology. In: Proceedings of the First IEEE International Conference on Cognitive Informatics, pp. 13–22 (2002)
6. Pfeifer, R., Scheier, C.: Understanding Intelligence. MIT press, Cambridge (1999)
7. Loewenstein, G., Lerner, J.S.: The role of affect in decision-making. In: Handbook of Affective Sciences, pp. 619–642. Oxford University Press (2003)
8. Elster, J.: Alchemies of the Mind: Rationality and the Emotions. Cambridge University Press, Cambridge (1999)
9. Damasio, A.R.: Descartes' Error: Emotion, Reason, and Human Brain. Putnam, New York (1994)
10. Darwin, C.: The Expression of the Emotions in Man and Animals, 3rd edn. Oxford University Press, New York (1998). Original Work Published in 1872
11. Chohra, A., Madani, K.: Biological regulation and psychological mechanisms models of adaptive decision-making behaviors: drives, emotions, and personality. In: Nguyen, N.-T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds.) ICCCI 2016. LNCS (LNAI), vol. 9875, pp. 412–422. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45243-2_38
12. Warren, W.H.: The dynamics of perception and action. Psychol. Rev. **113**(2), 358–389 (2006)
13. McFarland, D., Bosser, T.: Intelligent Behavior in Animals and Robots. MIT Press, Cambridge (1993)
14. Breazeal, C.: Emotion and sociable humanoid robots. Int. J. Hum Comput Stud. **59**, 119–155 (2003)
15. Walter, W.G.: The Living Brain. W. W. Norton & Co., New York (1963)
16. Ashby, W.R.: Design for A Brain: The Origin of Adaptive Behaviour, 2nd edn. Wiley/Chapman & Hall, New York/London (1960)

17. McDougall, W.: An Introduction to Social Psychology, 14th edn. Batoche Books, Kitchener (2001)
18. McAdams, D.P.: The five-factor model in personality: a critical appraisal. J. Pers. **60**(2), 328–361 (1992)
19. Izard, C.: Human Emotions. Plenum, New York (1977)
20. Plutchik, R.: The Emotions. University Press of America, Lanham (1991)
21. Pickering, A.: The Cybernetic Brain: Sketches of Another Future. The University of Chicago Press, Chicago and London (2010)
22. Moshkina, L., Arkin, R.C.: On TAMEing robots. In: International IEEE Conference on Systems, Man, and Cybernetics, vol. 4, pp. 3949–3956 (2003)
23. Egges, A., Kshirsagar, S., Magnenat-Thalmann, N.: A model for personality and emotion simulation. In: Palade, V., Howlett, Robert J., Jain, L. (eds.) KES 2003. LNCS (LNAI), vol. 2773, pp. 453–461. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45224-9_63
24. Elliot, A.J., Thrash, T.M.: Approach avoidance motivation in personality: approach and avoidance temperaments and goals. J. Pers. Soc. Psychol. Am. Psychol. Assoc. **82**(5), 804–818 (2000)
25. Meyer, J.A.: Artificial Life and the animat approach to artificial intelligence. In: Artificial Intelligence. Academic Press (1995)
26. Ashby, W.R.: Design for a Brain. Chapman & Hall, London (1952)
27. Sibly, R.M., McFarland, D.: On the fitness of behavior sequences. Am. Nat. **110**, 601–617 (1976)
28. Lorenz, K.: Foundations of Ethology. Springer, New York (1973). https://doi.org/10.1007/978-3-7091-3671-3
29. Cabanac, M.: Pleasure: the common currency. J. Theor. Biol. **155**, 173–200 (1992)
30. Brooks, R.: A robust layered control system for a mobile robot. IEEE J. Robot. Autom. **2**, 253–262 (1986)
31. McCulloch, W., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**, 115–133 (1943)
32. Minsky, M.: The Society of Mind. Simon et Schuster, New York (1988)
33. Blumberg, B.M., Todd, P.M., Maes, P.: No bad dogs: ethological lessons for learning. In: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior (SAB96), pp. 295–304. MIT Press, Cambridge (1996)
34. Abro, A.H., Manzoor, A., Tabatabei, S.A., Treur, Y.: A computational cognitive model integrating different emotion regulation strategies. In: Annual International Conference On Biologically Inspired Cognitive Architectures, vol. 71, pp. 157–168. Elsevier (2015)
35. Chohra, A.: Embodied cognitive science, intelligent behavior control, machine learning, soft computing, and fpga integration: towards fast, cooperative and adversarial robot team (RoboCup). Technical Fraunhofer (GMD) Report. Germany, no. 136 (2001)

Check for updates

SWRL-Based Recommendation System for Provision of the First Aid

Martina Husáková(✉)

Faculty of Informatics and Management,
University of Hradec Králové, Hradec Králové, Czech Republic
martina.husakova.2@uhk.cz

Abstract. The emergency management offers a collection of methods, strategies and frameworks how to efficiently solve particular emergency situation for saving human lives with minimum spending time. Fast and faultless decision making is inevitable in these situations. This requires huge amount of theoretical and practical experience which can be gained during studying of particular courses, study programmes, or where we are directly confronted with a reality. Non-professionals have a problem to decide about the best sequence of steps which has to be applied if a particular emergency event occurs. The main aim of this text is to present the SWRL-based prototype for decision making if the first aid is necessary.

Keywords: Emergency management · First aid · OWL ontology
SWRL

1 Introduction

Saving of human lives is a primary domain of the professional rescuers and the medical doctors. They apply deep theoretical knowledge and practical experience in solving critical situations that endanger human lives. They often deal with incomplete and vague pieces of knowledge. Stress and strict time limits complicate faultless and efficient decision making. If a situation is really serious, more rescuers should be available. In this case, cooperation and coordination activities have to be well organised. Professional rescuers are not the only ones who are in touch with sudden or unpredictable life endangering events. Common civilians can meet a situation when have to help an injured person. Accurate definition of the first aid is presented e.g. in [15]: "*First aid is defined as a set of simple actions which any person can take regardless of education level or a special training*." Ordinary situation is that common civilian did not pass any intensive rescue course or study programme. Somebody often supposes that provision of the first aid is a domain only for professionals. This incorrect confidence can lead to fatal impacts when our help will be inevitable in the future. The problem is that non-professional rescuers are not able to decide about a state of an injured person and which sequence of steps should be applied. This text is focused on a conceptualisation of the first aid application domain with the main aim to build a recommendation-based system providing advices in cases of the emergency incident. Ontologies are information and knowledge-based structures playing important role in modelling complex knowledge. Particular research papers prove that ontologies can be used for knowledge

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXIX, LNCS 10840, pp. 84–102, 2018.
https://doi.org/10.1007/978-3-319-90287-6_5

representation for the needs of the emergency management, but it is find out that this type of knowledge-based schema is not often applied in case of the first aid. The text introduces how the OWL (Ontology Web Language) ontology can be used for representation of this complex domain together with the rules providing particular recommendations. This text extends the research that has been introduced in [7, 18]. The paper [7] presents the approach (conceptual framework) for systematic development of the decision-support system for provision of the first aid. This framework is briefly mentioned in the Sect. 3 of this text. The presented text extends the paper [18] in the view of the usage of ontologies in praxis, in the emergency management, especially in provision of the first aid, see Sect. 2. It adds the comparison of the mind maps and concept maps which are considered for non-formal knowledge modelling, see Sub-sect. 3.2. Complete view on concept mapping used in development of the ontology-based recommendation system is provided, see Sub-sect. 3.2. The paper presents comparison of the various versions of the Protégé tool, see Sub-sect. 3.3 and briefly compares developed the FAO ontology with similar ontologies which are also focused on the first aid.

The text is structured as follows. The Sect. 2 presents ontologies as structures for knowledge modelling, especially in the context of the emergency management. The Sect. 3 presents development of the SWRL-based recommendation prototype. The Sect. 4 introduces discussion and future directions. The Sect. 5 concludes the text.

2 Ontological Engineering in Emergency Management

Ontologies are information and knowledge structures representing concepts and relations between them that occur in an application domain. Ontologies can be distinguished according to various criteria, e.g. according to their specificity. Upper-level ontologies model substantial entities which can occur across various domains, e.g. time, space, topological properties or relations between parts of a whole. Domain-specific ontologies are more focused on particular problem that is modelled in many details in comparison to the upper-level ontologies. Ontologies should facilitate communication between humans, computers or between humans and computers. General purpose of these structures is to provide an overview about particular application domain with the aim to offer a vocabulary of "things" that is approved by a group of interested parties or integrate it into a final application for problem solving or decision making. Necessary attributes of ontologies are mentioned in two of the most known definitions of ontologies:

- Gruber [8]: "*An ontology is an explicit specification of a conceptualisation.*"
  - *Explicit specification*: An ontology should not be hidden only in the heads of the experts, but it should be available for public.
  - *Conceptualisation* is a process of specification which concepts and relations between them are important for our problem solution.
- Borst [3]: "*An ontology is a formal specification of a shared conceptualisation*."
  - *Formal specification*: If an ontology should be used as a vocabulary of "things" that is machine-processable, a formal language has to be used for its representation.

– *Shared conceptualisation*: An ontology models particular fragment of our reality, but various people can have different views on this reality. An ontology can be perceived as a result of a consensus between different groups of people, i.e. how this fragment of our world looks like according to our notion. An ontological model should be available for next processing by various interested parties.

Ontologies are applied in various application domains. A lot of case studies and software solutions prove that ontologies are useful also for the emergency management. The emergency management is a group of processes, strategies, rules and technologies which are used for avoiding of emergency events or reducing their impacts. A review of the ontological applications in the emergency management is introduced below. On the basis of this review, it is found out that the research results in the following outputs:

1. *An ontology-based conceptual model*: An ontology-based formal model is proposed and its authors often suppose that it is going to be used in particular software solution in the future.
2. *Framework*: Authors propose a conceptual or a programming framework that integrates principles of the ontological engineering with the aim to solve problem in the emergency management.
3. *Practical application of an ontology*: Authors introduce ontologies which are used in their real applications. Building of these ontology-based solutions is in various stages of development (an initial study, almost complete or fully complete solution).

The following section introduces selected research studies which can be divided into the above mentioned thematic groups.

A medical OWL ontology is proposed in [9]. This ontology clarifies the most important concepts and relations between them in the emergency application domain. The ontology is built in the Protégé tool with the Methontology methodology [10] that can be used if we develop ontology from scratch or if we would like to reuse already existing ones for our purposes. A group of experts is used for ontology building and review of number of papers and manuals is realised for information extraction for conceptualisation. The main aim of the ontology is to support understanding, communication and knowledge sharing about the emergency domain. Authors of the paper [11] propose a domain ontology used for the disaster management where heterogeneous data have to be created, maintained and accessed for effective decision making. The ontology is developed with hybrid approach where a combination of a top-down and a bottom-up modelling technique is used for conceptualisation together with proposition of various ontology types, i.e. the data-based domain ontologies representing particular concepts occurring in the disaster management (cadaster data, topological data, etc.), the local data-based ontologies describing semantic content of data and organisational ontologies modelling organisational structure and responsibilities of each user of the ontology-based system.

Guidelines and frameworks can bring added value into decision making in the emergency management. They can provide solutions which are based on experience and deep knowledge which are valuable especially if a time-dependent problem solving is necessary. As an example, particular OWL ontology-based conceptual framework is proposed for improving shared situation awareness among teams of rescuers in case of emergency incidents [4]. Mass evacuation during tsunami event is the case study for demonstration of the framework. Han and Xu propose a framework-based decision support system used for solving of crisis situations. The framework combines four types of ontologies for provision of the final solution [5]. Practical usage of the framework is demonstrated with the Tianjin Port Explosion case study.

Telemedicine is a part of the eHealth initiative providing approaches how to offer medical services if a patient is far away from the medical centres. An OWL ontology-based telemedicine consultation system is proposed in [12]. The system helps with medical services for people which are far away, e.g. on seagoing vessels. Particular symptoms and patient medical history is represented in the OWL ontology. A JSP-based application is used for sending all required information for emergency assistance towards particular maritime telemedical centre. This ontology is built in the Protégé 3.4.1 where SPARQL queries are used for data retrieval from the ontology. SPARQL (Simple Protocol and Query Language) is a protocol and query language for accessing data represented by the RDF (Resource Description Framework) graphs that are used for data exchange and information representation on the web.

Well-prepared emergency plans are also crucial in the emergency management. Experience mentioned in [13] demonstrates that most of these plans are in plain-text documents which make difficult to realise efficient administrative work, e.g. plan generation, documentation and maintenance. A prototype of an emergency plan training system is proposed. The system supports efficient emergency training process for metro staff and its operations. Ontologies are used for formalisation of these emergency plans and related concepts.

Smart homes dispose heterogeneous hardware and software components which have to communicate with each other. The amount of components and intensive communication between them increase a complexity of these environments. An ontology-based system for anomalies detection is introduced in [14]. The ontology represents various types of anomalies, i.e. hardware, software, network, operator and context-based which can be detected by the smart home system. Proposed solution is designed as a framework for facilitation definition of anomalies and rules how to react at some inappropriate changes in the system. We can imagine that this system could be also applied in the emergency management where the elderly people can be monitored in their homes and if something unusual in their behaviour occurs the system detects it and provides a solution.

The first aid can be defined as a collection of more or less complex emergency activities and decisions which are used for ensuring homeostasis in an organism of a human. Any person should be able to provide the first aid regardless of an education level or a special training [15]. Implementation of an OWL ontology-based first aid assistance service (Smart-M3 platform) is introduced in [15]. The OWL ontology represents fundamental participants, their roles and relations between them in case of the first aid provision. The system is aimed at developing design of mobile smart applications which should be provided in the Android devices. Adaptive learning systems can bring added value into the educational process where an individual approach to students can be applied. An ontology-based adaptive learning system is proposed in [16]. The system integrates OWL domain ontology representing the first aid application domain. This ontology has an explanatory function – the content of the ontology is used as a tutorial, and an evaluative function – the content of the ontology is used for preparation of multiple choice tests. This case study demonstrates that ontology can be used in education for multiple purposes. Teaching and learning cannot be realised only in a real world. Virtual spaces are promising alternatives where the students can learn how to solve problems in a real world. A frame-based ontology is used for modelling pieces of knowledge how to apply the first aid [17]. This ontology is used by the automated scenario-based training system, i.e. for development of the scenario generator component of this system. This component uses the ontology especially for creation a back story consisting of events taking place prior to injury.

3 Conceptual Framework for Building of SWRL-Based Recommendation System

Provision of the first aid is a typical example of a complex domain where a lot of inputs have to be evaluated and particular pieces of information have to be combined with respect to an individual state of an injured person. A conceptual framework is proposed for systematic building of the SWRL-based recommendation system. The conceptual framework integrates six phases of development: *knowledge acquisition*, *concept mapping*, *knowledge base development* including *ontology development* and *implementation of rules*, *instance data modelling* and *testing of the software solution* [7]. These phases are visualised in the Fig. 1 and briefly explained in the following sub-sections.

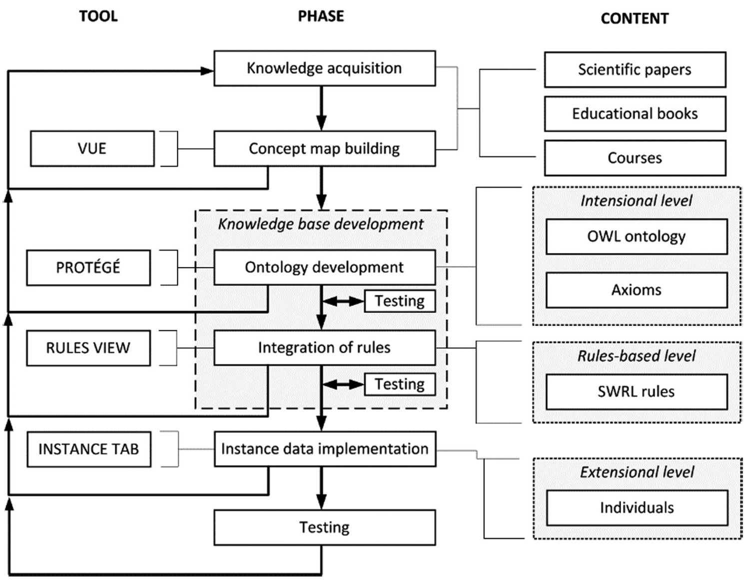


Fig. 1. Systematic building of the SWRL-based recommendation system [7]

3.1 Knowledge Acquisition

Before building of an ontology, it has to be obvious which application domain is in the centre of our interest and which part of a reality should be represented in a knowledge model. Knowledge acquisition is a non-formal phase when particular knowledge sources are identified and used for knowledge extraction and decision which facts are more or less important four our solution.

Non-professional rescuers have a problem to decide about the order of rescue operations. C-ABCDE algorithm offers “a systematic framework” consisting of steps ordered according to the priority with the aim to increase a chance for survival [1]. Each phase tells us what is necessary to do and what is inevitable to do after particular step. As the example, if somebody is not able to react on surrounding stimuli, it is necessary to check the presence of the haemorrhage together with the breath and whether airways are not blocked. The following steps should be applied to all critically ill patients. It is important to note, presence of the consciousness has to be monitored during each phase of the algorithm. Fundamental concepts of this algorithm are modelled with the OWL ontology named the FAO.owl (First Aid Ontology). The algorithm consists of the following phases in the extended version: (1) C - Catastrophic Haemorrhage Control, (2) A - Airway, (3) B - Breathing, (4) C - Circulation (monitoring of a blood circulation that is evaluated on a combination of consciousness, a breath and a hearthbeat), (5) D - Disability (checking the level of consciousness) and (6) E - Environment (global examination of a patient is applied, e.g. allergies, medication, past medical history, last oral intakes or events leading up to an illness or an injury are investigated).

3.2 Non-formal Knowledge Modelling

If we are interested in development of the ontology-based knowledge bases, we often work with complex knowledge. Ontological engineers suggest to organise these complex knowledge into the understandable shape that is easily readable and manageable also by the non-experts. These non-experts often assist in knowledge acquisition for building knowledge bases that integrate correct information and knowledge. The idea is simple and is based on commonly known English language-idiom where a picture is worth a thousand words. Two non-formal approaches for knowledge modelling were considered for the FAO ontology development – mind maps and concept maps.

Mind mapping is creative technique for visual organisation of ideas and information. This technique was proposed by Buzan in the late of 1960s [19]. Mind map is a tree-like structure containing a central subject – a topic which is in the centre of our interest. Next sub-ideas are added to the central topic on the branches of the map. A concept map is a non-formal structure providing visual view on the knowledge of particular application domain [2]. Graph structure of a concept map consists of concepts representing “objects/classes” of a domain and relations between these concepts model statements telling us something about a domain. This technique is efficient if we want to receive global view on the fundamental facts in a domain. The Table 1 briefly compares these two approaches for non-formal knowledge modelling. Finally,

Table 1. Mind maps and concept maps - comparison

Mind maps	Concept maps
Ideas are represented on the lines denoting branches of the map	Concepts are represented in box-like shapes
Relations between ideas do not have a name	Relations between concepts have a name
The most general ideas are closer to the centre of the map	The most general concepts are at the top of the map
Graphical additional illustrations are welcomed	Graphical additional illustrations can be added into the map, but it is not inevitable
Creative construction of the map is desirable	Logical interconnections between concepts are more important than creativity
Tree-like structure	Graph-based structure
Map is mainly used by its author, its sharing is not expected	Map is used by its author, but it is intended for its sharing
Subjectivity is a key aspect of the mind mapping	Objectivity is a key aspect of the concept mapping
They are mainly used for information organisation, remembering and learning (support of pedagogical activities) Mind maps support the brainstorming	They are mainly used for pedagogical purposes (improvement of understanding of particular topics, teaching, learning, remembering). Concept maps support brainstorming

concept mapping is used for the analysis of the application domain (the first aid) because represented knowledge has a graph-based structure instead of a “star-like” structure. Objectivity is also a key aspect in modelling of provision of the first aid. Subjective point of view is mainly applied during mind mapping where it does not matter what public thinks about a mind map. The mind map is mainly built for personal usage of the “mind mapper”.

Four concept maps are built in the VUE (Visual Understanding Environment) tool. Each concept map is focused on the one central topic of the C-ABCDE algorithm, i.e. *Heamorrhage*, *Breathing*, *Blood circulation* and *Consciousness*. Concept maps cover the first five phases of the C-ABCDE algorithm (C-ABCD). One global concept map integrating these four partial maps is not easily readable. This is the reason why the global concept map is visualised only in a human mind. These concept maps are depicted in the figures below. The Fig. 2 represents fundamental facts about disorders which can occur during breathing. The Fig. 3 depicts concept map representing knowledge about a blood circulation. The Fig. 4 visualises facts related to the unconsciousness and the Fig. 5 models knowledge about a haemorrhage. Represented knowledge by the concept map and the FAO ontology is based on the intensive study of literature sources and the first aid course that has been provided by the Red Cross subdivision in the Czech Republic.

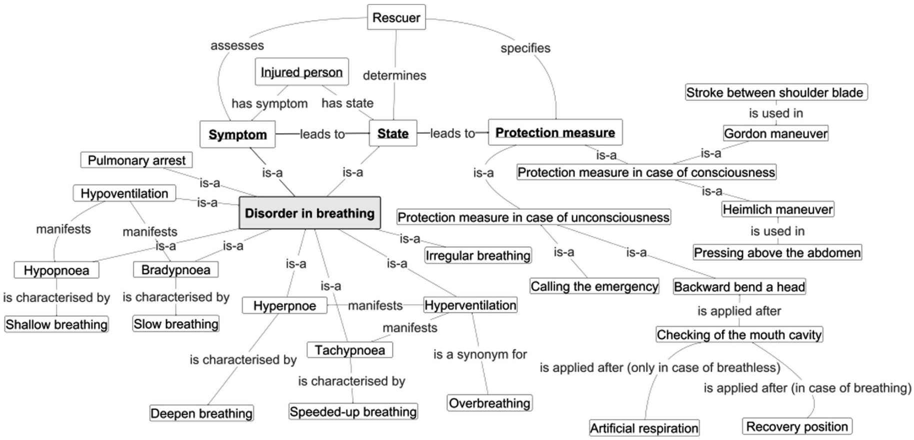


Fig. 2. Concept map representing knowledge about disorders during breathing

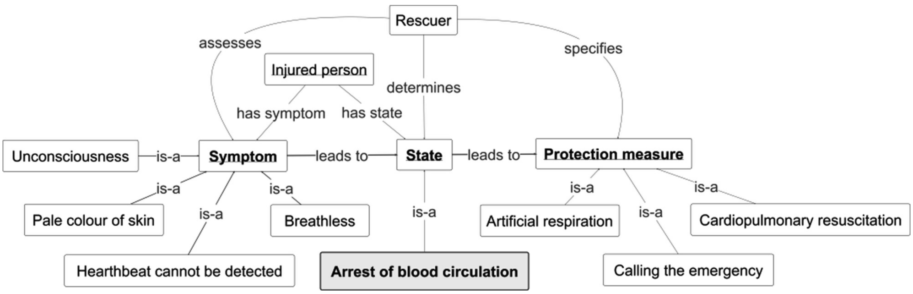


Fig. 3. Concept map representing knowledge about blood circulation

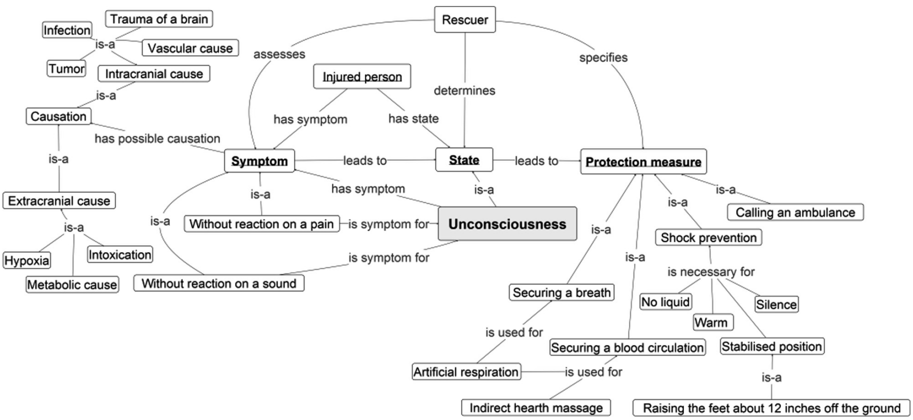


Fig. 4. Concept map representing knowledge about unconsciousness

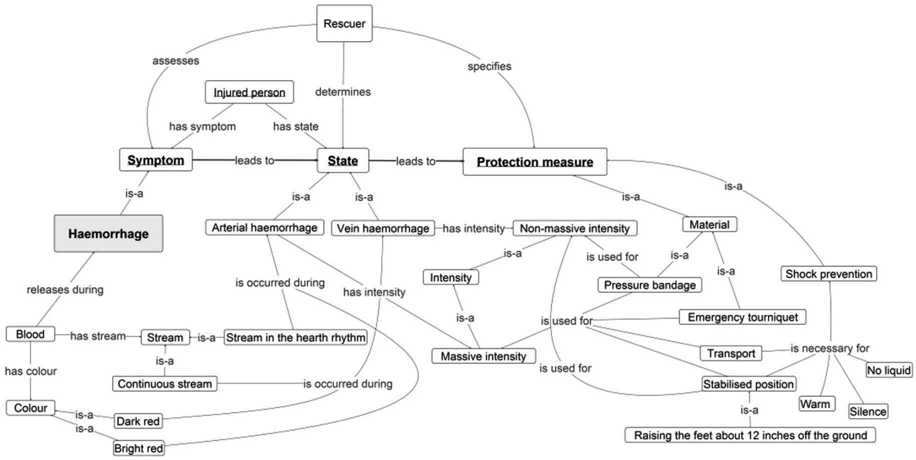


Fig. 5. Concept map representing knowledge about haemorrhage [18]

3.3 OWL Ontology Development

Concept mapping facilitated the OWL ontology development, especially in the identification of the core of the OWL ontology. This core consists of the concepts which are underlined in the figures above (2–5): *Injured person*, *Symptom*, *State* and *Protection measure*. These concepts are used as a basis for the OWL ontology development in a modified form, see the Fig. 6. The state (the OWL class *State*) of the injured person (the OWL class *Emergency case*) is judged according to the symptoms (the OWL class *Symptom*). The state is evaluated and the suitable protection measures (the OWL class *Protection measure*) are used. The core of the OWL ontology is extended with the OWL class representing a cause of symptoms or states (the OWL class *Cause*) and with the fundamental concepts of the C-ABCD algorithm: *Heamorrhage*, *Breathing*, *Blood circulation* and *Consciousness*, see the Fig. 6.

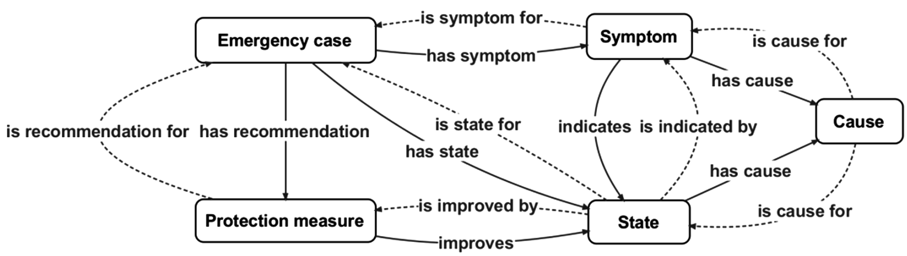


Fig. 6. The core of the OWL ontology

The Protégé environment is the most cited tool used by the ontological engineering community for ontologies development, visualisation, querying, inference, testing and their complete management. This tool is java-based, free, open source and provided for

Windows, MacOS and Linux platform. It has a long history. The initial version was released in 1999 by the Stanford Center for Biomedical Informatics Research. In time of writing of this text, five generations of the Protégé occurred. Almost each version of the Protégé brings something new thanks to the plugin-based architecture and this is the reason why these versions were investigated and compared for decision which of them should be used for the OWL ontology development. Comparison of particular Protégé versions is depicted in the Table 2. Versions were tested with the following computer configuration: CPU Intel Core i5-4460, frequency 3.20 GHz, operational memory 16 GB RAM, 64-bit system, operating system Win 10 Education, JAVA 8 (update 121). It is found out that the Protégé 4.3.0 (build 304) is the most stable version for the used computer configuration. The Protégé 5.0.0 and 5.2.0 causes the exception because of the missing additional libraries for the FACT++ reasoner. FACT++ is the open-source OWL-DL (OWL Description Logics) reasoner implemented in C++. It supports reasoning with the OWL 1 and the OWL 2 language with some restrictions. It became default reasoner for the Protégé 4 and the Protégé 5. The problem with missing additional libraries is usual and frequently mentioned in the Protégé community. Removing of this problem is not so trivial and this is the reason why the most actual version of the Protégé is not used also with respect to the user friendly modelling. The OWL ontology is modelled in the OWL 2 version that is used by the Protégé 4 and by the 5th generation. It extends the OWL 1 in the view of the new "semantics constructs", see [6]. The OWL ontology is normalised because of the easily maintainable structure. The FAO ontology is visualised with the OWLViz plugin, see the Fig. 7.

Table 2. Possibilities for modelling ontologies in the Protégé environment – comparison

Parameter	Protégé version							
	Protégé 3.3.1	Protégé 3.5.0	Protégé 4.0.0	Protégé 4.1.0	Protégé 4.2.0	Protégé 4.3.0	Protégé 5.0.0	Protégé 5.2.0
Frames	Yes	Yes	No	No	No	No	No	No
RDF support	RDF/XML RDF	RDF/XML RDF	RDF/XML Turtle	RDF/XML Turtle	RDF/XML Turtle	RDF/XML Turtle	RDF/XML Turtle	RDF/XML Turtle
OWL support	OWL 1	OWL 1	OWL 2	OWL 2	OWL 2	OWL 2	OWL 2	OWL 2
Available reasoner	Manual installation (Pellet)	Pellet	Fact++	Fact++ Hermit ELK	Fact++ Hermit Pellet	Fact++ Hermit Pellet	Fact++ Pellet Jcel	Fact++ Hermit Pellet ELK
Visualisation	Jambalaya OWLViz OntoViz	Jambalaya OWLViz OntoViz	OWLViz OWLPropViz	OWLViz Ontograph	Ontograph OWLViz	Ontograph OWLViz	Ontograph OWLViz	Ontograph OWLViz
Rules support	SWRLTab (Jess)	SWRLTab (Drools)	Rules tab (SWRL)	SWRL rules	SWRL rules	SWRL rules	SWRLTab SWRL rules	SWRLTab (Drools)
Querying	Queries tab	Queries tab String search	DL Query	DL Query	SPARQL DLQuery	SPARQL DLQuery	DL Query SPARQL SNAP S. SQWRLTab Exist. Query	DL Query SPARQL SNAP S. SQWRLTab Exist. Query
Debugger	Yes	Yes	No	No	No	No	Yes	Yes

Fig. 7. Visualisation of the OWL ontology named FAO in the OWLViz (Protégé 4.3.0)

Ontologies can be used for knowledge bases development consisting of classes, properties, axioms and additional constructs specifying the semantic of the implemented statements. Instances of the OWL classes named individuals play the role of the real data that are part of these classes. OWL individuals represent particular symptoms, states, causes and protection measures in the FAO ontology modelling the first aid domain. The statistics of the ontology is the following:

- number of OWL classes: 32,
- number of OWL object properties: 12 (including inverse object properties),
- number of OWL individuals: 41.

3.4 Integration of SWRL Rules into OWL Ontology

The FAO normalised ontology contains information and knowledge about situations which can occur if somebody needs the first aid, e.g. a massive haemorrhage (arterial or vein), the unconsciousness, complications during breathing or problems with a hearth rhythm. It is important to mention that presented ontology models static aspects of the investigated system. It does not model dynamic characteristics, e.g. behaviour or transitions between states of a system. It reflects a part of a real world - existing entities and relations between them together with their semantics. The purpose of developed application is to propose particular recommendation with respect to an actual state of an injured person. The FAO ontology is partially able to provide recommendations, i.e. it is possible to infer that a particular subclass can have particular parental class(es) or an individual can be a part of a class. This type of inference is not needed for the intended application, because specific recommendation (an individual) has to be provided for concrete emergency case (an individual). It means that the relations between individuals have to be inferred. SWRL (Semantic Web Rule Language) is a rule-based language able to extend the OWL ontology with semantic rules. These rules model situations in the shape of condition-action structure. This structure is based on OWL classes and OWL properties which are combined together with variables. Relations between OWL individuals are represented due to this condition-action structure where they are included in variables. SWRL rules of the two types are proposed for the FAO ontology:

- SWRL rules facilitating to decide about a state of an emergency incident. The following OWL object properties are mainly used in this category of rules: *hasSymptom*, *hasState*.
- SWRL rules facilitating to select suitable protection measures. The following OWL object properties are mainly used in this category of rules: *hasState*, *hasRecommendation*.

SWRL rules represent decision making of a rescuer. SWRL rules are designed with respect to the most fundamental concepts of the C-ABCD algorithm, i.e. haemorrhage, breathing, blood circulation and consciousness. The OWL class named EmergencyCase represents injured person requiring the first aid, and its symptoms indicating particular problem which has to be solved by a rescuer. These symptoms are represented as OWL individuals that belong to particular OWL classes. These actual symptoms play a role of input data for the inference process. Inference is carried out a reasoner. This software uses pieces of knowledge represented by the OWL ontology and SWRL rules for decision making and provision of the recommendations, i.e. a reasoner judges a state of an injured person according to its symptoms and provides suitable protection measures which should be used for the first aid. The Protégé 4.3.0 integrates a plugin for representation of SWRL rules – SWRL editor (view Rules). Recommendations are provided by the HermiT open source reasoner supporting the OWL 1 and the OWL 2. It implements "hypertableau" calculus, similarly as the FACT ++. It provides efficient and fast reasoning in comparison to previously-known algorithms, for more details please see [20]. This reasoner is a part of the Protégé 4.3.0, see Table 2.

The following examples demonstrate which SWRL rules can be applied for decision making in case of the first aid. These SWRL rules reflect typical emergence situations that can occur in a reality. The first example represents how a type of haemorrhage can be evaluated. A vein haemorrhage can be characterised by dark red colour and continuous stream of a blood, in comparison to an arterial haemorrhage which is indicated by a bright colour of a blood and pulsation of a blood.

```
hasSymptom(?ec, ContinuousStreamOfBlood),
hasSymptom(?ec, DarkRedColourOfBlood)
->
hasState(?ec, VeinHaemorrhage)
```

Recommendation is based on an actual state of a person (an emergency case). If the vein haemorrhage is present, it is necessary to evaluate the intensity of bleeding, see the following SWRL rule providing a recommendation if bleeding is non-intensive.

```
hasState(?ec, VeinHaemorrhage),
hasSymptom(?ec, NonMassiveIntensityOfHaemorrhage)
->
hasRecommendation(?ec, PressureBandage),
hasRecommendation(?ec, RecoveryPosition)
```

The above mentioned example demonstrates that a pressure bandage and recovery position is necessary. In case of the intensive (massive) haemorrhage, particular recommendations are different in comparison to the previous SWRL rule, i.e. calling the emergency and transportation is suitable. An emergency tourniquet or a pressure bandage can be used for suppressing an intensity of bleeding. It depends what kind of equipment a rescuer currently has.

```
StateOfHaemorrhage(?state),
hasState(?ec, ?state),
hasSymptom(?ec, MassiveIntensityOfHaemorrhage)
->
hasRecommendation(?ec, Emergency),
hasRecommendation(?ec, EmergencyTourniquet),
hasRecommendation(?ec, PressureBandage),
hasRecommendation(?ec, Transport)
```

3.5 Testing of SWRL-Based Prototype

The Protégé 4.3.0 is not only used for ontology and rules modelling. It is applied as an experimentation tool for verification whether and how the semantic web technologies can be used for building of the application providing advices in case of the first aid. The FAO ontology actually integrates 19 SWRL rules which are applied for the inference process by the HermiT reasoner. The reasoner also checks a consistency of the OWL ontology and detects possible missing elements used for semantics expression, e.g. disjointness between classes.

The following scenarios demonstrate inference process realised by the HermiT reasoner. The first scenario solves the emergency case (the Case02-UnresponsiveBreathless, see the Fig. 8) where the following symptoms are present: disability to provide a response and breathless. There are only two symptoms, but they are serious. The Fig. 8 depicts the output of the inference process realised by the HermiT.

These two symptoms lead to an obvious decision about a state of a person. A person has a problem with breathing and is in unconsciousness state. These two conclusions are used as the inputs for an initiation of the next SWRL rule that provides

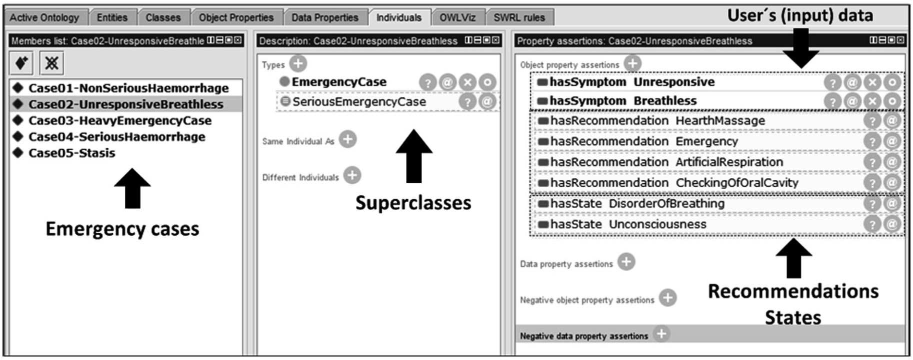


Fig. 8. Test scenario 01: no response and breathless

the final recommendations: calling the emergency because of the serious situation, checking the oral cavity whether particular obstacle is not present inside. If the air passages are free, combination of the hearth massage and the artificial respiration should be applied.

The second scenario solves the emergency case (the Case04-SeriousHaemorrhage, see the Fig. 9) where more symptoms are present: a person is able to respond to stimuli, breathing is normal, but the patient bleeds massively in pulses where a colour of a blood is bright. The HermiT correctly identifies that a person is fully conscious, breathing is normal, but the arterial haemorrhage occurs. According to this deduction, the HermiT proposes the following recommendations: calling the emergency and transport, usage of the pressure bandage or the emergency tourniquet should be used.

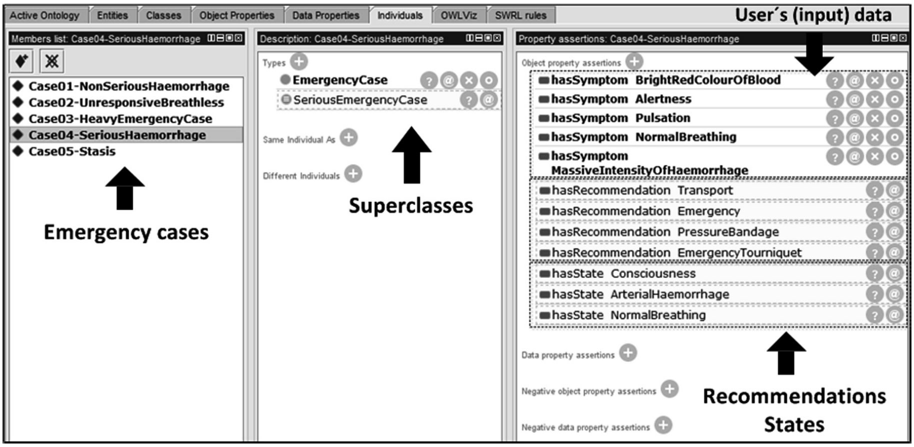


Fig. 9. Test scenario 02: massive haemorrhage

The third scenario describes the emergency case (the Case05-Stasis, see the Fig. 10) where four symptoms are localised: a person has irregular breath and

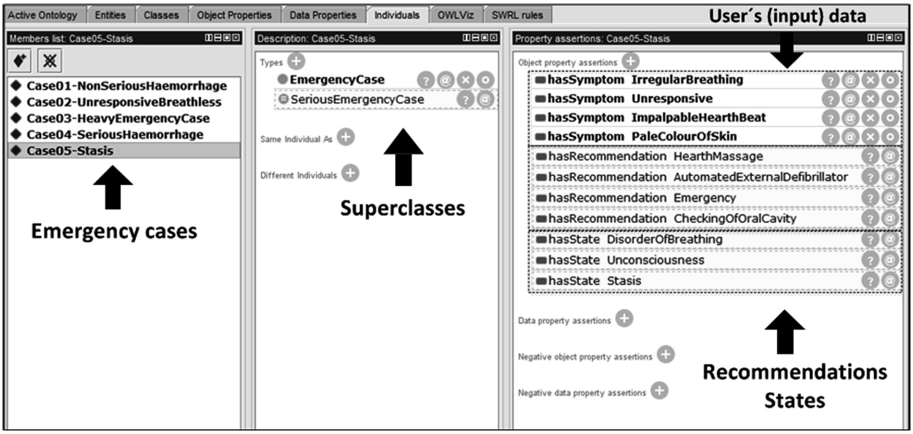


Fig. 10. Test scenario 03: stasis

impalpable hearth beat, colour of a skin is pale and is not able to respond on stimuli. The HermiT reasoner deduces that a patient has a problem with breathing. The patient is unconscious and the stasis state is detected. According to these conclusions, the HermiT recommends to call the emergency, to check the oral cavity, to use the automated external defibrillator together with the hearth massage.

4 Discussion and Future Directions

The Sect. 2 of the text introduces various applications of ontologies in the emergency management. It is find out that usage of the semantic web technologies for development of the systems for the first aid is not common practice. Researchers attempt to find approaches for representation and sharing knowledge about emergency, crisis situations and how to solve them in different domains, including ontological engineering. The research studies [15–17] are the closest to the presented research in this text. These studies are briefly compared with the FAO ontology in the Table 3.

Precise comparison of the ontologies introduced in the above mentioned research studies is difficult because of their inaccessibility, but the shared results of these authors help to improve the FAO ontology. The research presented in [16] clearly mentions the ontological structure in the view of the concepts related with the first aid. The FAO ontology models more OWL classes related with the first aid, e.g. how to deal with: unconsciousness, disorders during blood circulation and irregular breathing. The research paper [16] is mainly focused on the bleeding and burning, but the authors mention that additional extensions are going to be done in the future. The FAO ontology does not contain any information about particular causes of injuries, personal details of an injured person or a location where the problem occurred. Personal details, roles, locations and competencies are included in the ontologies introduced in [15, 17]. The FAO ontology is going to be extended, mainly with the causes of disorders and personal details, especially an age of a person. The age is very important characteristic

Table 3. Ontologies for provision of the first aid (the papers are chronologically ordered)

Article	Characteristics of ontologies		
	Ontological format (tool)	Content of the ontological hierarchy	Support of rules for the inference (reasoning)
[16]	OWL (Protégé)	Bleeding, burning, intensity, symptoms, advices, treatments, body parts	No
	Non-formal model of the domain/ontology: non-formal model of the application domain or ontology is not proposed		
	Results: OWL ontology is used for education of the first aid and questionnaires preparation		
[17]	Frame-based ontology (Protégé-Frames 3.5.0)	Event, object, task, role, agent, action, competency	Reasoning rules are implemented (Theoretical explanation of their usage is provided)
	Non-formal model of the domain/ontology: non-formal model of the ontology is depicted (own visualisation)		
	Results: ontology is used for automated scenarios generation supporting teaching and education of the first aid		
[15]	OWL (not mentioned)	7 classes, 15 data properties, 5 object properties (e.g. Person, Patient, Volunteer, Location, Profile, Alarm, AuthorisationRequest)	No
	Non-formal model of the domain: non-formal model of the ontology is depicted (own visualisation)		
	Results: proposed ontology unifies the smart space information about participants in case of the first aid		
The FAO ontology	OWL (Protégé 4.5.0)	Consciousness, haemorrhage, blood circulation, breathing, intensities, recommendations, symptoms	Yes, SWRL rules are applied
	Non-formal model of the domain: concept mapping is used for non-formal modelling of the application domain		
	Results: the prototype of the ontology-based recommendation system for the first aid is proposed		

for provision of the first aid. As the example, breathless is solved differently in case of children and adults. Valuable advices were also received thanks to the consultations with the expert. Actual knowledge base contains small collection of situations which can occur during emergency incident. These situations are represented by 19 SWRL rules, but our lives can be under attack of more dangerous situations.

Main purpose of the Protégé environment is to support knowledge modelling with the ontologies and to support knowledge mapping, visualisation and inference. This tool cannot be used as a final software solution for the needs of the end users. The Protégé facilitates the ontology development, especially because its source code is hidden for a knowledge modeller. This tool is not only used for building of the FAO ontology. It is also used for ontology validation, visualisation and as the experimental tool. It is possible to “simulate” various scenarios by running SWRL rules and to explore the impacts on the ontological structure. This tool is used for exploration whether and how it is possible to use particular technologies (approaches) of the semantic web initiative for modelling of a complex domain – the first aid. Development of sophisticated recommendation system for the first aid cannot be done with the Protégé itself. As the example, if the particular SWRL rules are initiated, the HermiT reasoner provides recommendations which are not ordered, e.g. according to their priority. The order of steps is really important if we provide the first aid. This could be solved by particular ontology-based application that can be built with a specialised framework. Various semantic web-based frameworks exist that provide methods for ontologies building, ontologies management and usage of ontologies as knowledge bases for knowledge sharing, e.g. OWL API, Jena, Sesame, OWLReady or CubicWeb. Future directions are going to be focused on the FAO ontology extension together with the SWRL rules and their usage by the particular semantic web-based framework that is able to operate with the OWL ontologies. In this point of view, OWL API and Jena are mainly taken into account.

5 Conclusion

The text introduces the initial research of the application of the semantic web technologies for building of the recommendation system used for provision of the first aid. Systematic approach for modelling of complex knowledge is inevitable. This is the reason why the conceptual framework is applied for representation of the first aid domain. Concept mapping is used in the initial phase of the knowledge mapping. Concept maps help with knowledge organisation, visualisation and for building of the knowledge base. The knowledge base is built with the OWL ontology and SWRL rules. These alternatives are selected because of their ability to model complex knowledge, represent context and provide intelligent reasoning and inference with existing knowledge structures. Various versions of the Protégé environment are compared and the version 4.3.0 is used for OWL ontology modelling. The FAO ontology is designed and compared with ontologies which are developed for similar purpose, i.e. how to provide the first aid. The FAO ontology is going to be extended with more concepts, especially causes of injuries and with SWRL rules covering more real situations. The OWL ontology-based prototype using SWRL rules is proposed, but it is inevitable to build more sophisticated solution that could be used in practice. The OWL API and Jena framework is going to be deeply investigated for building of such application. Necessity to build this system is evident. Everybody should be able to know what to do if someone encounters with sudden health incident.

Acknowledgements. Czech Science Foundation project 18-01246S is kindly acknowledged.

References

1. Thim, T., et al.: Initial assessment and treatment with the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach. Int. J. Gen. Med. **5**, 117–121 (2012)
2. Novak, J.: Concept maps and Vee diagrams: two metacognitive tools to facilitate meaningful learning. Instr. Sci. **19**, 29–52 (1990)
3. Borst, W.N.: Construction of engineering ontologies for knowledge sharing and reuse. Ph.D. thesis, Universiteit Twente (1997). http://doc.utwente.nl/17864/
4. Javed, Y., Norris, T., Johnston, D.: Ontology-based inference to enhance team situation awareness in emergency management. In: Proceedings of the 8th International ISCRAM Conference, Lisbon, Portugal, pp. 1–9 (2011)
5. Han, Y., Xu, W.: An ontology-oriented decision support system for emergency management based on information fusion. In: Proceedings of the 8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, Bellevue, USA, p. 8 (2015)
6. W3C OWL Working Group: OWL2 Web Ontology Language Document Overview (Second Edition), W3C Recommendation (2012). https://www.w3.org/TR/owl2-overview/
7. Husáková, M.: Systematic development of ontology-based decision support system for solving emergency incidents. In: Proceedings of the 27th International Business Information Management Association Conference (IBIMA), Milan, Italy, pp. 152–156 (2016)
8. Gruber, T.R.: A translation approach to portable ontologies. Knowl. Acquis. **5**(2), 199–220 (1993)
9. Zashan, F., Mohamad, R.: Medical ontology in the dynamic healthcare environment. Procedia Comput. Sci. **10**, 340–348 (2012)
10. Fernández, M., Gómez-Pérez, A., Juristo, N.: METHONTOLOGY: from ontological art towards ontological engineering. AAAI Technical report SS-97-06 (1997)
11. Xu, W., Zlatanova, S.: Ontologies for disaster management response. In: Li, J., Zlatanova, S., Fabbri, A.G. (eds.) Geomatics Solutions for Disaster Management. Lecture Notes in Geoinformation and Cartography, pp. 185–200. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72108-6_13
12. Carletti, G., Giuliodori, P., Di Pietri, V., Peretti, A., Amenta, F.: An ontology-based consultation system to support medical care on board seagoing vessels. Int. Marit. Health **67**(1), 14–20 (2016). https://doi.org/10.5603/IMH.2016.0004
13. Luo, H., Peng, X., Zhong, B.: Application of ontology in emergency plan management of metro operation. Procedia Eng. **164**, 158–165 (2016). https://doi.org/10.1016/j.proeng.2016.11.605. Creative Construction Conference (CCC 2016)
14. Pardo, E., Espes, D., Le-Parc, P.: A framework for anomaly diagnosis in smart homes based on ontology. Procedia Comput. Sci. **83**, 545–552 (2016). https://doi.org/10.1016/j.procs.2016.04.255
15. Lebedev, N., Timofeev, I., Zavialova, I.: Design and implementation of the first aid assistance service based on Smart-M3 platform. In: The 18th of the Conference of Open Innovations Association and Seminar on Information Security and Protection of Information Technology, St. Petersburg, pp. 174–180 (2016). https://doi.org/10.1109/fruct-ispit.2016.7561524
16. Ahmed, S., Parsons, D., Ryu, H.: Supporting adaptive learning interactions with ontologies. In: Conference Proceedings of the 11th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction, Auckland, New Zealand, pp. 17–24 (2010). https://doi.org/10.1145/1832838.1832842
17. Peeters, M.M.M., et al.: An ontology for automated scenario-based training. Int. J. Technol. Enhanced Learn. **6**(3), 195–211 (2014)

18. Husáková, M.: Ontology-based education support system for solving emergency incidents. In: Nguyen, N.-T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds.) ICCCI 2016. LNCS (LNAI), vol. 9876, pp. 313–322. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45246-3_30
19. Buzan, T., Buzan, B.: The Mind Map Book: How to Use Radiant Thinking to Maximize Your Brain's Untapped Potential. Plume, New York (1996). Reprint edition 1st March 1996. ISBN 978-0452273221
20. Shearer, R., Motik, B., Horrocks, I.: HermiT: a highly – efficient OWL reasoner. In: Ruttenberg, A., Sattler, U., Dolbear, C. (eds.) Proceedings of the 5th International Workshop on OWL: Experiences and Directions (OWLED 2008), Karlsruhe, Germany, 26–27 October 2008, vol. 432 (2008)

Queue Lengths Management for Deterministic Queuing Systems

Martin Gavalec and Zuzana Němcová(✉)

Faculty of Informatics and Management, University of Hradec Králové,
Rokitanského 62, 50003 Hradec Králové, Czech Republic
{martin.gavalec,zuzana.nemcova}@uhk.cz

Abstract. The paper discusses two proposed methods for the cost optimization of the deterministic queuing systems based on the control of the queue lengths. The first method uses the evaluation of actual states at the particular service places according to their development. The decision is then based on the comparison of the criteria of productivity and the expended costs. The suggested change in the system setting with the highest priority is then accomplished. The second method is based on the simulation of the future states and on this basis the appropriate time and type of the modification of the system setup is suggested.

Keywords: Optimization · Deterministic queuing system
Tandem network

1 Introduction

This paper is an extended version of [7], a paper entitled "Cost optimizing methods for deterministic queuing systems" published in the Proceedings of the International Conference on Computational Collective Intelligence (ICCCI 2016). The main extension lies in the detailed description of different system types that can be considered for optimization. The formulas for computation of basic characteristics of these systems are given. Also, the examples are interpreted in more detail.

Queuing network modeling have been investigated by many researchers. The general overview and bibliography material concerned with modeling the production and transfer lines using queuing networks can be found in [10]. The paper provides a bibliography material concerned with modeling such systems. Stochastic approach in queuing theory have been introduced in [3,5,11]. In [2] is provided the deep insight into the flow problems in networking using deterministic approach. In [6] the authors use the objective function of the model as a fitness function operator. The objective function is composed of the expended manufacturing costs (such as machine purchasing costs, repair service costs) rather than the costs connected with queue development. In [8] the objective of the optimization is to maximize the probability of fulfilling orders within the quoted fulfilment time. The model assumes fixed number of servers.

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXIX, LNCS 10840, pp. 103–122, 2018.
https://doi.org/10.1007/978-3-319-90287-6_6

This paper offers different perspective on the cost optimization in sense of optimizing the queue lengths through the change in the system setting. It is designed for systems that allows either to flexibly operate with the number of machines (or workers) or to flexibly change the service intensity at particular service places.

Queuing systems can be found in many real systems and many of us come into contact with them in everyday life. Moreover, in today's information age society, in context of the interconnected communication and data sharing, the importance of the understanding and prediction of the system behavior increases. We can find the applications, for example, in the packet network design and optimization, the design of the traffic lights at crossroads and in the manufacturing process of the production systems. The first application of the queuing theory can be traced back to the beginning of the 20th century, when the telephone networks has been designed and analyzed.

The simulation of a system in discrete time means modeling the system as an interlaced network consisting of dynamic and static objects. The objects picture the individual entities described by specific characteristics, which influence the passage of the entity through the system. Despite the fact that the simulated time is continuous, the changes of the system state are implemented at discrete moments of time. These systems have in most cases a mass character. Such systems can be seen in real life, for example, a delivery of ordered products, manufacturing products, providing a customer service at the post counter, or a communication network. Naturally, queues occur in these systems; the knowledge of the development of these queues is important for the process and/or also for possible cost optimization. Waiting costs certainly represent a significant part of the total costs. These can be either some kind of penalization for staying in the queue that is inadmissible due to the applied technology, or can be represented by the dissatisfaction of customers standing in a long queue. Therefore the control of the lengths of the queues can lead to considerable savings and thus to an increase in the efficiency of the system.

Naturally, it is supposed that it is beyond the human strength to model or simulate the system perfectly. From some level of complexity of the system it is impracticable to use the analytical methods, and thus it is necessary to be able to describe or investigate the system with certain simplification.

The state transitions of a *discrete event system* (abbreviated as DES) are initiated by events (a DES can also be called an event-driven system although this depends on whether the state transitions are synchronized by a clock or occur asynchronously) and these events occur at discrete moments of time. In other words, with each transition of the system, some event can be associated. The event can represent the start/end of some activity, for example, a completion of the product, a customer's arrival, or a machine breakdown. Between these events, time lags of different length can be observed (the lags can be of a deterministic or a stochastic character). The basic characteristics of DES can be found in [4,13].

On the other hand, the so-called *discrete time systems* can be distinguished, where the states generally change with the passage of time (because the passage of time is the cause of the change, the system is also called “time driven”). In other words, with every “clock tick” the state is expected to change. To complete the idea, continuous state systems are considered as time-driven by their very nature. But a discrete state system can be of either type, depending on whether the state transitions are synchronized by a clock or occur asynchronously.

In general, there are two types of units in such systems, the requests and the service places. The basic idea of this concept is that the requests come into the system in order to be served at particular service places. The intensity of the requests arrival to the system can naturally vary in time. The service times of the servers at the service places can also differ. In dependence on the ratio of these quantities, the queues can appear and on the other hand, some of the servers in the system can become idle. Sometimes, the occurrence of the queues and/or the idle servers is undesirable because it can pose additional indirect costs. This situation can result in searching for better possible system setting.

The arrangement of the service places is important as well; the service places can be concatenated in both, the series and/or the parallel manner. Queuing systems can have various structures, from those with the simplest layout (for example the cash desk at the gasoline station) to those with the complex organization (for example assembly lines).

2 Motivation

The motivation for studying DES are problems that can arise during the operation of such systems. Lets look more closely at a production system, an assembly line manufacturing some product. The assembly line can consist of several concatenated processing units, the servers. These servers serve the incoming requests and machine the product. The product has to pass through a sequence of servers. The passage of the product through the system depends on the connections of the servers. They can be interconnected in an arbitrary combination of parallel and/or serial connections. Each server offers its service during a given service time, and then the request is transported to the next server. The server can usually start working on a new product if it has finished processing the previous one. The server can serve at most one product in one time instant. It is also assumed that each server starts working as soon as all the needed parts are available. In other words, the server has to wait until the service of the request by the preceding server has ended. Two aspects of such systems can be seen here. Since the service times differ, queues are formed in such a system. The service intensity of a faster server is higher than the service intensity of a slower server and it is clear that the queues in front of a slower server can increase, whereas a faster server can easily become idle. Both states are undesirable, especially from the economic point of view, because they can mean indirect costs. The second aspect is that after the system starts, a transient time period can be seen: after this transient time period, the system can become steady. The parameters of the system can become constant or some differences can appear, but in predictable periods. Whether the system gets into a steady state or not depends on the specific setting of the parameters of the system (with some settings, it can happen that the stability of the system is never achieved).

Particular characteristics describing the behavior of the system can be computed during the system analysis in dependence on the user's intentions. The characteristics can be divided into several categories [1]:

- time characteristics of the requests,
- characteristics related to the number of the requests,
- probabilistic characteristics,
- cost characteristics.

Assessing of these quantities is important especially by the system design or modernization. It helps to project the corresponding extent of the system so that its performance is optimal.

Discrete event systems can be studied by several methods. Computer simulation is quite widely used to study such systems. One of the disadvantages of simulation is that to create an appropriate model with the corresponding details and functions that operate properly is a demanding task (but of course, not unachievable). However, it is without doubt that the model is in accordance with reality (with a high degree of similarity).

Another way to understand or describe the system is to make a mathematical model of it. Such models also allow the prediction of the system behavior, but in general, the mathematical analysis is more exact and allows an easy derivation of some properties of the system and the deduction of general conclusions. But it depends on the type of the system and also on the purpose of the study. To use a mixture of analytic and simulation methods thus seems to be advantageous.

In further sections the deterministic concept of the system is considered. It can be understood as a special case of stochastic approach. Using this approach it is possible to compute the parameters of the system, and thus it is also possible to predict the system future development.

3 Deterministic Open Linear Queuing System

Consider an open linear queuing system (in literature also known as the tandem queuing network, see [9]) with $n + 1$ service places, see Fig. 1. The incoming requests have to pass through the series of all servers and then leave the system.

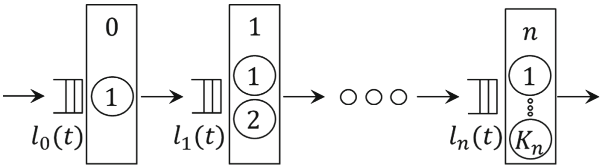


Fig. 1. Open linear system

There can be arbitrarily long queues in front of each server before the system starts. The system works in the so-called stages; the stage is the time period during which the system setting stays unchanged. The time variable during stage takes values $t = 1, 2, ..., T$, where T is the length of the stage. Manager of the system has following information about the system setting:

- z_i, the basic service time of the server(s) at service place i,
- K_i, the number of identical servers at the service place i,
- l_i, the length of the queue at the service place i.

The service intensity at particular service place is then computed as $\sigma_i = K_i/z_i$. The intensity of arrivals is represented by the service intensity of the entrance server (indexed by $i = 0$). This server can be considered as a gatekeeper, which controls the access of the requests. It is assumed, that the queue in front of the entrance server is long enough; the gatekeeper admits the requests to the system with a constant speed.

Service times of the servers can be changed. The cause of the change can be either internal or external. The internal reason means that the manager of the system decide to change some of the service times, quantities, that can be directly influenced (the change in the value z_i, where $i > 0$). The external reason represents the change in the arrivals intensity of the requests coming into the system (the change in z_0, respectively K_0, in the model). It is important to note that the possibility of unexpected change in the intensity of arrivals as well as the need to cope with the different types of queues can foil the effort to set the service intensities to the intensity of arrivals at the start of the system.

4 Types of the Queues

Within the stage the queues can behave differently depending on the sequence of service intensities at particular service places. Some of them remain stable, some of them increase or decrease. The change in queue length is dependent not only on the service intensity of previous service place but also on the situation at all preceding service places.

The time variable plays its role as well. In the so-called transient time, it can happen that the queue is temporarily increasing or decreasing. After the transient time, in which the queues can possibly change the sign of their growth, the steady state is considered.

Then the queues do not change their character, i.e., each queue either remains stable, or constantly increases or decreases (except in the situation where the queue becomes empty).

In the steady state, three types of queues are distinguished:

- constant queue,
- decreasing queue,
- increasing queue.

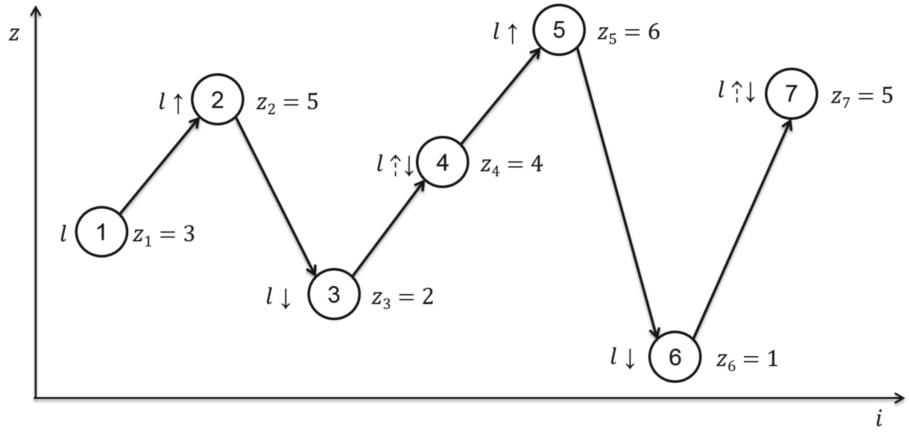


Fig. 2. Queues in transient period and in long term

An empty queue emerges as a result of a decreasing queue, or it is a special case of the constant queue.

For example see Fig. 2, the queue in front of the ith service place can temporarily increase in transient time (depicted in the Figure by the dashed arrow) if the service place $i-1$ is faster but some of the preceding service places is even slower than the ith service place, i.e. $z_{i-1} < z_i < z_{i-k}$. In this case, the queue can increase for some time period (depending on the actual lengths of the queues), because the preceding service place sends the requests in higher intensity. In the long term (in Fig. 2 long term tendencies are presented by continuous arrows) the queue will decrease, because the intensity of arrivals to the ith service place is then considered as the service intensity of the slowest preceding service place.

5 Production Costs

The aim of the optimization is to adapt service times to the arrival intensities and to control the lengths of the queues as well as the number of empty queues to minimize the total cost of the system.

The optimization is made in each stage separately - with respect to actual system setting and the function of total production costs of the stage. The function comprises the components that create important factors for the system performance.

The manager of the system can then significantly affect the overall production costs of the system by making the decision to change various service intensities.

To make the model as simple as possible, we consider as the components of the total production costs only such factors, that can be determined directly during the optimization process. In other words, for simplicity, the costs for material, maintenance, fixed costs and other are omitted in this model.

The production cost P consists of the work costs W, the costs of inaction on the part of some of the servers I, the costs resulting from long queues Q, and the costs of changing the number of servers in some of the service places C, i.e.,

$$P = W + I + Q + C \tag{1}$$

Let us denote the constituent costs: work, idle, queuing and change by W_i, I_i, Q_i, C_i. Each queue at the ith service place can have different maximal limit denoted by M_i; it means that the queue which is shorter or equal to M_i is still tolerable.

Work costs are proportionally dependent on the actual number of servers at the ith service place. Idle costs are spent in case when the service capacity at the service place i is not fully used. It is proportionally dependent on the degree of such idleness. Queuing costs result from the excessively long queues, i.e. that are over the maximal tolerable limit M_i. These costs are also proportionally dependent on the degree of exceeding this limit. The last, change costs, are one-shot costs (added at the end of the stage) that are spend for the factual change in the system setting (at the service place j).

The possible reactions of the manager of the system depend on the ratios of the constituent costs. If the queuing costs are several times higher than the idle costs (which means that staying in a queue that is considered to be long is undesirable, for example, because the surface of the varnished product will become dry and that makes further processing impossible: the products are spoiled), then there will be a tendency to maintain the lengths of the queues under the threshold level rather than take some action towards eliminating servers that are idle.

Denote by $l_i(t)$ the length of the queue in front of service place i in time t. The total production costs of the stage are then computed according to the following formula:

$$P(T,j) = \sum_{t=1}^{T} \left(\sum_{i=1}^{n} W_i K_i + \sum_{\substack{i=1 \\ l_i(t) < \frac{K_i}{z_i}}}^{n} I_i \left(\frac{K_i}{z_i} - l_i(t) \right) + \sum_{\substack{i=1 \\ l_i(t) > M_i}}^{n} Q_i \left(l_i(t) - M_i \right) \right) + C_j \tag{2}$$

6 System Types

Every real system can be considered as unique with its specific behavior. Main difference between systems is in the way they use to process the requests and possibilities how can system manager affect the system setting (if possible).

Both the service times and the request flow can be of two types: continuous or discrete. Continuous service time means that the manager of the system can change the setting of the system by accelerating or decelerating the service times of the servers so that the condition $z_i > 0$ is fulfilled. Considering the discrete service times approach, it is possible to add or take away another identical server to some service place in order that $K_i \geq 1$. A continuous request flow means that there is no need to wait until the whole request has been served: the served part of the request can leave the service place and fall in another queue. If the request flow is discrete, the request can fall into the next queue as soon as it is completed.

Thus four types of system can be distinguished:

1. a system with continuous service time and continuous request flow,
2. a system with continuous service time and discrete request flow,
3. a system with discrete service time and continuous request flow,
4. a system with discrete service time and discrete request flow.

Method for queue lengths computation can vary for different system type. The computation of basic characteristics as well as more detailed description follows.

6.1 System with Continuous Service Times and Continuous Request Flow

An example of such a system is processing the letters at a post office, where the requests are represented by packages of letters and at every time unit someone carries away the already processed letters from each worker. In case of need, workers can slow down or speed up the pace of work.

For this type of system it is supposed that there is one server at each specific service place, $K_i = 1$. Then $\sigma_i = 1/z_i$ in this type of system. The given service time z_i can be changed within certain limits. In other words, the server can be accelerated or decelerated. Only a deceleration where $z_i > 0$ can be taken into account: the server is still working (it is not possible to put the server out of service).

Because the served part of the request can fall into the next queue, the quantity of the already served part of the request can be computed.

Henceforth, the following notation will be used: the served quantity of request in service place i at time t is denoted by $f_i(t)$, and the number of requests present at the ith service place is denoted by m_i, $m_i \geq 0$. The remaining part of the request that is being served at service place i is denoted by s_i, $0 \leq s_i \leq 1$. The notation s_i will be used in Theorem 2 (for the systems with continuous service time with one server at each service place).

Theorem 1. *Consider a queuing system with continuous service times and continuous request flow. Then the parameters $f_i(t)$, $m_i(t)$ and $l_i(t)$ are computed as follows.*

$$f_i(t) = \min(m_i(t-1), \sigma_i) \tag{3}$$

$$m_i(t) = m_i(t-1) - f_i(t) + f_{i-1}(t-1) \tag{4}$$

$$l_i(t) = m_i(t) - 1 \tag{5}$$

Proof. Formulas (3), (4) and (5) follow directly from the definition.

Remark 1. The served quantity is described as a minimum of a number of requests in time $t-1$ and service intensity of ith server. Due to the continuity of the flow of requests, it can happen that the capacity of the server is not fully used, i.e., $0 < m_i(t-1) < \sigma_i$. It can be seen that m_i, the number of requests, need not be an integer.

Remark 2. Having $l_i(t) \geq 0$ indicates that there is a queue, whereas $-1 < l_i(t) < 0$ indicates that the server is not fully used. But $l_i(t) = -1$ means that the server is not working.

6.2 System with Continuous Service Times and Discrete Request Flow

An example of such a system is an assembly line manufacturing spare parts for cars. The spare part can move ahead to the next service place after it has been fully completed. The speed of the machines can be adjusted.

Similarly to the previous type of system, there is one server at each service place i in this type and the manager can change the service times of the servers. The request flow is discrete: the request remains in the service until it has been completed.

Theorem 2. *Consider a system with continuous service times and discrete request flow. The amount of requests served at service place i at time t is*

$$f_i(t) = \min(s_i(t-1), \sigma_i) \tag{6}$$

The number of requests present at the service place i at time t, $m_i(t)$, and the value s_i are computed by the following recursion rules:

$If(s_{i-1}(t) = 1 \ or \ s_{i-1}(t) = 0) \ and \ s_{i-1}(t-1) > 0$

$\quad If(s_i(t) = 1 \ or \ s_i(t) = 0) \ and \ s_i(t-1) > 0$

$\quad\quad m_i(t) = m_i(t-1); \quad else \quad m_i(t) = m_i(t-1) + 1;$

$\quad endif;$

$else$

$\quad If \ (s_i(t) = 1 \ or \ s_i(t) = 0) \ and \ s_i(t-1) > 0$

$\quad\quad m_i(t) = m_i(t-1) - 1; \quad else \quad m_i(t) = m_i(t-1);$

$\quad endif;$

$endif.$

$If \ m_i(t) > 0$

$\quad If \ s_i(t-1) = 0$

$\quad\quad s_i(t) := 1;$

$\quad else$

$\quad\quad s_i(t) := s_i(t-1) - \sigma_i;$

$$
\begin{aligned}
&\qquad endif;\\
&\quad else\\
&\qquad s_i(t) := 0;\\
&\quad endif.
\end{aligned}
\tag{7}
$$

Proof. These recursion rules follow from the properties of s_i and m_i.

Theorem 3. *The length of the queue is*

$$l_i(t) = m_i(t) - 1 \tag{8}$$

Proof. The length of the queue is dependent on the number of working servers and there is only one server in each service place in this type of system.

6.3 System with Discrete Service Times and Continuous Request Flow

An example of such a system is similar to the first example of a system (with continuous service time and continuous request flow). The only difference is that in case of need, the manager can add or recall another clerk to some service place.

In the case of discrete service times, the manager can change the number of servers at the ith service place, K_i, by adding or taking away some additional server(s), preserving the condition $K_i \geq 1$. All the servers at service place i have the same service time z_i.

The remaining part of the requests that are being served at service place i by the kth server is denoted by s_{i_k}, $0 \leq s_{i_k} \leq 1$. The notation s_{i_k} will be used in Theorems 4 and 5 (for the systems with continuous service time with K_i servers at each service place).

Theorem 4. *Consider a queuing system with discrete service times and continuous request flow. Then the parameters s_{i_k}, $f_i(t)$, $m_i(t)$ and $l_i(t)$ are computed according to the following formulas.*

$$
\begin{aligned}
&If\ m_i(t) > 0\\
&\quad For\ k = 1 : K_i\\
&\qquad If\ s_{i_k}(t-1) > 0\\
&\qquad\quad s_{i_k}(t) := s_i(t-1) - \sigma_i;\\
&\qquad endif;\\
&\qquad If\ s_{i_k}(t-1) \leq 0\ and\ l_i > 0\\
&\qquad\quad s_{i_k}(t) := 1;\\
&\qquad\quad l_i(t) := l_i(t-1) - 1;\\
&\qquad endif;\\
&\quad endfor;\\
&endif.
\end{aligned}
\tag{9}
$$

$$f_i(t) = \sum_k \min(s_{i_k}(t-1), \sigma_i) \tag{10}$$

$$m_i(t) = m_i(t-1) - f_i(t) + f_{i-1}(t-1) \tag{11}$$

$$l_i(t) = m_i(t) - \sum(s_{i_k}) \tag{12}$$

Proof. The value s_{i_k} is changed according to the service phase. The quantity of requests served, $f_i(t)$, is the minimum of the remaining part of the request and the service intensity because due to the continuity of the flow of requests, it can happen that $s_{i_k}(t-1) < \sigma_i$ (the server is not fully used). The value $m_i(t)$ is computed with respect to the incoming and outgoing served parts of the requests. The length of the queue is the difference between the number of the present requests and the sum of the remaining parts at all servers in the service place.

6.4 System with Discrete Service Times and Discrete Request Flow

An example of such a system is similar to the assembly line manufacturing spare parts, except that changing the settings of the system is different. To change the speed of the process in a service place, another identical machine is added or taken away.

Similarly to the previous type of system, the manager can change the number of servers in the service place i, however the request flow is discrete.

The remaining part of the request, s_{i_k}, is computed by rules in (9).

Theorem 5. *Consider a system with discrete service times and discrete request flow. The values m_i, l_i are computed by the following rules:*

$$
\begin{aligned}
&For\ k = 1 : K_{i-1}\\
&\quad If\ (s_{i-1_k}(t) = 1\ or\ s_{i-1_k}(t) = 0)\ and\ s_{i-1_k}(t-1) > 0\\
&\qquad x := x + 1;\\
&\quad endif;\\
&endfor;\\
&For\ k = 1 : K_i\\
&\quad If\ (s_{i_k}(t) = 1\ or\ s_{i_k}(t) = 0)\ and\ s_{i_k}(t-1) > 0\\
&\qquad y := y + 1;\\
&\quad endif;\\
&endfor;\\
&m_i(t) = m_i(t-1) + x - y.
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
&For\ k = 1 : K_i \\
&\quad If\ s_{i_k}(t) > 0\ or\ s_{i_k}(t-1) > 0 \\
&\qquad o := o + 1; \\
&\quad endif; \\
&endfor; \\
&l_i(t) = m_i(t) - o.
\end{aligned} \tag{14}
$$

Proof. The value $m_i(t)$ is computed with respect to incoming requests from the previous $(i-1)$ service place (variable x) and the completed requests in service place i (variable y). The length of the queue is calculated by subtracting the number of working servers (variable o) from the number of the requests present at the service place.

For simplicity we assume that any change of the service times is performed exactly at one server by one time unit, and all more complex changes are performed as a series of such simple changes. With respect to computational complexity we consider the change costs as constant. In the following text two optimization methods are suggested. For the examples and calculations the system with continuous service time and continuous request flow is considered. The model of the system was designed using the language VBA and the tool MS Excel.

7 Suggested Methods

7.1 Method Based on the Evaluation of the System States

This method applies the Markovian property, i.e. the future development of the system states depends on its current state and finite number of previous states. The past history has been completely summarized in the current state and the system has no memory to the intent that it is not known how the current state was reached.

The method uses the evaluation of the system states. The evaluation criteria take into account just the present state and the previous one (but not the sequence of preceding events), thus we can say, that considered system is memoryless.

The suggestions for the changes in system settings are given every time unit (the so-called turn), so there is no need to penalize the change of the system setting via the addition of the change costs (in the calculation the function of total production costs), thus the parameter $C_j = 0$. Also due to continuous changes each stage is of length $T = 1$.

Remark 3. Note, that the length of the stage is adjusted to 1, therefore the turn and the stage are equivalent concepts for this method.

The following sequence of steps is performed at the end of each stage for each service place:

1. compute the characteristics, i.e. length of the queue $l_i(t)$, queue tendency $d_i(t)$, particular costs for turn, total costs for turn,
2. weight the urgency of the change, weights are denoted by $G_i(t)$,
3. suggest strategies,
4. choose the appropriate action,
5. apply the changes.

The length of the ith queue is computed according the type of the system, it depends on the above mentioned character of the requests. In general, the queue length in time (t) is equal to the queue length in time $(t-1)$ subtracted by the quantity of requests that left the queue and added by the quantity of requests that have fallen into this queue.

Queue tendency, $d_i(t)$, reflects how will be the actual queue changed with respect to the current setting; it is dependent on the $l_i(t)$, $l_{i-1}(t)$ and also on the $\sigma_i(t)$ and $\sigma_{i-1}(t)$. The value expresses the increment or decrement of actual queue and also gives the information about the intensity of this variation. If $d_i(t) > 0$, the queue is increasing; for $d_i(t) < 0$ it holds that the queue is decreasing and $d_i(t) = 0$ means that the length of the queue will not change in next turn. Computation of this parameter is again dependent on the character of the requests.

The value of particular costs for turn (considering the idle and the queuing costs) will be positive if either the $l_i(t) < K_i(t)/z_i$ in case of idle costs, or if $l_i(t) - M_i > 0$ in case of queuing costs. The total costs for turn are then computed as a sum of particular costs.

After the computation of the above mentioned basic characteristics is made, the particular situations at the service places are weighted. The weights are expressed by the multiple of Q_i,I_i according to the importance of the situation. If the queue tends to grow and the limit M_i will be exceeded in u turns (the value depends on the need to provide prompt reactions upon changes in the system), then the weight $G_i(t) = Q_i \cdot x'$, where $x' \in \langle 0, 1\rangle$ is a coefficient expressing the urgency of the reaction. If the queue tends to grow and the limit M_i is already exceeded, then the weight is intensified by the addend. $G_i(t) = Q_i \cdot (l_i(t) - M_i) + Q_i$. Formulas for weights of the queue tending to fall are constructed similarly. The overview of formulas for computation $G_i(t)$ and suggested strategies (do nothing, accelerate z_i, accelerate preceding (z_{i-1}), decelerate) for different combinations of the values $l_i(t)$ and $d_i(t)$ are shown in the Table 1.

Remark 4. Values of coefficients in the Table 1 should fulfill conditions $x' < x''$ and $y' > y''$ (according to the urgency of the situations).

With the knowledge of particular weights $G_i(t)$ it is possible to decide which of the situations is the most urgent - it is the situation with the highest value of weight, and according to this comparison, the suggested strategy can be implemented.

Table 1. Weights and strategies

	$l_i(t) = 0$	$0 < l_i(t) < M_i$	$l_i(t) > M_i$
$d_i(t) > 0$	$Q_i \cdot x'$ do nothing	$Q_i \cdot x''$ accelerate	$Q_i \cdot (l_i(t) - M_i) + Q_i$ accelerate
$d_i(t) < 0$	$I_i \cdot y'$ acc. preceding	$I_i \cdot y''$ decelerate	$Q_i \cdot (l_i(t) - M_i)$ accelerate
$d_i(t) = 0$	$I_i \cdot z'$ acc. preceding	0 do nothing	$Q_i \cdot (l_i(t) - M_i)$ accelerate

Example 1. The work of the evaluation method is illustrated in the following figures. The system with the continuous service times and requests flow is considered. During the computation of particular and total costs we can omit the work costs, because these costs are proportionally dependent on the number of servers at the service place and for this type of the system there is exactly one server at each service place. Initial parameters are set to $z_i(0) = \{3, 5, 4, 3, 6, 6, 7, 6, 4, 4\}$, $l_i(0) = \{2, 5, 6, 7, 6, 6, 8, 7, 6, 6\}$, $M_i = 3$, $Q_i = I_i = 10$.

To show the method in action, Fig. 3 depicts the 38th step of the optimization and values of computed parameters. One of the queues tends to grow (the queue in front of the fifth server), some of them tend to fall and some remain stable (the queues in front of the servers $6, 8, 10$). We can observe, that the greatest weight is $G_5(38) = 45,5$. It means that the service time at the fifth service place will be accelerated in the next turn because the maximal tolerable limit of the queue length ($M_5 = 3$) have been already exceeded ($l_5(38) = 6,55$) and queue tends to grow.

38	$z_i(38)$	$l_i(38)$	queue tendency	Idle costs	Queuing costs	Total costs for turn	$G_i(38)$	Strategy	Total	
1	3	2,							Total cumulative costs	10 893,5
2	2	5,65	-0,1667	0,	26,5	26,5	26,5	accelerate		
3	1	4,6	-0,5	0,	16,	16,	16	accelerate	Total for turn	127,6667
4	0,5	5,25	-1,	0,	22,5	22,5	22,5	accelerate		
5	2	6,55	1,5	0,	35,5	35,5	45,5	accelerate		
6	2	3,	0,	0,	0,	0,	0	do nothing		
7	1	3,6333	-0,5	0,	6,3333	6,3333	6,3333	accelerate		
8	1	2,6667	0,	0,	0,	0,	0	do nothing		
9	0,5	2,9	-1,	0,	0,	0,	7,5	decelerate		
10	0,5	5,0833	0,	0,	20,8333	20,8333	20,8333	accelerate		

Fig. 3. 38. stage of the optimization

The development of the service times at particular service places during the optimization for $t = \{0, 1, \ldots, 17\}$ and also the reaction of the queue lengths in time connected with costs are shown in Fig. 4. The starting system setting can be seen in the column with index 0. As we can see, the longest queue in time $t = 0$ is in front of the 7th server (red numbers in the table of queue lengths). Costs for turn are also high (50 units). Again, according to the previous considerations,

$z_i(t)$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	3	3	3
3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
5	6	6	6	6	6	6	5	4	3	3	3	3	3	3	3	3	3	3
6	6	6	6	6	6	6	6	6	6	6	6	5	4	3	3	3	3	3
7	7	6	5	5	4	4	4	4	4	3	3	3	3	3	3	3	3	3
8	6	6	6	5	5	4	4	4	4	4	3	3	3	3	3	3	3	2
9	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3
10	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
$l_i(t)$																		
1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	5	5,13	5,27	5,4	5,53	5,67	5,8	5,93	6,07	6,2	6,33	6,47	6,6	6,73	6,82	6,82	6,82	6,82
3	6	5,95	5,9	5,85	5,8	5,75	5,7	5,65	5,6	5,55	5,5	5,45	5,4	5,35	5,35	5,43	5,52	5,6
4	7	6,92	6,83	6,75	6,67	6,58	6,5	6,42	6,33	6,25	6,17	6,08	6	5,92	5,83	5,75	5,67	5,58
5	6	6,17	6,33	6,5	6,67	6,83	6,97	7,05	7,05	7,05	7,05	7,05	7,05	7,05	7,05	7,05	7,05	7,05
6	6	6	6	6	6	6	6,03	6,12	6,28	6,45	6,62	6,75	6,83	6,83	6,83	6,83	6,83	6,83
7	8	8	7,97	7,93	7,85	7,77	7,68	7,6	7,52	7,35	7,18	7,05	6,97	6,97	6,97	6,97	6,97	6,97
8	7	7	7,03	7,03	7,08	7,08	7,08	7,08	7,08	7,17	7,17	7,17	7,17	7,17	7,17	7,17	7,17	7
9	6	5,92	5,83	5,78	5,73	5,73	5,73	5,73	5,73	5,73	5,82	5,9	5,98	6,07	6,15	6,23	6,23	6,4
10	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6,08	6,17
costs for turn																		
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	20	21,3	22,7	24	25,3	26,7	28	29,3	30,7	32	33,3	34,7	36	37,3	38,2	38,2	38,2	38,2
3	30	29,5	29	28,5	28	27,5	27	26,5	26	25,5	25	24,5	24	23,5	23,5	24,3	25,2	26
4	40	39,2	38,3	37,5	36,7	35,8	35	34,2	33,3	32,5	31,7	30,8	30	29,2	28,3	27,5	26,7	25,8
5	30	31,7	33,3	35	36,7	38,3	39,7	40,5	40,5	40,5	40,5	40,5	40,5	40,5	40,5	40,5	40,5	40,5
6	30	30	30	30	30	30	30,3	31,2	32,8	34,5	36,2	37,5	38,3	38,3	38,3	38,3	38,3	38,3
7	50	50	49,7	49,3	48,5	47,7	46,8	46	45,2	43,5	41,8	40,5	39,7	39,7	39,7	39,7	39,7	39,7
8	40	40	40,3	40,3	40,8	40,8	40,8	40,8	40,8	41,7	41,7	41,7	41,7	41,7	41,7	41,7	41,7	40
9	30	29,2	28,3	27,8	27,3	27,3	27,3	27,3	27,3	27,3	28,2	29	29,8	30,7	31,5	32,3	32,3	34
10	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30,8	31,7

Fig. 4. The development of service times, queue length, and costs

because the queue tends to grow and the costs are the highest, the strategy with the highest weight was to accelerate the 7th server (we can see it from the table of service times, in time $t = 1$ the service time $z_7 = 6$.

Figure 5 depicts the evolution of costs for turn and the total costs (total costs are connected to the secondary axis in the graph) in comparison to the evolution of the system costs of unoptimized system. At the start, the costs are nearly the same, and then the optimization takes effect. It can be seen, that after 67 turns the system reaches the state where the queues are of acceptable length (not too long and/or not empty) and thus do not create any undesirable costs.

7.2 Method Based on the Simulation of Future States

The system can be affected by the manager by changing the service intensity at some service place. The main reason is to reduce total costs and optimize the manufacturing process. The decision is divided into two parts. First, it must be decided at which moment the change in the service intensity will be done in order

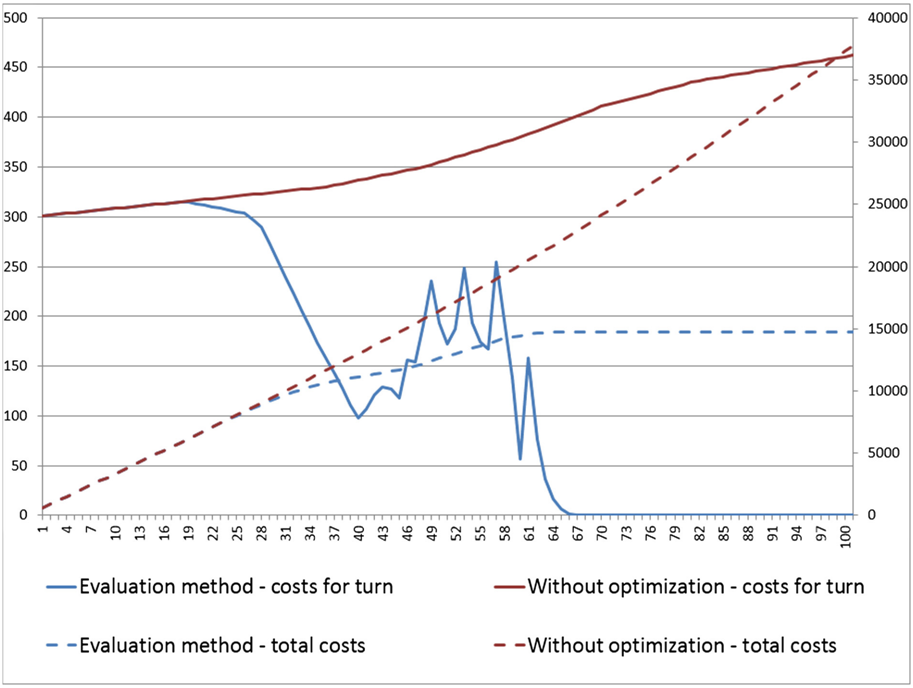


Fig. 5. The evolution of costs

to minimize the average production costs. Second, it has to be decided which of the service intensities will be changed. If the queue at service place i is too long and is still increasing, then the service intensity z_i should be accelerated. This can be done, for example, by adding a parallel service machine at position i. On the other hand, if the server(s) in the service place i is (are) idle, then z_i can be diminished.

Second method is based on the simulation of the system future development for several possible settings and on the computation of the function of average production costs for each of these simulations. Unlike the previous method, the suggestion to change the system setting is not given every time unit (for some types of systems it is undesirable to change the system setting continuously), thus the length of the stage is $T \geq 1$. The intention is to let the system perform for some time period and influence it only in time when necessary - not too soon but also not too late - therefore also $C_j \geq 0$. For computational simplicity we assume the change at exactly one service place at the same time.

Assuming that the stage will end in $T = t$, the average production costs during the stage are computed according to the following formula:

$$E(T, j, \delta) = \frac{P(T, j)}{T}. \tag{15}$$

Parameter j indicates the service place where the change in the system setting is made, δ expresses the intensity of this modification (for example, $j = 5$ and $\delta = 3$ means, that the change is considered at fifth service place and in dependence on the type of service times described above, either the service time of the server is decelerated by $+3$ time units or the number of servers is increased by $+3$ units).

At the beginning of each stage the rth set of functions $E(T, j, \delta)^{(r)} = \frac{P(T,j)}{T}$ is computed (in other words, index r indicates the number of stage). This set represents the evolution of the system average production costs for all possible intended settings. Optimal setting is the one that corresponds to the function of the set which contains the global minimum of the rth set. This minimum is very important, because the time when this function reaches its minimum is the time convenient for next change of the system setting (because from this moment the average costs are increasing) - and the question What will be the change like? will be answered by choosing the right function from next, $(r+1)$th, set of functions.

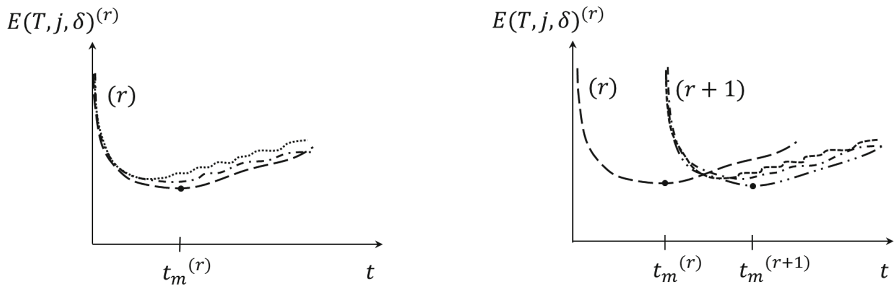


Fig. 6. The decision process

The process is illustrated by Fig. 6. The curves of the graphs are not smooth. This is a result of the variation in the development of the queues. The functions marked as (r) (the left part of the figure) represent the rth stage and were computed by calculating the function of average costs if the stage ends at time T for each possible change. The result of the first step of the process is the time $t_m^{(r)}$ (global minimum of the set of functions), which answers the question When to make the change? because the continued work of the system will lead to the growth of the average production costs.

Remark 5. For the first stage it holds that the $E(T, j, \delta)^{(1)}$ is the one and only one because it was assumed the change costs are constant.

Example 2. The method is again illustrated on the example of the system with continuous both, the service times and request flow. Initial parameters are set to $\sigma_i(0) = \{3, 5, 4, 3, 6, 6, 7, 6, 4, 4\}$, $l_i(0) = \{2, 5, 6, 7, 6, 6, 8, 7, 6, 6\}$, $M_i = 3$, $I_i = Q_i = 10$, $C_i = 30$. For simplicity, the changes in service times are considered to be unit ($\delta = \pm 1$).

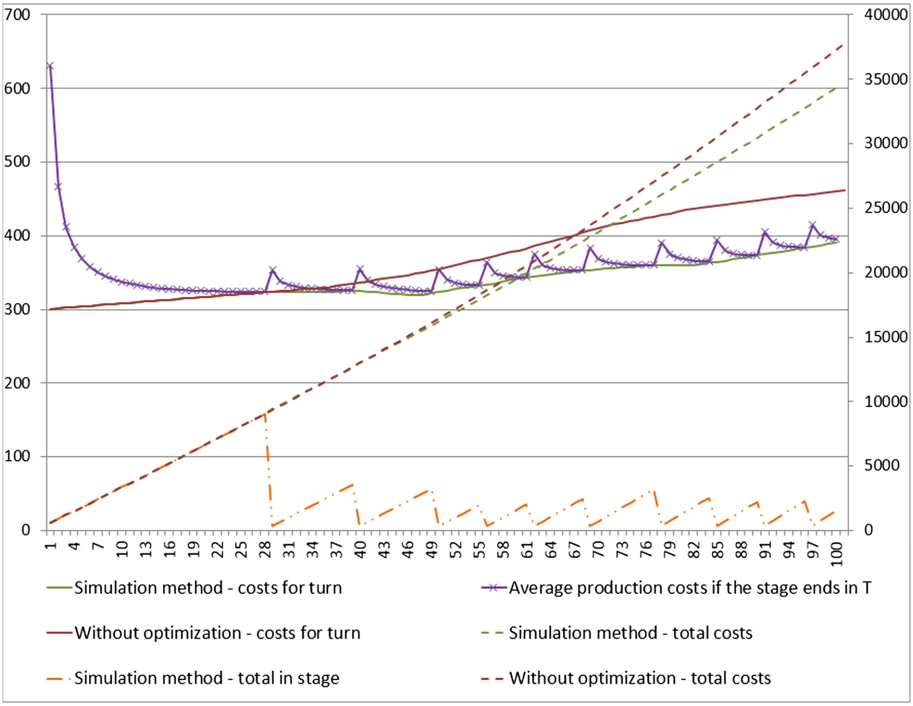


Fig. 7. The evolution of costs

The time evolution of costs for 100 turns are depicted in Fig. 7. The graph shows costs for turn and total costs in comparison to the evolution of costs of the system without optimization (total costs are connected to secondary axis in the graph).

The graph also shows the function of average production costs if the stage ends in T (purple one with marks). This function consists of parts of particular curves related to the chosen settings during the optimization process. During the first stage the function reaches its minimum after 28 turns. For the time $t = 29$, the next set of functions was computed and one of the settings was chosen. The new stage began. The part of the curve related to chosen setting corresponds to the part of the purple curve for $29 < t < 39$. Again, for time $t = 40$ the next set of functions was computed.

8 Conclusions and Future Work

The study in this paper focuses primarily on the deterministic discrete event systems and possible issues connected with the development of the queues during the system run. Four types of such systems were distinguished. Focus of the study lies in the optimization of the production process.

Two methods for cost optimization of the tandem queuing network are suggested. Each method is applicable for different system type. These two methods should not be mutually compared due to different frequency in changing the system setting and penalization in the form of change costs, that are completely omitted in case of the first method. It can be seen, that in case of the Method based on the evaluation, the reactions on the issues that can arise is really very quick and pointed. The optimization with use of the Method based on the simulation is suitable especially for systems with long run, where frequent changes in setting are unwanted (the system setting is influenced only in time, when the average production costs reach the minimum).

The motivation to study the DES primarily resulted from the creation of the agent-based economy model developed as an experimental study of autonomous behavior of the virtual economy system. As the system of virtual economy is complex, there are several types of agents in the system. Besides the so-called production agents, we can find here mining agents, service agents, transportation agents, consumer agents and others. In [12] you can find the description of the multiagent subsystem - the production unit. The next step of our research is to include the methods suggested in this paper to the virtual economy model.

Acknowledgments. The financial support of the Specific Research Project Socio-economic models and autonomous systems at FIM UHK is gratefully acknowledged. The authors would like to thank Petr Blecha for his insightful comments and help.

References

1. Baum, D., Lothar, B.: An Introduction to Queueing Theory and Matrix-Analytic Methods. Springer, Dordrecht (2005). https://doi.org/10.1007/1-4020-3631-0
2. Le Boudec, J.Y., Thiran, P.: Network Calculus: A Theory of Deterministic Queuing Systems for the Internet, vol. 2050. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45318-0
3. Buzacott, J.A., Shanthikumar, J.G.: Stochastic Models of Manufacturing Systems, vol. 4. Prentice Hall, Englewood Cliffs (1993)
4. Cassandras, C.G., Lafortune, S.: Introduction to Discrete Event Systems. Springer, USA (2008). https://doi.org/10.1007/978-0-387-68612-7
5. Cinlar, E.: Introduction to Stochastic Processes. Prentice Hall, Englewood Cliffs (1975)
6. Delgoshaei, A., Gomes, C.: A multi-layer perceptron for scheduling cellular manufacturing systems in the presence of unreliable machines and uncertain cost. Appl. Soft Comput. **49**, 27–55 (2016)
7. Gavalec, M., Němcová, Z.: Cost optimizing methods for deterministic queuing systems. In: Nguyen, N.-T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds.) ICCCI 2016. LNCS (LNAI), vol. 9875, pp. 335–344. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45243-2_31
8. Hum, S.H., Parlar, M., Zhou, Y.: Measurement and optimization of responsiveness in supply chain networks with queueing structures. Eur. J. Oper. Res. **264**, 106–118 (2018)
9. Lawrence, J.A., Pasternack, B.A.: Applied Management Science. Wiley, Hoboken (1997)

10. Papadopoulos, H.T., Heavey, C.: Queueing theory in manufacturing systems analysis and design: a classification of models for production and transfer lines. Eur. J. Oper. Res. **92**(1), 1–27 (1996)
11. Tempelmeier, H.: Practical considerations in the optimization of flow production systems. Int. J. Prod. Res. **41**(1), 149–170 (2003)
12. Tucnik, P., Nemcova, Z., Nachazel, T.: Multiplant production design in agent-based artificial economic system. In: Nguyen, N.T., Papadopoulos, G.A., Jędrzejowicz, P., Trawiński, B., Vossen, G. (eds.) ICCCI 2017. LNCS (LNAI), vol. 10448, pp. 371–380. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67074-4_36
13. Zeigler, B.P., Praehofer, H., Kim, T.G.: Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems. Academic press, Cambridge (2000)

WordNet and Wiktionary-Based Approach for Word Sense Disambiguation

Mohamed Ben Aouicha[1(✉)], Mohamed Ali Hadj Taieb[1], and Hania Ibn Marai[2]

[1] Multimedia InfoRmation System and Advanced Computing Laboratory, Sfax University, Sfax, Tunisia
mohamed.benaouicha@fss.usf.tn
[2] Faculty of Sciences, Sfax University, Sfax, Tunisia

Abstract. Word sense disambiguation (WSD) is the ability to identify the meaning of words in context in a computational manner. WSD is considered as a task whose solution is at least as hard as the most difficult problems in artificial intelligence. This is basically used in application like information retrieval, machine translation, information extraction because of its semantics understanding. This paper describes the proposed approach W3SD (This paper is an extended version of our work [4] published in the 8th International Conference on Computational Collective Intelligence.) which is based on the words surrounding the polysemous word in a context. Each meaning of these words is represented by a vector composed of weighted nouns using WordNet and Wiktionary features through the taxonomic information content from WordNet and the glosses from Wiktionary. The main emphasis of this paper is feature selection for disambiguation purpose. The assessment of WSD systems is discussed in the context of the Senseval campaign, aiming at the objective evaluation of our proposal to the systems participating in several different disambiguation tasks.

Keywords: Word sense disambiguation · WordNet · Wiktionary
Information content · Gloss

1 Introduction

This work is an extended version of a previously published work [4] in the 8th International Conference on Computational Collective Intelligence. It proposed a method for Word Sense Disambiguation (WSD) based only on the WordNet and the information content approach. In the present work, we integrate the Wiktionary as additional ressource for enriching the WordNet information in the weighting mechanism for providing weights for nouns. In fact, we proceed through a matching process between the glosses of WordNet and the glosses in Wiktionary. Then, the Wiktionary glosses were pretreated for identifying the nouns which will be exploited, in conjunction with the nouns in WordNet, in the quantification of the information content of a word.

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXIX, LNCS 10840, pp. 123–143, 2018.
https://doi.org/10.1007/978-3-319-90287-6_7

WSD is a problem encountered in most natural language processing systems for resolving the ambiguity of polysemous words. This concept was introduced for the first time by Warren in 1949 [1]. He stressed that the ambiguity of words must be solved to enable automatic translation between languages. This ambiguity may be syntactic or semantic. The first was largely solved by using the markers parts of speech predicting the grammatical category of words and with very high accuracy as the approach of Brill [2]. The problem is that words often have more than one meaning. The identification of the correct meaning of a word is called Word Sense Disambiguation (WSD). The objective, therefore, of a WSD system is to choose the right sense for a word in a context, usually a sentence, if that word has more than one meaning.

WSD existing approaches are usually based on various principles. They can be divided into three categories: the corpus-based, the knowledge-based and hybrid approaches which are based on the combination of both resources (corpus and knowledge resource). One of the problems with the conventional methods of WSD is the need to use an annotated corpus. In this respect, our goal was to develop a new approach to WSD without needing to an annotated corpus, based on the words surrounding the target word and a weighting mechanism applied on the nouns using taxonomic Information Content (IC) method through the WordNet [3] “is a” taxonomy.

The rest of the paper is organized as follows: Sect. 2 provides an overview about the WSD approaches. Section 3 presents our proposal W3SD including the weighting mechanism and the disambiguation process based on co-occurring words. Section 4 reports on the evaluation and comparison of our approach through different benchmarks. The final section is devoted to presenting our conclusions and future work.

2 Related Works

Several approaches have been developed as part of the WSD. This section contains a description of a number of existing approaches in the field of WSD and the characteristics of each of them. We can distinguish three categories of approaches: corpora-based, knowledge-based and hybrid.

The performance of a WSD system is related to the knowledge source. Furthermore, the importance of this problem is related to the fact that many applications based on it in their treatment processes. Solving the WSD problem in an efficient way will have a positive effect on the efficiency of systems such as machine translations, document summarization, question answering systems and Information Retrieval (IR).

The disambiguation of text can provide a major push in the treatment of large-scale amounts of data, thus participating via a solid contribution to the realization of the so-called semantic Web, “an extension of the current Web, in which information is given well-defined meaning, better enabling computers and people to work in cooperation” [5]. The potential of WSD is also clear when we deal with the problem of machine translation because the polysemous words have multiple entries in the target language. In next paragraphs, we will discuss some applications exploiting the WSD.

- Information retrieval:

In IR, WSD can bring two ways of benefits [6, 7]. First, queries may contain ambiguous words (terms), which have multiple senses. The ambiguities of these query words can have a negative effect on the retrieval precision. So, the identification of the correct meaning of the polysemous words in both queries and documents can improve retrieval precision. Secondly, query words may have semantic related meanings with other words not in the query. The enrichment by these relations between words can improve retrieval recall.

However, in relation with the impact of WSD on IR, different researchers arrived at conflicting observations. Some of the early research showed a drop in retrieval performance by using word senses. Some other experiments observed improvements by integrating word senses in IR systems.

Guyot et al. think that WSD-based indexing is a promising approach for language-independent indexing and retrieval systems [8]. Although, an efficient WSD is important for creating conceptual indexes, they demonstrated that ambiguous indexes (several concepts assigned to same terms) are often sufficient to reach a good multilingual retrieval performance.

Sanderson shows that an IR system will improve its performance if the documents it retrieves are represented by word senses rather than words [9]. Recently, research was conducted to investigate this method of document representation. The researchers used a word sense disambiguator to disambiguate an IR test collection. However, experiments using this disambiguated collection showed a drop in retrieval performance.

Stokoe et al. investigated the use of a state of the art automated WSD system within a web IR framework. The focus of this research is to perform large scale evaluation of both the automated WSD algorithm and the IR system. They demonstrate relative performance of an IR system using WSD compared to a baseline retrieval approach such as the vector space model [10].

- Machine translation:

Capuat and Wu [11] were among the first to treat the common assumption that WSD should improve upon machine translation performance, demonstrating that many of the contextual features that are important in performing WSD were in fact already implicit in the language models that are trained to perform machine translation. Although the rich semantic data exploiting WSD algorithms lead to better predictions of lexical choice. They showed that by training a machine translation system on complete parallel sentences they could obtain higher BLEU scores than a comparative system using the output of WSD for isolated words into the translation model [11]. These remarks were motivated the researchers to reformulate the WSD process in such a way as to make it useful for machine translation systems.

The main work adapting the WSD for machine translation came from [12], who proposed that multi-word phrases as opposed to single words be considered as ‘phrase-sense disambiguation’. Their approach is based on the fact that machine translation models are already trained using contextual features from full sentences.

They suggested that phrase-based sense disambiguation could be more useful to the sentence-based translation models used for machine translation.

Neale et al. performed highly-efficient WSD over an accurate graph-based representation of the knowledge base WordNet [13]. Training these models over a large, open domain corpus, they have obtained statistically significant improvements in BLEU score (translating from English to Portuguese) when compared to a baseline version of their machine translation system.

- Text categorization:

Word ambiguity is a hard problem in the keywords-based applications. For example, if '*bank*' occurs several times in a document, should the file be classified to "*geological formation*" or "*financial institution*"? Each word may have multiple senses, and multiple words may have the same sense. It is not trivial for a computer to know which sense the keyword is using in a given context.

Liu et al. proposed a text classification method based on sense disambiguation [14]. Each keyword in a given document is mapped to the concept hierarchy where each sense maintains a counter. This algorithm is applied to Brown Corpus. The effectiveness of their automatic text classification method is evaluated by comparing the classification results with the classification results using manual disambiguation offered by Princeton University.

- Ontologies mappings:

Lexical similarity measures are a main component in ontology mapping procedures. Schadd and Ross proposed the inclusion of word-sense disambiguation techniques into lexical similarity metrics to disambiguate ontology concepts [15]. So, they proposed a concept disambiguation method which identifies corresponding senses using a virtual document model. For each given concept and possible senses, a parametrized model is used to gather relevant information into separate virtual documents, such that their document similarities indicate which sense is the most likely to denote the meaning of the given concept.

- User Profiling:

Semeraro et al. proposed a method for annotating senses to terms in short queries, and also described an approach to integrate senses into an IR task [16]. In the experiment on four query sets of TREC collection, they compared the performance of a supervised WSD method and two WSD baseline methods.

Their experimental results showed that the integration of senses improved a state-of-the-art baseline. The performance of applying the supervised WSD method is better than the other two WSD baseline methods. They also proposed a method to further integrate the synonym relations to the learning machine approaches.

With the integration of synonym relations, their best performance setting with the supervised WSD achieved an improvement of 4.39% over the baseline method, and it outperformed the best participating systems.

- Semantic role labeling:

Che and Liu adopted the all word sense features into Semantic Role Labeling (SRL). The performed experiments show that the determination of the right sense using WSD system can improve the performance significantly [17]. They presented a Markov logic model that jointly models all word sense disambiguation and SRL. They conclude that it is easy to implement the joint model of all word sense disambiguation and SRL with Markov logic.

- Vocabulary learning:

Kulkarni et al. show that performing sense disambiguation for homonyms helps vocabulary learning in English as a Second Language (ESL) students [18]. They demonstrated that the task of disambiguating homonyms can be automated through the learning classifiers. In fact, they can assign the appropriate sense to a homonym in a specific context with high accuracy. A study, focusing on users, reveals that students equipped with WSD-enabled vocabulary tutor perform better than students using vocabulary tutor without the WSD capabilities.

- Entity Linking:

Moro et al. presented Babelfy[1], a novel, integrated approach to Entity Linking and Word Sense Disambiguation. This solution is based on three key steps: firstly, the automatic creation of semantic signatures, i.e., related concepts and named entities, for each node in the reference semantic network; secondly, the unconstrained identification of candidate meanings for all possible textual fragments; and finally, linking based on a high-coherence densest subgraph algorithm [19]. They used BabelNet[2] as multilingual semantic network.

Two key features characterize BabelNet, its multilinguality and its integration of lexicographic and encyclopedic knowledge. Therefore, it is possible to run the two tasks of Entity Linking and WSD in any of the languages covered by the semantic network.

However, they also demonstrated that BabelNet in itself does not lead to state-of-the-art accuracy on both tasks, even when used in conjunction with a high-performance graph-based algorithm like Personalized PageRank.

2.1 Corpus-Based Approaches

The corpus-based approaches exploit the textual content of the corpus. This approach seeks to find the correct meaning of an ambiguous word based on the context formed by the neighboring words (co-occurrence). In what follows, we present a study of some works belonging to such approaches.

Véronis [20] proposed HyperLex, an approach to automatically determine the various uses of a word in a corpus. The correct meaning of an ambiguous word is determined using the concept of co-occurrence graphs. A co-occurrence graph is

[1] http://babelfy.org.

[2] http://babelnet.org/.

constructed from the context where the word appears (nouns and adjectives). For a polysemous word, a corpus is created through the search of the word in the web using the meta-search engine Copernic Agent via both singular and plural forms of the word. They select the component receiving the highest weight as disambiguating.

Nameh et al. approach [21] is intended to affect the right sense in an ambiguous word by comparing the context in which it appears with the existing texts in the corpus annotated by the senses. This algorithm is divided into two main steps; the first is the extraction of feature vectors. In fact, for the ambiguous word, the vector of characteristics is represented by the set of words that appear in context. Each component of the vector is represented by the number of times that a word appears in the context of ambiguous word. Then, they compare the vector of ambiguous word with any other vector already built using the cosine measure.

2.2 Knowledge-Based Approaches

The rise of electronic dictionaries and lexical database offered another direction in the development of automatic disambiguation field. This new perspective has materialized through the methods based on the knowledge that try to extract from these resources automatically the information necessary for disambiguation. These approaches use an external resource (typically WordNet) to extract useful information for disambiguation.

Lesk [22] calculates the number of words in common between two definitions from a dictionary. Applied to the disambiguation, Lesk inputs a context (sentence) and the word to disambiguate; it compares the gloss of each sense of the target word with the glosses of any other word in the sentence. The correct meaning is the one assigned to the definition that shares the greatest number of words with the definitions of other words.

Banerjee and Pedersen [23] proposed a new approach derived from Lesk improving the definitions of words through their enrichment by the definitions of concepts which are linked to a given hierarchized concept (hyperonymy, metonymy, holonomy, etc.).

Sinha and Mihalcea [24] have proposed an approach based on the graphs to represent the meaning of ambiguous words. Initially, they created the graph representing the words, their meaning and the dependencies between them. Thereafter, they used a combination of three semantic similarity measures (Jiang and Conrath [25], Leacock and Chodorow [26] and Lesk [22]) for calculating the proximity between the different grammatical categories.

Lopez-Arevalo et al. presented an approach for domain-specific WSD for the lexical sample task. The basic idea is that words co-occurring with an ambiguous word are used to identify words from contextual definitions of such ambiguous word. Then, semantic similarity values are generated and added to its association degree to disambiguate each instance of the ambiguous word in the corpus. The results of applying the approach demonstrate that integrating such information it is possible to construct a competitive method for domain-specific WSD according to results reported in the literature [27].

Panchenko et al. proposed an unsupervised knowledge-based approach to WSD based on the hybrid aligned resource [28] introduced by [29]. The key difference of this approach from hybrid methods based on sense embeddings, e.g. [30], is the fact that they exploit lexical representations making the sense representation interpretable and allow using this representation for word sense disambiguation. Their method requires no mapping of texts to a sense inventory and thus can be applied to larger text collections.

2.3 Hybrid Approaches

To assign the correct meaning of words, hybrid approaches combine several types of resources (word frequency, context, definition and relationships).

Hessami et al. proposed an approach [31] which uses the notion of trees and graphs of co-occurrence. They use WordNet as an external resource to extract the meaning of words. The algorithm is applied to a text corpus by treating its sentences. It allows the disambiguation of all words of context (phrase). They create a graph G (V, E) with V is the set of vertices and E is the set of edges. Then, they create trees for each sense in G from the relationships "is a" of WordNet; then they search the tree node that exists in the graph: If it exists, an edge is created in the graph between the root of the tree and this node.

Basile et al. [32] inspired from the Lesk algorithm. Their basic idea was to improve the Lesk algorithm by replacing the notion of overlap between the glosses by measuring cosine similarity. They use Babelnet [33] as inventory senses. Having w_i (target word), the first step is to find its way from Babelnet. Creating the context: the words right and left words of w_i in context. Then, the glosses are expanded with linked concepts to a particular sense.

3 W3SD: The Disambiguation Process

Figure 1 illustrates the different steps in the proposed method W3SD. Our basic idea was to introduce the concept of IC in WSD process which is divided into two main modules. The first module concerns the pre-processing including the extraction of grammatical categories, simple and compound nouns and their transformation into their canonical forms. The second module focuses on the steps of the disambiguation process. It begins by representing each target word by a feature vector. A feature vector is formed by the weighted nouns co-occurring with the target word. They are created using the nouns of glosses (extracted from WordNet and Wiktionary) and synset glosses that are related to the target concept. The weight is computed using an IC-computing method. Compared to the related works, our proposal is considered a hybrid approach because it exploits the textual information (glosses) and the knowledge structure through the WordNet «is a» taxonomy for calculating the information content. The following paragraphs describe the exploited resources.

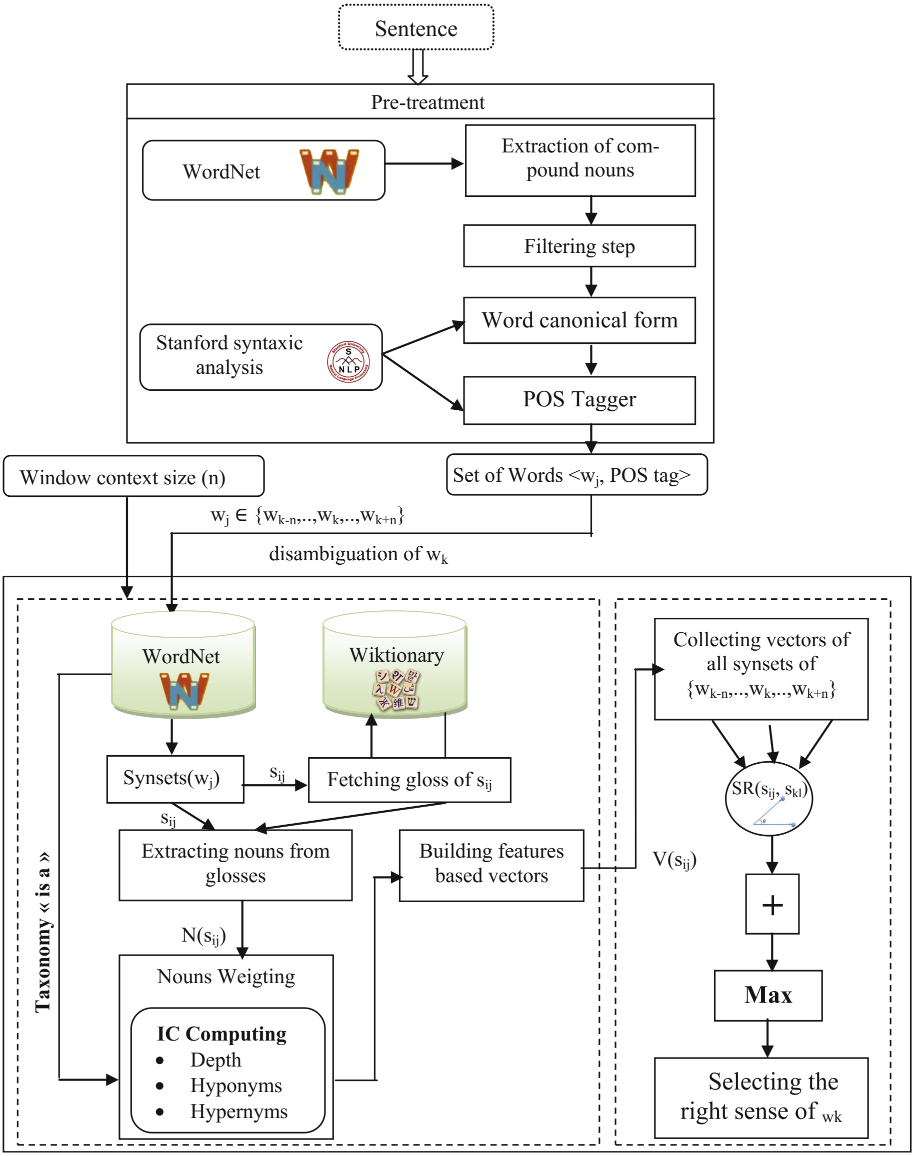


Fig. 1. The W3SD process for word sense disambiguation

- WordNet

The WordNet [3] organizes the lexical information in meanings (senses) and synsets (set of synonym words in a specific context). Each synset has a gloss that defines the concept. For example the words {car, auto, automobile, and motorcar} are a synset that represents the concept defined by the gloss "four wheel Motor vehicle, usually propelled by an internal combustion engine". Hypernymy is a relation that organizes noun synsets into lexical inheritance taxonomy (Fig. 2). In this taxonomy, a subordinate term inherits the basic features from the superordinate term and adds its distinctive features to form its own meaning.

- Wiktionary

Wiktionary[3] is a free online content dictionary written and edited collaboratively by volunteers. It includes a wide range of lexical and semantic information, such as definitions, pronunciation, translations, inflected forms, and glosses. The Wiktionary dump is handled via the API JWKTL package [34]. A lexical entry in Wiktionary is represented by a set of senses which can be aligned with the synsets in WordNet. The exploited version in this work is the dump dated March 28, 2015.

The basic idea is the computing of semantic relatedness degree between all senses of the target word and any sense of any word belonging to its context using the cosine measure. Each combination is represented by the sum of the values of semantic similarities between a sense s_{kj} and a set of senses each from a word (pertaining to the context). The combination having the highest sum will be chosen as representative of the senses s_{kj}. The last step concerns the choice of the sense with the highest value as the correct sense of the word that will be annotated. The corresponding sense to an ambiguous word is done using WordNet 3.0. However, before embarking on the disambiguation process, a pretreatment phase composed of two steps is necessary.

Extraction of Simple and Compound Nouns: It is used to browse the sentence and seek all simple and compound nouns. This is important because it affects the results of the disambiguation process due to the fact that the meaning of a compound noun differs from the meaning of its composed words. For example, the words "*computer*", "*science*" and "*computer science*" have different senses.

Determining Part of Speech: In this step, we considered that the words with grammatical category noun, verb, adjective or adverb. Every other word is eliminated. Grammatical categories of words are extracted using the Stanford parser because it treats the sentence as a whole and determines grammatical categories of words through the complete structure of the sentence. Researching the canonical forms of words is necessary to gain access to WordNet.

The disambiguation process takes as entry the target word and its context. Let C be the context constituted from n words: $C = \{w_1, w_2, w_3, \ldots, w_n\}$ with w_k is the word to disambiguate ($1 \leq k \leq n$). The first step is to check if the target word has one or more senses. If it has only one sense, it will be the disambiguating sense. Else, it will follow the disambiguation process. The steps of this process are described by the following Algorithm:

[3] https://en.wiktionary.org/wiki/Wiktionary:Main_Page.

```
Algorithm : W3SD
Input
w: word to be annotated,
Vs: vector formed by the words composing the sentence s. It contains the pairs
<wi, POSwi> where wi is a word and POSwi is its part of speech.
Output
Idx: the index of the selected sense of the word w
Begin
Senses← sensesInWordNet(w) // the senses of w in WordNet 3.0
If (|Senses| = = 1) Then
choose the singleton s1 ∈ Senses as the target sense of w
Idx←1
Else
Cw← extractContext (s, w, pos, n) // pos is the position of w in the sentence s
//and n the size of the context
For each wi ∈ Cw do
SF ← extractSenseFeatures(wi) //SF is a set of vectors representing the senses
//of the word wi and  composed from the extracted nouns
For each s ∈ SF do
     computeNounsWeights (s) // compute the weight of each noun in vector s
End For
End For
Idx← findCorrectSense (SF, w) // the index of the selected sense of the word w
End If
End
```

The method "*extractSenseFeatures(*w_i*)*" is intended to represent every meaning of a word present in the context of the word w by a feature vector. Nouns are generally the least ambiguous words with respect to verbs, hence our choice of nouns is only to represent the feature vectors of the meaning of a word. For a given sense of the word, its feature vector consists of all words extracted from glosses and the glosses of synsets that are connected to it. Since the synset glosses are generally short, we have chosen to enrich them by synset glosses that are related to them.

This method aims to extract for each sense assigned to a word in WordNet. The same way is applied to different grammatical categories of words by changing only the relations according to each category. For a given synset, its gloss is divided into two parts: the definition and the example. Each part follows a parsing step to determine the part of speech of each word. Thereafter, only the nouns that are extracted from glosses of the concerned synset and its related are grouped in a single vector. Any word belonging to this vector will be weighted using a mechanism based on the information content.

3.1 Weighting Mechanism Based on IC

The weighted mechanism is based on noun gloss overlap quantified using nouns' weights computed with our IC computing method. All the nouns pertaining to the set of glosses of the ancestors' subgraph assigned to target sense are extracted. In the next paragraphs, we detail the steps of the computing process in order to provide the weight of the word w_i.

3.1.1 Computing Noun Weight

In this section, we describe the method used to compute the weight assigned to each noun pertaining to the set of glosses representing a specific concept. A concept is represented by a set of nouns extracted from the glosses of the concepts pertaining to its ancestors' subgraph. The weight attributed to each word *w* is based on the IC computing method proposed by Hadj Taieb et al. [35] by using the set of synsets in WordNet *Syn(w)* ascribed to the target word *w* as follows:

$$weight_W(w) = \sum\nolimits_{c \in Syn(w)} IC(c) \tag{1}$$

Next, this function is exploited to quantify the weight assigned to a concept $weight_C(c_i)$. The function $weight_C(c_i)$ is expressed as follows:

$$weight_C(c) = \sum\nolimits_{w \in Nouns(c)} weight_W(w) \tag{2}$$

Where *Nouns(c)* is the set of nouns extracted from the glosses representing the concept *c* in WordNet and in Wiktionary. In fact, from WordNet, we use the glosses of the concepts pertaining to *Hyper(c)* representing the hypernyms of *c* including itself.

In fact, glosses are split into tokens using Stanford[4]. Only the words determined as nouns are kept after obtaining their canonical form:

$$Nouns(c) = \bigcup_{g \in Gloss(c)} \{nouns\ in\ the\ gloss\ g\} \tag{3}$$

where *Gloss(c)* is composed of two subsets as follows:

$$Gloss(c) = Gloss_{WNet}(c) \cup Gloss_{Wikt}(c) \tag{4}$$

The subset $Gloss_{WNet}(c)$ extracted from WordNet is formed by the glosses of the concepts forming the ancestors' subgraph *Hyper(c)* including *c, and* the glosses pertained to neighborhood of the concept *c* through the semantic relations (e.g. hyponymy, hypernymy, holonymy, pertainym, etc.). The subset $Gloss_{Wikt}(c)$ contains the glosses pertaining to Wiktionary and corresponding to the concepts of *Hyper(c)*. As defined above, the function *Hyper(c)* returns the set of ancestors of a concept *c* in the "*is a*" taxonomy of WordNet. $Gloss_{WNet}(c)$ *and* $Gloss_{Wikt}(c)$ are defined as follows:

[4] http://nlp.stanford.edu/software/.

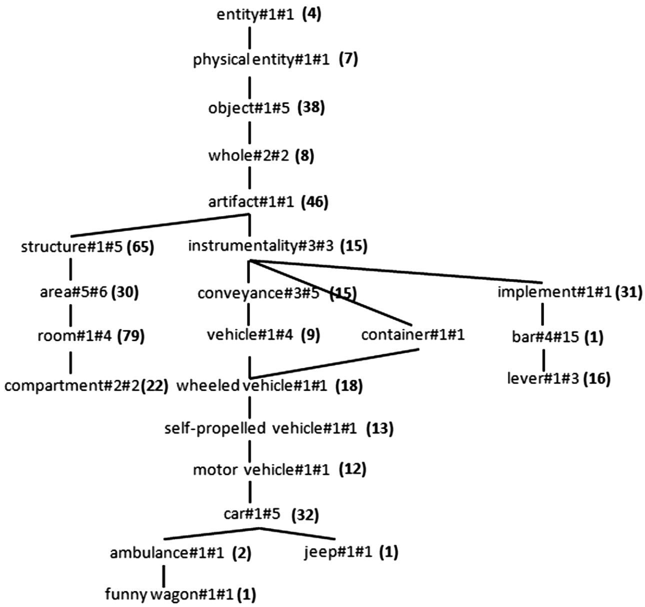


Fig. 2. WordNet "*is a*" taxonomy fragment. The number between parentheses is the number of direct hyponyms of each concept including itself. The notation *w#i#j* refers to the synset number *i* of the word *w* among its *j* synsets.

$$Gloss_{WNet}(c) = \bigcup_{s \in Ancestors(c)} gloss_{WNet}(s) \tag{5}$$

and

$$Gloss_{Wikt}(c) = \bigcup_{s \in Ancestors(c)} gloss_{Wikt}(s) \tag{6}$$

$gloss_{Wikt}(s)$ returns the set of glosses in Wiktionary that correspond to the concept s pertaining to WordNet. In fact, the gloss g_i in Wiktionary corresponds to the gloss g_j in WordNet if they share the maximum of nouns. This choice can be explained by the fact that glosses in both resources are short sentences containing precisely chosen nouns. The nouns used in the definition of a target concept will, therefore, be the same in most cases.

3.1.2 Computing Information Content

Hadj Taieb et al. [35] proposed a method that quantifies the concept's IC based on the subgraph of ancestors. The ancestors' subgraph extracted from the taxonomy "*is a*" represents the propagation of features from a parent to the descendant by adding some specificities.

The IC of a concept is computed based mainly on the contribution of each ancestor according to its depth and descendants' number. The IC value of a concept C is computed as follows:

$$IC(C) = \sum_{c \in Hyper(C)} Score(c) \tag{7}$$

where *Score(c)* indicates the contribution of each ancestor pertaining to the set *Hyper (C)*. This score is computed as follows:

$$Score(c) = \left(\sum_{c' \in DirectHyper(c)} \frac{Depth(c')}{|Hypo(c')|} \right) \times |Hypo(c)| \tag{8}$$

where c and c' are concepts, *DirectHyper(c)* is the set of direct parents of c, and *Hypo (c)* is the set of direct and indirect descendants, including the concept c.

The IC-values are computed using the WNetSS API[5] [36] that offers a great variety of WordNet based semantic similarity measures.

3.2 Disambiguation Process

Figure 3 shows that the target word w_k is surrounded by the words pertaining to its context. Each word w_i is represented by a set of m_i senses in WordNet: $\{s_{i1}, s_{i2}, s_{i3}, \ldots, s_{im_i}\}$ where $i \in [1, n]$, and $i \neq k$ with each sense is defined by a vector of weighted nouns.

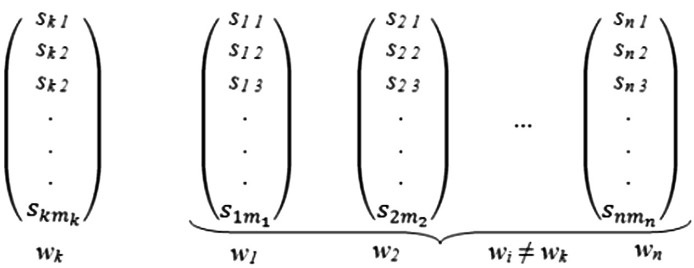


Fig. 3. Representation of the words in the context of w_k with their meanings in WordNet

$SR(s_{kj}, s_{il})$ represents the semantic relatedness between the senses s_{kj} and s_{il} where $1 \leq j \leq m_k$, $1 \leq i \leq n$, $i \neq k$ and $1 \leq l \leq m_i$.

$SR(s_{kj}, s_{il})$ is a function exploited in computing semantic relatedness between two senses s_{kj} and s_{il}. The value is calculated using the cosine measure between the vectors representing each sense (A sense is represented by the weighted words using the IC).

[5] http://wnetss-api.smr-team.org/.

The first step is to calculate, for each sense of the target word, the values of its semantic relatedness with any sense of any word belonging to the context of the word. Figure 4 illustrates all proximity values to calculate. Each vector represents all the values of proximity between a sense s_{kj} of the target word and the meaning of a word belonging to its context.

The next step is to generate all possible combinations between semantic relatedness values obtained in the previous step as it is illustrated in Fig. 4. For each of the obtained combinations, the semantic relatedness values are summed into one value. Each sense of the word w_k is represented by a value. Among the meanings of the target word, the sense chosen as the disambiguating sense is the one assigned to the highest value.

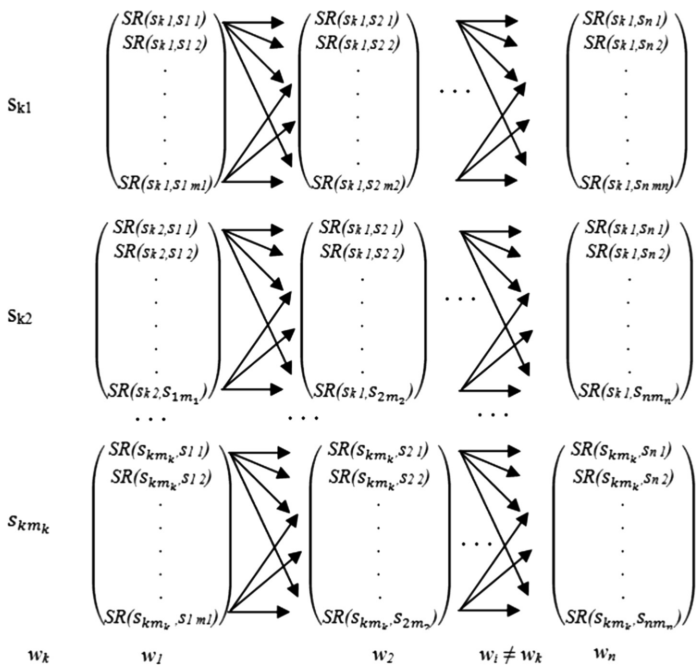


Fig. 4. Representation of different semantic relatedness computed values using the words of the context and the target word w_k.

4 Results and Interpretations

In this section, we introduce all the experiments conducted to evaluate our approach. JSemcor[6] is used for interfacing the *SemCor* and *Senseval* data. Two different datasets are used: Senseval2 and Senseval3. At the beginning, we present the exploited datasets. Subsequently, we present the evaluation metrics. At the end, we detail the results and

6 http://projects.csail.mit.edu/jsemcor/.

their interpretations. For handling WordNet, we exploit the package JWNL (Java WordNet Library) [37] and for Wikitionary the package JWKTL (Java-based Wiktionary Library) [34].

4.1 Datasets and Evaluation Metrics

Several datasets, named as benchmarks are used in the evaluation of systems disambiguation of words.

Senseval2 is a dataset interested in evaluating word sense disambiguation of the meaning of words under the guidance of both associations ACL and SIGLEX. Senseval2 is a set of manually annotated data by experts. This set is annotated with WordNet 1.7 that has been standardized in WordNet 3.0. It consists of three files with 2518 words composed from 475 adjectives, 299 adverbs, 1163 nouns and 581 verbs.

Senseval3 has been annotated by experts using the WordNet. It consists of three files containing 2081 words. These 2081 words are divided into 364 adjectives, 15 adverbs, 951 nouns and 751 verbs.

Navigli [38] stated that a WSD approach can be assessed using the metrics recall, precision and F-measure as follows:

Recall is used to return the percentage of words correctly annotated by the approach relative to the total number of words (of all dataset) to annotate (to disambiguate).

$$Recall = \frac{number\ of\ words\ coorectly\ annotated\ by\ the\ proposed\ apporach}{number\ of\ total\ words\ annotated\ in\ dataset} \tag{9}$$

Precision refers to the measure of the effectiveness of an approach. It is defined as the number of words correctly annotated relative to the total number of words annotated by the approach.

$$\text{Precision} \ = \frac{number\ of\ words\ correctly\ annotated\ by\ proposed\ approach}{number\ of\ words\ annotated\ by\ the\ proposed\ approach} \tag{10}$$

F-measure is based on the combination of both precision and recall measures. It is defined as follows:

$$\text{F-measure} = \frac{2 * recall * precision}{recall + précision} \tag{11}$$

4.2 Results

In the following section, we introduce a comparison of our approach with some other unsupervised approaches that have been evaluated on the same benchmarks that we used. The results are used from the published ones according to the Senseval2 and Senseval3 events.

Table 1. Results of all data senseval2 "*all words*" using W3SD for different *n*

n	Precision	Recall	F-measure
1	0.4715	0.4872	0.4792
2	0.4825	0.50	0.4911
3	0.5283	0.549	0.538
4	0.5625	0.5744	0.5684
5	0.5686	0.6041	0.5856
6	0.5713	0.6062	0.5882

Table 2. Results of all data senseval3 "*all words*" using W3SD for different *n*

n	Precision	Recall	F-measure
1	0.3877	0.3958	0.3917
2	0.4218	0.4369	0.4292
3	0.4439	0.4592	0.4514
4	0.4756	0.4821	0.4788
5	0.4980	0.5082	0.503
6	0.5048	0.5166	0.5106

Table 1 shows the results found for different window context size n for the dataset senseval2. The results are provided using the proposed approach W3SD exploiting the resources WordNet and Wiktionary.

Table 2 shows the results found for different window context size *n* for the dataset senseval3.

Table 3 focuses on the benchmark Senseval2[7]. It shows a comparison of our approach with some works that have been evaluated using the task "*all words*" [39]. The values are sorted in ascending order according to the recall metric. We got the best results for a size of context window (n = 6). The cited approaches, for comparison, are unsupervised methods. IITl, IIT2, and IIT3 are different versions of a semantic tagging system exploiting WordNet examples for sense discriminations. The system uses WordNet relations in identifying examples of words with similar lexical constraints and matching those examples to the context. Susssex system [40] uses automatically acquired selectional preferences to sense tag subject and object head nouns, along with the associated verbal predicates. The preferences include probability distributions over WordNet nouns. These distributions are conditioned on WordNet verb classes.

UNED-AW-U2 [41] is an unsupervised system based on mutual information measured over a large corpus (277 million words) and some additional heuristics.

CL Research-DIMAP [42] is a word-sense disambiguation (WSD) system which is part of the DIMAP dictionary software, designed to use dictionary for unsupervised disambiguation.

[7] www.hipposmond.com/senseval2/Results/all_graphs.xls.

Table 3. Comparison of results of all data senseval2 "*all words*"

System	Precision	Recall	F-measure
IIT 1	0.287	0.033	0.059
IIT 3	0.294	0.034	0.060
IIT 2	0.328	0.038	0.068
Sussex - sel	0.598	0.14	0.226
Sussex - sel-ospd-ana	0.545	0.169	0.257
Sussex - sel-ospd	0.566	0.169	0.260
Universiti Sains Malaysia 2	0.36	0.36	0.36
CL Research-DIMAP	0.451	0.451	0.451
W3SD (WordNet only)	**0.471**	**0.450**	**0.460**
UCLA - gchao	0.5	0.449	0.473
UNED-AW-U2	0.556	0.55	0.552
W3SD (WordNet and Wiktionary)	**0.5713**	**0.6062**	**0.5882**

Table 3 shows that our system has resulted in a higher recall than many other unsupervised systems. The maximum values are obtained by supervised systems. These values are explained by their use of annotated corpus. But the development of this type of corpus is expensive. This represents a disadvantage for supervised approaches. Also, we can note that some approaches have high precision values and in against part their recall values are low such as *IIT* and *Sussex-sel*.

Senseval3 [43] is considered more difficult to disambiguate by [31] due to the high number of verbs which are fine-refined in WordNet. Table 4 displays the recall, precision and F-measure values obtained by our system and a number of other systems that have been evaluated on this benchmark. The values are sorted in ascending order of the recall.

The methods autoPS and autoPSNVs [44] investigate the performance of an unsupervised first sense heuristic where predominant senses are acquired automatically from raw text. IRST [45] tried to refine the domain driven disambiguation that they presented at Senseval-2. The refinements consist of both exploiting the technique domain relevance estimation for domain detection in texts. KNULP system [46] disambiguates senses of a target word in a context by selecting a substituent among WordNet relatives of the target word, such as synonyms, hypernyms, meronyms, etc. The decision is made based on co-occurrence frequency between candidate relatives and each of the context words. UPV system [47] is an unsupervised system based on conceptual density, frequencies of WordNet senses, and WordNet domains.

The problem here is the fine granularity of WordNet which can be resolved through the use of the most frequent senses. Moreover, our proposal shows high accuracy with compound nouns. Tables 3 and 4 show that the exploitation of the Wiktionary glosses leads to a clear enhancement in the performance of the proposed approach W3SD. In fact, for the dataset Senseval2, the precision passes from 0.471 with WordNet to 0.5713 when integrating the Wiktionary.

Table 4. Comparison of results of all data senseval3 "*all words*"

System	Recall	Precision	F-measure
DLSI-UA-all-Nosu	0.275	0.343	0.305
merl.system2	0.352	0.48	0.406
autoPSNVs	0.354	0.563	0.434
autoPS	0.433	0.49	0.459
IRST-DDD-09	0.441	0.729	0.549
upv-unige-CIAOSENSO2-eaw	0.451	0.608	0.517
W3SD (WordNet only)	**0.454**	**0.471**	**0.462**
KUNLP-Eng-ALL	0.496	0.51	0.502
IRST-DDD-LSI	0.496	0.661	0.566
W3SD (WordNet and Wiktionary)	**0.5048**	**0.5166**	**0.5106**
DFA-Unsup-AW	0.546	0.557	0.551
IRST-DDD-00	0.582	0.583	0.582

5 Conclusion and Future Work

This paper presents the W3SD as an approach for word sense disambiguation. The proposal is exploited for determining the correct meaning of an ambiguous word using the words surrounding it in the context of the target word and all relationships provided by the hierarchical structure of WordNet. To determine the meaning of a word to disambiguate, we proposed to use a semantic relatedness measurement which is based on the Informational Content (IC) of concepts. Each sense of a polysemous word is represented by a feature vector, each of which composed of weighted nouns through an IC computing method. The IC computing method quantifies the specificity of a concept according to its topological parameters in the taxonomy «is a». These parameters are the depth, the hyponyms and the graph of ancestors exploited in the quantification process of the IC. These nouns are taken from the glosses of a synset in WordNet and those corresponding to the same synset in Wiktionary. These glosses are extended by those in relationship with them (e.g. hyponymy, hypernymy, holonymy, pertainym, etc.). The cosine measure is then used to calculate the distance between the meaning of the target word and the meaning of words in the same context. The taken sense is the one of the target word leading to the maximum of semantic relatedness between the words pertaining to the context window. W3SD is assessed using the benchmarks Senseval2 and Senseval3 and compared to the unsupervised approaches. Experiments are performed on two steps: firstly, only WordNet is used and secondly, WordNet and Wiktionary are exploited. The results are encouraging and they show an enhancement through the use of Wiktionary. Considering the promising results yielded this work, further studies are needed to the use of other semantic relatedness measures.

References

1. Weaver, W.: Translation. In: Locke, W.N., Boothe, A.D. (eds.) Machine Translation of Languages, pp. 15–23. MIT Press, Cambridge (1949)
2. Brill, E.: Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging. Comput. Linguist. **21**, 543–565 (1995)
3. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database (Language, Speech, and Communication), illustrated edn. The MIT Press, Cambridge (1998)
4. Ben Aouicha, M., Hadj Taieb, M.A., Ibn Marai, H.: WSD-TIC: word sense disambiguation using taxonomic information content. In: Nguyen, N.-T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds.) ICCCI 2016. LNCS (LNAI), vol. 9875, pp. 131–142. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45243-2_12
5. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Sci. Am. **284**, 34–43 (2001)
6. Zhong, Z., Ng, H.T.: Word sense disambiguation improves information retrieval. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Long Papers, vol. 1, pp. 273–282. Association for Computational Linguistics, Jeju Island (2012)
7. Ng, H.T.: Does word sense disambiguation improve information retrieval? In: Proceedings of the Fourth Workshop on Exploiting Semantic Annotations in Information Retrieval, ESAIR 2011, Glasgow, United Kingdom, 28 October 2011, pp. 17–18 (2011)
8. Guyot, J., Falquet, G., Radhouani, S., Benzineb, K.: Analysis of word sense disambiguation-based information retrieval. In: Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-language Evaluation Forum, CLEF 2008, Aarhus, Denmark, 17–19 September 2008, Revised Selected Papers, pp. 146–154 (2008)
9. Sanderson, M.: Word sense disambiguation and information retrieval. In: Croft, B.W., van Rijsbergen, C.J. (eds.) SIGIR 1994, pp. 142–151. Springer, London (1994). https://doi.org/10.1007/978-1-4471-2099-5_15
10. Stokoe, C., Oakes, M.P., Tait, J.: Word sense disambiguation in information retrieval revisited. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 159–166. ACM, Toronto (2003)
11. Carpuat, M., Wu, D.: Word sense disambiguation vs. statistical machine translation. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 387–394. Association for Computational Linguistics, Ann Arbor (2005)
12. Carpuat, M., Wu, D.: How phrase sense disambiguation outperforms word sense disambiguation for statistical machine translation. In: Proceedings of the 11th Conference on Theoretical and Methodological Issues in Machine Translation (TMI) (2007)
13. Neale, S., Gomes, L., Agirre, E., de Lacalle, O.L., Branco, A.: Word sense-aware machine translation: including senses as contextual features for improved translation models. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, 23–28 May 2016 (2016)
14. Liu, Y., Scheuermann, P., Li, X., Zhu, X.: Using WordNet to disambiguate word senses for text classification. In: Shi, Y., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2007. LNCS, vol. 4489, pp. 781–789. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72588-6_127
15. Schadd, F.C., Roos, N.: Word-sense disambiguation for ontology mapping: concept disambiguation using virtual documents and information retrieval techniques. J. Data Semant. **4**, 167–186 (2015)

16. Semeraro, G., Degemmis, M., Lops, P., Basile, P.: Combining learning and word sense disambiguation for intelligent user profiling. In: Twentieth International Joint Conference on Artificial Intelligence, 6–12 January 2007, Hyderabad, India (2007, to appear)
17. Che, W., Liu, T.: Using word sense disambiguation for semantic role labeling. In: 4th International Universal Communication Symposium, IUCS 2010, Beijing, China, 18–19 October 2010, pp. 167–174 (2010)
18. Kulkarni, A., Heilman, M., Eskénazi, M., Callan, J.: Word sense disambiguation for vocabulary learning. In: Proceedings of the 9th International Conference Intelligent Tutoring Systems, ITS 2008, Montreal, Canada, 23–27 June 2008, pp. 500–509 (2008)
19. Moro, A., Raganato, A., Navigli, R.: Entity linking meets word sense disambiguation: a unified approach. TACL **2**, 231–244 (2014)
20. Véronis, J.: HyperLex: lexical cartography for information retrieval. Comput. Speech Lang. **18**, 223–252 (2004)
21. Nameh, M., Fakhrahmad, S.M., Jahromi, M.Z.: A new approach to word sense disambiguation based on context similarity. In: Proceedings of the Workshop of the World Congress on Engineering (2011)
22. Lesk, M.: Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In: Proceedings of the 5th Annual International Conference on Systems Documentation, pp. 24–26. ACM, Toronto (1986)
23. Banerjee, S., Pedersen, T.: Extended gloss overlaps as a measure of semantic relatedness. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence, pp. 805–810. Morgan Kaufmann Publishers Inc., Acapulco (2003)
24. Sinha, R., Mihalcea, R.: Unsupervised graph-based word sense disambiguation using measures of word semantic similarity. In: Proceedings of the International Conference on Semantic Computing, pp. 363–369. IEEE Computer Society, Washington, DC (2007)
25. Jiang, J.J., Conrath, D.W.: Semantic similarity based on corpus statistics and lexical taxonomy. CoRR cmp-lg/9709008 (1997)
26. Leacock, C., Chodorow, M.: Combining local context and WordNet similarity for word sense identification. In: Fellfaum, C. (ed.) WordNet: An Electronic Lexical Database, pp. 265–283. MIT Press, Cambridge (1998)
27. López-Arévalo, I., Sosa, V.J.S., López, F.R., Tello-Leal, E.: Improving selection of synsets from WordNet for domain-specific word sense disambiguation. Comput. Speech Lang. **41**, 128–145 (2017)
28. Panchenko, A., Faralli, S., Ponzetto, S.P., Biemann, C.: Using linked disambiguated distributional networks for word sense disambiguation. In: EACL 2017 Workshop on Sense, Concept and Entity Representations and their Applications: Proceedings of the Workshop, 4 April 2017, Valencia, Spain, pp. 72–78. Association for Computational Linguistics, Stroudsburg (2017)
29. Faralli, S., Panchenko, A., Biemann, C., Ponzetto, S.P.: Linked disambiguated distributional semantic networks. In: Groth, P., Simperl, E., Gray, A., Sabou, M., Krötzsch, M., Lecue, F., Flöck, F., Gil, Y. (eds.) ISWC 2016. LNCS, vol. 9982, pp. 56–64. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46547-0_7
30. Rothe, S., Schütze, H.: AutoExtend: extending word embeddings to embeddings for synsets and lexemes. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, 26–31 July 2015, Beijing, China, Long Papers, vol. 1, pp. 1793–1803 (2015)
31. Hessami, E., Mahmoudi, F., Jadidinejad, A.H.: Unsupervised graph-based word sense disambiguation using lexical relation of WordNet. Int. J. Comput. Sci. Issues **8**(6), 225–230 (2011). No. 3

32. Basile, P., Caputo, A., Semeraro, G.: An enhanced lesk word sense disambiguation algorithm through a distributional semantic model. In: COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, 23–29 August 2014, Dublin, Ireland, pp. 1591–1600 (2014)
33. Navigli, R., Ponzetto, S.P.: BabelNet: the automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artif. Intell. **193**, 217–250 (2012)
34. Zesch, T., Müller, C., Gurevych, I.: Extracting lexical semantic knowledge from Wikipedia and Wiktionary. In: Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May–1 June 2008, Marrakech, Morocco (2008)
35. Hadj Taieb, M.A., Ben Aouicha, M., Ben Hamadou, A.: A new semantic relatedness measurement using WordNet features. Knowl. Inf. Syst. **41**, 467–497 (2014)
36. Ben Aouicha, M., Hadj Taieb, M.A., Ben Hamadou, A.: SISR: system for integrating semantic relatedness and similarity measures. Soft. Comput. **22**, 1855–1879 (2016)
37. Didion, J.: JWNL (Java WordNet Library) (2004)
38. Navigli, R.: Word sense disambiguation: a survey. ACM Comput. Surv. **41**, 10:1–10:69 (2009)
39. Palmer, M., Fellbaum, C., Cotton, S., Delfs, L., Dang, H.T.: English tasks: all-words and verb lexical sample. In: Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems, pp. 21–24. Association for Computational Linguistics, Toulouse (2001)
40. McCarthy, D., Carroll, J., Preiss, J.: Disambiguating noun and verb senses using automatically acquired selectional preferences. In: The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems, pp. 119–122. Association for Computational Linguistics, Toulouse (2001)
41. Fernández-Amorós, D., Gonzalo, J., Verdejo, F.: The UNED systems at Senseval-2. In: The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems, pp. 75–78. Association for Computational Linguistics, Toulouse (2001)
42. Kenneth, C.L.: Use of machine readable dictionaries for word-sense disambiguation in SENSEVAL-2. In: The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems, pp. 75–78. Association for Computational Linguistics, Toulouse (2001)
43. Snyder, B., Palmer, M.: The English all-words task. In: Mihalcea, R., Edmonds, P. (eds.) Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, pp. 41–43. Association for Computational Linguistics, Barcelona (2004)
44. Mccarthy, D., Koeling, R., Weeds, J., Carroll, J.: Using automatically acquired predominant senses for word sense disambiguation. In: Proceedings of the ACL SENSEVAL-3 Workshop, pp. 151–154 (2004)
45. Strapparava, C., Gliozzo, A., Giuliano, C.: Pattern abstraction and term similarity for word sense disambiguation: IRST at Senseval-3. In: Mihalcea, R., Edmonds, P. (eds.) Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, pp. 229–234. Association for Computational Linguistics, Barcelona (2004)
46. Seo, H.-C., Rim, H.-C., Kim, S.-H.: KUNLP system in Senseval-3. In: Mihalcea, R., Edmonds, P. (eds.) Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, pp. 222–225. Association for Computational Linguistics, Barcelona (2004)
47. Buscaldi, D., Rosso, P., Masulli, F.: The upv-unige-CIAOSENSO WSD system. In: Mihalcea, R., Edmonds, P. (eds.) Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, pp. 77–82. Association for Computational Linguistics, Barcelona (2004)

An MP/CP-Based Hybrid Approach to Optimization of the Resource-Constrained Scheduling Problems

Paweł Sitek(✉) and Jarosław Wikarek

Department of Information Systems,
Kielce University of Technology, Kielce, Poland
{sitek, j.wikarek}@tu.kielce.pl

Abstract. Constrained scheduling problems are quite common in project management, manufacturing, distribution, transportation, logistics, supply chain management, software engineering, and computer networks etc. The need to use integer and binary decision variables representing the allocation of different resources to many activities and numerous specific, universal and additional constraints on these decision variables are typical components of the resource-constrained scheduling problems (RCSPs) modeling. It is often necessary to model additional resources and constraints. For these reasons, models are becoming computationally demanding. This is particularly noticeable when methods of operations research (mathematical programing (MP), network programing, and dynamic programming) are used. On the other hand, most RCPSs can be easily modeled as instances of the constraint satisfaction problems (CSPs) and solved using constraint programming (CP) methods. In contrast to the MP-based environment, the CP-based environment methods deal well with binary constraints but worse in optimization. Therefore, the hybrid approach based on integration mathematical programming and constraint logic programming to optimization resource-constrained scheduling problems has been proposed. To evaluate the applicability and computational efficiency of the proposed approach and its implementation programming framework, the illustrative examples of optimization resource-constrained scheduling problems are implemented separately for mathematical programming and hybrid method.

Keywords: Constraint satisfaction problem · Constraint logic programming
Mathematical programming · Resource-constrained scheduling problem
Optimization

1 Introduction

Nowadays in highly competitive marketplace and global context, very often, the decision-making problems include the optimal allocation of different types of constrained resources to activities over the time. This allocation must satisfy a set of different constraints (capacity, precedence, resource, time etc.), where some them may conflict with each other. In the above issue, we can talk about a general form of resource-constrained scheduling problem (RCSP), which covers different organizational

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXIX, LNCS 10840, pp. 144–162, 2018.
https://doi.org/10.1007/978-3-319-90287-6_8

forms and environments such as project and multi-project, flow-shop, job-shop, open-shop, and hybrid environments. The RCSP consists of deciding when, where and how to execute each activity (job). In modern technical and organizational systems, the RCSPs are characterized by numerous additional constraints, often nonlinear, logical, etc. This is a result of technological, business, legal, and security conditions.

As these problems appear very often in many real life decisions in logistic, transportation, supply chain, distribution, manufacturing, project management, computer networks, and construction engineering, it is key issue to find method of flexible modeling and efficient optimization of RCSPs. The methods based on operational research, such as mathematical programming (MP) [1], dynamic programming and network programming seem to be a natural choice, as confirmed by numerous publications [2, 3]. However, the number of decision variables including integer variables in the models and 0–1 allocation decision variables increases computational complexity of the problems and complicate their structure. Moreover, mathematical programming allows modeling only linear and integer constraints [1], whereas complex industrial and management problems may contain constraints with different structures and features usually non-linear, logical, etc.

On the other hand, the most resource-constrained scheduling problems can easily be represented as instances of a Boolean satisfiability problem (SAT) [4] or a constraint satisfaction problem (CSP) [5, 6]. Constraint satisfaction problems (CSPs) are the subject of intense study in artificial intelligence, operations research and soft computing [5]. The result of this study is the development and implementation of CP (Constraint Programming), CCP (Concurrent Constraint Programming) and CLP (Constraint Logic Programming) environments.

CP-based programming environment is very effective for solving binary constraints (binding at most two decision variables). Unfortunately, discrete optimization is not a strong point of CP-based environment.

Based on [5–9] and our earlier studies on the use and integration of MP and CLP for different problems [10–13] some strengths and weaknesses of the aforementioned environments have been observed and detected. It was found that, the integration of CLP and MP can help modeling and solving decision and optimization problems for resource-constrained scheduling which are difficult to solve with one of two methods alone. The proposed integration (called hybridization) has been complemented with model transformation [10–12].

The main contribution as well as the motivation of this study was to apply a hybrid approach to flexible modeling and effective optimization of the resource-constrained scheduling problems. The proposed approach is enriched with the model transformation based on facts used as a presolving method. Moreover, the new combined framework which integrates constraint programming and mathematical programming for optimizing resource-constrained scheduling problems has been developed. A hybrid model for RCSP as a set of CSPs which consist of different types of constraints has been proposed. Finally, the universal structure of the facts of the RCSPs is also presented.

This article is an expanded version of the paper [14] presented at ICCCI 2016. The updates include a comparison of hybrid approach implementation environments

(Sect. 3), a description of fact representation of the problem (Sect. 4) and a new CSP representation of the model (Sect. 5.1). Also, transformation of the model at the fact level has also been added (Sect. 5.1). The number of numerical experiments has been increased to cover instances P8 and P9 with new objective functions (Sect. 5.2).

2 Resource-Constrained Scheduling Problems (RCSPs)

Constrained scheduling processes are common to many different organization, engineering and management areas. Scheduling process answers "*When*", "*What*", "*How*" and/or "*Where*" and it should be done. One of the most important type of scheduling problems is a resource-constrained scheduling problem (RCSP).

This problem in most general form [15–17] is defined as follows. There are a given set of activities (e.g., jobs, tasks, orders, machine operations, etc.) that must be done, a set of main resources (e.g., machines, processors, workstations, etc.) and additional resources (e.g., tools, workers, equipment, memory, etc.) to execute activities, a set of different types of constraints (e.g., allocation, capacity, precedence, time etc.) must be satisfied, and a set of objectives (e.g., makespan, average flow time, average lateness etc.) to evaluate a schedule's execution with (Fig. 1).

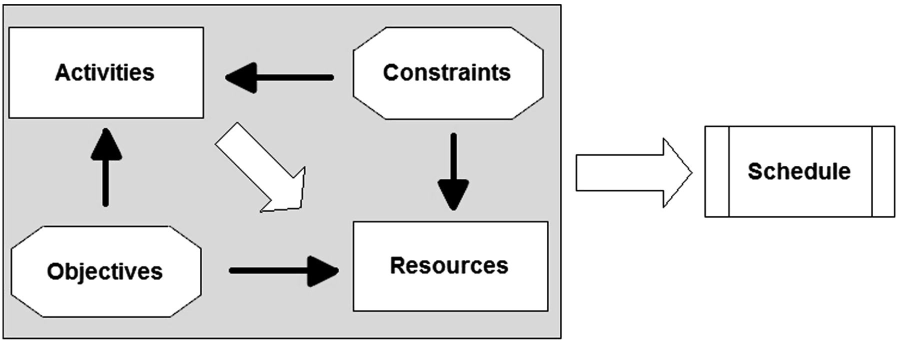


Fig. 1. The components of scheduling process

In real industrial and management problems logical conditions (constraints) may exist, e.g., the condition that two additional resources (workers, tools, etc.) or machines/processors cannot be used simultaneously, which result from legal, technological, business - or safety-related requirements. Two main concerns have to be resolved in RCSP: the choice of the best way to assign available resources (machines/processors/workstations) and additional resources (workers/tools/memory) to the activities (jobs/tasks/orders) within given processing times such that all constraints are satisfied and the choice of the best objective measures (makespan/average flow time). Constraints and objectives are defined during the construction of the model. Constraints define the "feasibility" while objectives define the "optimality" of the generated schedule.

Constraints in scheduling problems appear in many areas and forms. Precedence constraints define the order in which activities can be executed. Temporal constraints limit the times at which resources may be used and/or activities may be done. Resource constraints define their limited availability. There are linear, integer, binary and logic constraints. Hence, it is crucial to find the principles and methods for their efficient modeling and quick resolution.

3 Basic Methods and Principles for Modeling and Optimization RCSPs

3.1 MP-Based Environments

The methods based on operational research, such as mathematical programming (mixed integer linear programming, integer programming, integer linear programming, etc.) and network programming seem to be a natural choice, as confirmed by numerous publications [2, 3] to modeling and solving RCSPs. The most versatile and practical of MP-based formalization is MILP (Mixed Integer Linear Programming) [1]. The MILP formulation is based on algebraic representation of a set of feasible alternative feasible solutions, as well as the objective criterion/goal for comparing alternative solutions. This comparing is realized by [18]:

- Introducing integer decision variables (often 0–1);
- Expressing the criterion/goal as a linear function of decision variables;
- Representing the set of feasible alternatives as the solutions to a conjunction of linear equations and inequalities on decision variables.

For the above reasons, MILP provides a general framework for modeling and solving a large variety of different scheduling problems. However, MILP models remain challenging from a computational standpoint: they are NP-complete or NP-hard. In addition, a significant limitation of MP-based approach is that only linear and integer constraints can be used in the modeling process. There are many solvers on the market that allow you to model and solve MILP models such as LINGO, CPLEX, AMPL, Xpress etc.

3.2 CSP and CP-Based Environments

The most RCSPs can easily be modelled as instances of the constraint satisfaction problems (CSPs). CSP is defined as follows: given a set of decision variables, a set of possible values (decision variable domain) for each decision variable, and a set of different constraints between the decision variables, assign a value to each decision variable, so that all the constraints are satisfied. CSPs are typically determined on a constraint network (Fig. 2). Formally, a CSP is defined as a triple $(X_i, D^i_{om}\ C^i_{st}$, where X_i is a set of decision variables, D^i_{om} is a domain of values of decision variables, and C^i_{st} is a set of constraints. The solving of CSP based on the specific deductions that result in the removal of values from the decision variables domains (called domain reductions). The set of values in the domain of a decision variable that are not invalidated by constraint propagation is called the current domain (indomain) of that decision variable. The mechanisms of domain reduction and constraint propagation (both method are incomplete) be applied to reduce the search space of decision variables. However, while this may determine whether a model is infeasible, it does not necessarily find solutions to the model. In such situations the backtracking mechanism must be used [5]. Unlike previous methods, backtrack is the fundamental 'complete' search method for CSPs, in this sense that one is guaranteed to find a solution if only exists. This search process is often represented as a search tree, where each node (below the root) represents a choice of a value for a decision variable, and each branch represents a candidate partial solution. Backtracking methods are unfortunately time-consuming and resource-consuming. Therefore, in many cases, makes it difficult or impossible to solve the CSP at an acceptable time. This is due to the nature of the CSP solving algorithm (Fig. 3), which in extreme cases with large variable domains can generate a very large number of backtrackings.

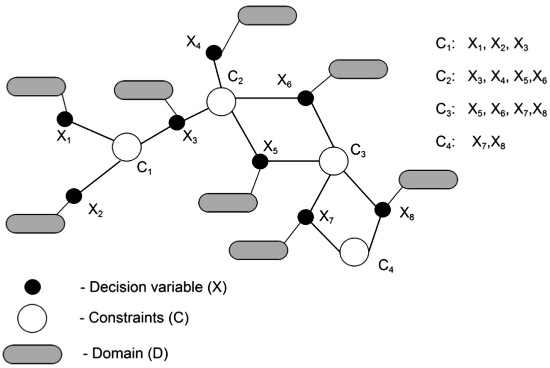


Fig. 2. Constraint network in CSP

```
CSP_solver (Problem_i)
  Create CSP_i For Problem_i
while not(solved or infeasible) do
Remove_Inconsistent_Values   } %Constraint Propagation method
Select_Decision_Variable     }
Select_Value_For_Variable    } %Distribution variables
end
                                                    ↑ backtracking
```

Fig. 3. CSP solution algorithm

The most popular and promising environments based on CSPs are CP (Constraint Programming), CCP (Concurrent Constraint Programming) and CLP (Constraint Logic Programming). In the above environments, the set of constraints is not only used to test the validity of a feasible solution, as in conventional mathematical programming

(Sect. 3.1), but also in an active mode: constraints are used to remove values from the decision variables domains and detect inconsistencies.

CP-based environments can be defined as programming methods based on the following main assumptions [5]:

- The problem to be solved is explicitly modelled in terms of decision variables and constraints over these variables.
- In a CP-based declarative program, the explicit problem definition is clearly separated from the method used to solve the problem.
- The modelled problem (description, decisions and objectives) are translated into a set of constraints.
- A deductive process referred to as "constraint propagation" is used to propagate the impacts and influences of the constraints.
- Constraint propagation is applied each time a new decision is made, and this process also is clearly separated from the solving method/algorithm.

3.3 Comparison MP-Based and CP-Based Approaches in the Context of Modeling and Solving RCSPs

Although developed in different communities with different roots and perspectives, CP-based (CLP, CP) and MP-based (MILP, ILP) methods share the same fundamental principles: search, inference and relaxation, though implemented in different ways and contexts. These principles guide to opportunities for integration. Table 1 shows the strengths and weaknesses of both approaches in the context of modeling and solving of scheduling problems.

Table 1. Advantages and disadvantages of both environments in the context of modeling and solving scheduling problems

MP (MILP)	CLP
The main advantages	
Relaxation methods - very effective when constraints or objective function contain many variables (very common in RCSPs)	The more constraints, the better. But just having a specific structure, e.g. binary constraints
Tools for filtering. Duality theory	Modeling easy
Numerical calculations	Build in methods for removes infeasible values from variable domains (filtering, domain consistency maintenance, constraint propagation)
The main disadvantages	
Only linear constraints	Lacks relaxation technology - It can be critical when solving large-scale problems
Artificial methods of modeling constraints allocation by the introduction of 0–1 variables	Weaker for continuous variable (due to lack of numerical techniques)
Not very high efficiency of the combinatorial problems with large sizes	May fail when constraints contain many variables. These constraints don't propagate well

The weaknesses on both sides are largely eliminated by the proposed CLP/MP integration through the introduction of unique problem modeling procedures and model transformations capable of reducing the solution space.

4 A Hybrid Approach to Modelling and Optimization RCSPs - the Concept and Implementation Programming Framework (IPF)

The proposed hybrid approach to modeling and solving (optimization) RCSPs is able to fill the gaps and eliminate the shortcomings that occur in both MP-based and CP-based approaches when they are used separately for such problems (Table 1).

The presented approach is based on the property that the definition of the problem is clearly separated from the methods used to solve it. A given problem is modeled with the use of CLP as a set of CSPs (Sect. 3.2), than pre-solved using model transformation method and the end solved/optimized with the use of MP (MILP).

One of the characteristic features of the hybrid approach is that the problem is represented in the form of facts (facts-based representation). Decision and optimization problems are typically written as algebraic expressions with coefficient matrices and vectors of decision variables, as in mathematical programming. The new representation (facts-based) reduces the size of the problem. This reduction can be referred to as "hidden" transformation resulting only from the change in problem representation. With algebraic notation, the status is usually shown as a matrix of coefficients and the decisions as a vector of decision variables. If a status is unacceptable, its corresponding parameter is "0". Then the corresponding decision parameters are zeroed but still present in the algebraic model representation. This increases the size of the model, thereby increasing the efforts necessary for solving it, and in extreme cases, leading to the situation in which the model is too big to be solved by a given MP solver. The change of representation (from algebraic to facts-based) provides the same effect "for free", with no need to use advanced methods or algorithms. The change of representation is shown symbolically in Fig. 4.

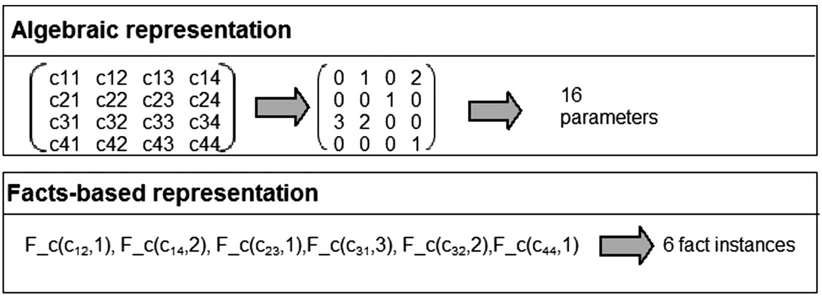


Fig. 4. Algebraic representation versus fact-based representation.

To support and implementation the concept of proposed approach to resource-constrained scheduling problems, an implementation programming framework (IPF) was proposed (Fig. 5).

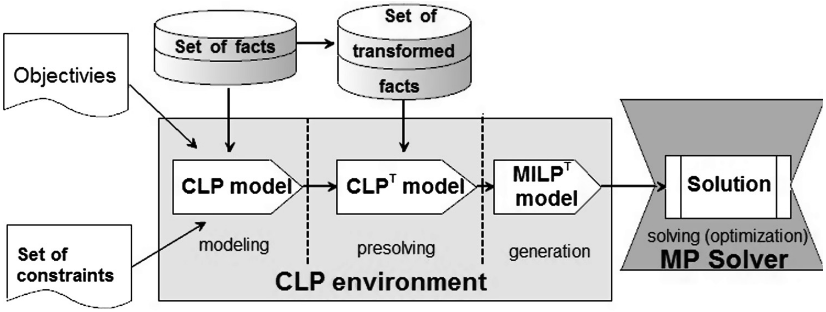


Fig. 5. Conceptual diagram of the implementation of hybrid approach in the form of IPF (Implementation Programming Framework)

The fact-based representation was used to propose another, "non-hidden", transformation of the problem. The transformation reduced the number of decision variables and constraints of the modeled problem, which results in the reduction and transformation of solution space relative to the model prior to transformation. It is based on the analysis of the instances of the facts that represent the problem (Sect. 5.1). In short, it involves the removal of unacceptable points form solution space before using constraint propagation.

The important assumptions of the IPF include the following. Specification of the problem (CLP model) can be stated as the sets of integrity, binary, linear, logic and symbolic constraints that have been grouped into CSPs. Additionally, in the case of optimization, the model may include objectives. Data instances of the problem are stored as sets of facts. Two environments, i.e. CLP and MP are integrated. The implementation framework supports the model transformation [12, 13] as a presolving method. The transformation changes the representation of the problem using data instances (facts) and the characteristics and possibilities of CLP. The result of the transformation is to reduce the number of decision variables and constraints of the problem, which is particularly important and useful for optimization because it reduces the search space. After the transformation, a CLPT model is obtained (the model size is reduced). On the basis of the CLPT model, a MILPT (mixed integer linear programming) [1] model after transformation is automatically generated and solved in the MP environment.

The schema of implementation programming framework for the hybrid approach has been shown in Fig. 5. Using this framework, you can solve any resource-constrained scheduling problems. From a variety of tools for the implementation of the CLP part in the framework the ECL^{i}PS^e software [19] is used. The area of mathematical programming (MP) in the framework was implemented using the SCIP [20].

4.1 Modeling in Hybrid Approach

The declarative CLP model constitutes a natural form of knowledge representation. It specifies the state of the modeled problem defined by an assignment of values to some or all of the decision variables without violating any constraints. This means that each form of a CSP can be interpreted as the representation of a certain portion of knowledge of the modeled problem. The question is whether CSPs are sufficient to describe the problem. It seems that the answer is "yes" in many cases, because according to the maxim, if something is not forbidden, it is permitted. Nevertheless, many of real-life problems, including scheduling problems, require that the manner and the goal of constraint satisfaction be defined (objectives). Therefore, architecture of the CSP-based models is flexible and diverse as shown in Fig. 6. Individual CSPs can be combined in any way and, just as importantly, some CSPs can be deleted or added, depending on the situation and data (facts).

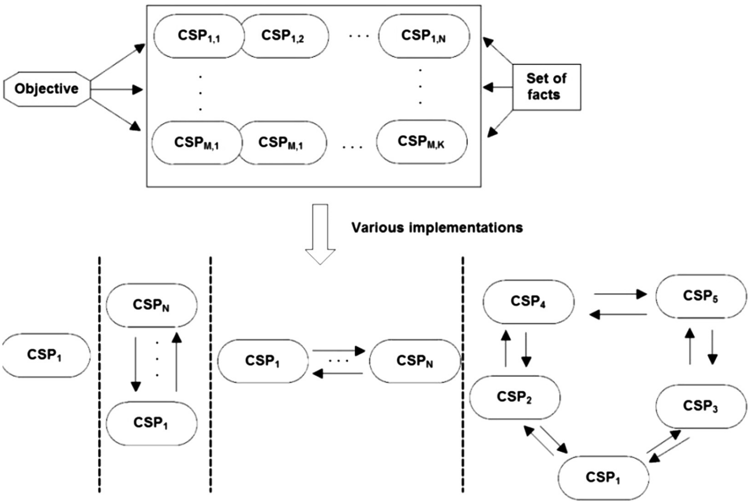


Fig. 6. Sample architectures of models that are a set/collection of CSPs

5 Illustrative Example – Implementation and Computational Experiments

The illustrative example shows a scheduling problem with additional resources (employees, tools, etc.) in job-shop system. Job-shop scheduling has been extensively studied in the literature reporting problems of varied degree of difficulty and complexity [15–17]. To evaluate the presented implementation programming framework for decision support and optimization, a number of numerical experiments were performed for the illustrative example (Sect. 5.2). The formal mathematical model for illustrative example in the form of indices, decision variables, and the main constraints (o2)..(o12), are shown in Table B1 (Appendix B). The model introduced additional constraints specifying the logical simultaneous use of machines (o13) and additional resources (o14). Three objective functions have been proposed for the illustrative example: to minimize the makespan (f1), to minimize the cost of using additional resources (f2) and to minimize the frequency of use of selected additional resources (f3).

5.1 Implementation of the Illustrative Example

The illustrative example was modeled and implemented in two environments: MP (SCIP) and IPF (SCIP and Eclipse). In the first scenario, all constraints and objective functions were directly implemented in the MP solver environment. In the second scenario, the problem was modeled using the set of CSPs with architecture as in Fig. 7. The CSPs in the model are described in Table 2. Subsequent CSPs cover different areas of the problem being modeled (see description under Table 2).

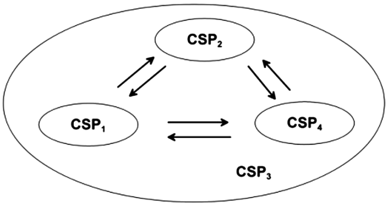


Fig. 7. Architecture for CLP model of illustrative example.

Table 2. Model CLP - the set of CSPs

CSP	Description
CSP_1	$CSP_1 = (C = \{c2, c8, c9\}, X = \{G_{j,f}, C_{max}, X_{j,f,a,t}\}, D = \{D^1_G, D^1_C, D^1_X\})$
CSP_2	$CSP_2 = (C = \{c3, c4, c5, c6, c7\}, X = \{X_{j,f,a,t}, Z_{j,f,a,t}\}, D = \{D^2_X, D^2_Z\})$
CSP_3	$CSP_3 = (C = \{c10, c11, c12\}, X = \{X_{j,f,a,t}, Z_{j,f,a,t}\}, D = \{D^3_X, D^3_Z\})$
CSP_4	$CSP_4 = (C = \{c13, c14\}, X = \{X_{j,f,a,t}\}, D = \{D^4_X\})$

CSP_1 - represents the scheduling time, CSP_2 - represents the occupancy of machinery and additional resources, CSP_3 - represents binarity and integrity of variables, CSP_4 - represents logic constraints (exclusion of simultaneous execution/use).

Knowledge about the problem was represented in the form of facts. The structure of the facts and their relationships for the illustrative example are shown in Fig. 8. Figures 8 and 9 illustrate the fact transformation technique, being the basis for the transformation of the model.

The facts are transformed as follows. Analysis of the set of instances of fact *Or_Fa (#J,..)* is the basis for establishing a set of attributes *#J* intended for realization. The instances related only to the elements of this set remain in facts *T_Fa(#J,#F,..), AL_Fa (#A,#F,#J,..), J_Fa(#J,#F1,#F2,..)*. In practice, this corresponds to the situation in which if product/job with an index from the *#J* set is not performed, information about

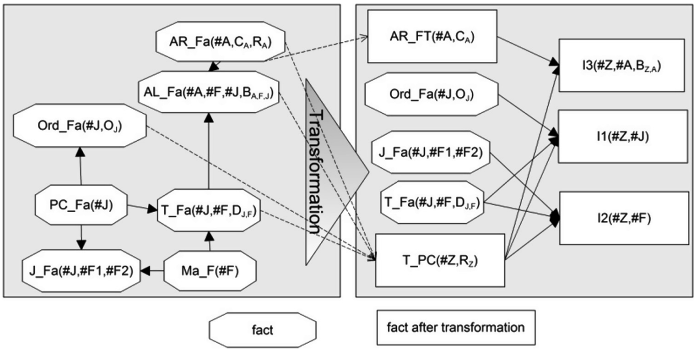


Fig. 8. The set of facts before and after transformation for illustrative example (# - ID of the fact, B - the number of allocated additional resources, C - limit of additional resources, D - execution time, O - order quantity, R - cost of use additional resources).

technologies, additional resources, etc. relevant for its realization are not needed. The *AL_Fa(#A,#F,#J,..)* is transformed in the next step. The transformation involves aggregating its key attributes, that is, replacing attributes *#A,#F,#J* with attribute *#Z* defining the so-called implementation. A new implementation fact is created, *T_PC(#Z, …),* as a basis for creating the so-called "existence" facts *I1(#Z,#J), I2(#Z,#F), I3 (#Z, #A*) determining respectively whether given implementation with attribute *#Z* refers to a product with attribute *#J*, uses a machine with attribute *#F* and needs an additional resource with attribute *#A*. Symbolic illustration of the processes of transformation and aggregation of facts and their attributes is shown in Fig. 9 The fact *AR_Fa(#A,…),* which will be replaced by *AR_FT(#A,…)* without attribute change will also be transformed. Transformation of facts means that the whole model with its variable, constraints, etc. is transformed.

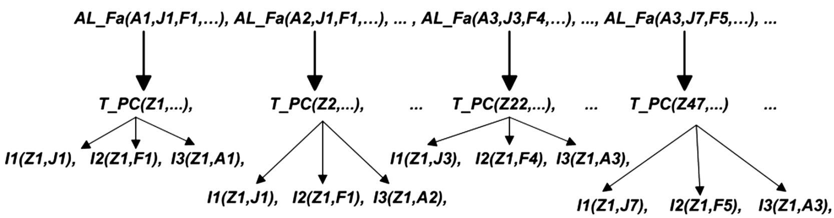


Fig. 9. Transformation technique for selected facts in the illustrative example.

5.2 Numerical Experiments

The numerical experiments are performed for the data instance consisting of twelve products/jobs (J1..J12), six machines (F1..F6), and four (A1..A4) different sets of the additional resources (workers with different skills). The set of facts for illustrative example is shown in Appendix A.

All numerical experiments were performed using an ordinary PC, Intel core (TM4), i-3, 3.1 GHz, and 6 GB RAM. At the beginning of the study, efficiency of the presented constraint-based implementation framework was evaluated relative to the MP-based implementations ($P_1..P_3$) for illustrative model (o2)..(o12), (f1). These experiments were conducted in the context of calculation time and the number of constraints and decision variables. Numerous experiments were performed with varied parameter C_a (additional resources). In the second phase of the study, the experiments were performed for hybrid model (o2)..(o15), (f1) with logical constraints ($P_4..P_7$). In these examples the two additional resources (P_6, P_7) or machines (P_4, P_5) cannot be used simultaneously. In the final stage of experiments, other objective functions were used (f2), (f3). For examples P_{8a} and P_{8b}, minimal costs of the additional resources with varied parameter C_a (additional resources) were found (f2). In other experiments (P_{9a}, P_{9b}), the frequency of use of additional resource C_3 at time T was optimized. The obtained results are presented in Table 3 and the corresponding schedules for examples P_1, P_3, P_4, P_7, P_{8b}, and P_{9b} in Figs. 10, 11, 12, 13, 14, 15, 16 and 17.

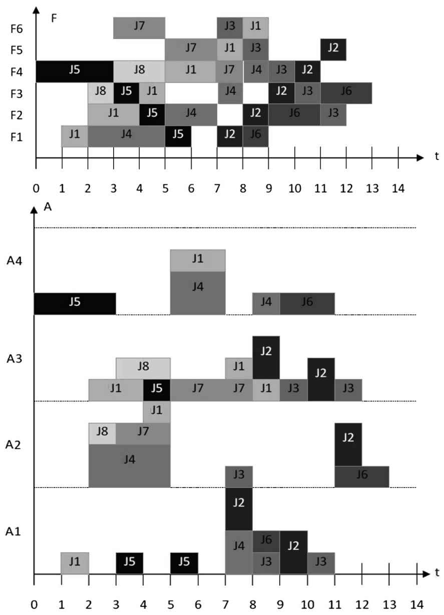


Fig. 10. Schedules (Gantt charts) for machines (top) and additional resources (down) corresponding to the example P_1 for $C_1 = C_2 = C_3 = C_4 = 10$

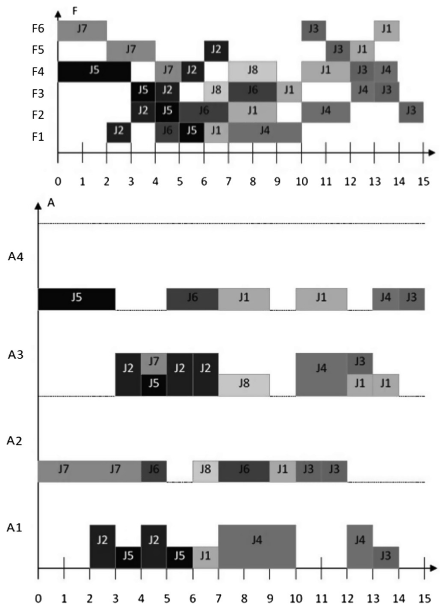


Fig. 11. Schedules (Gantt charts) for machines (top) and additional resources (down) corresponding to the example P_3 for $C_1 = C_3 = 4$ $C_2 = C_4 = 2$

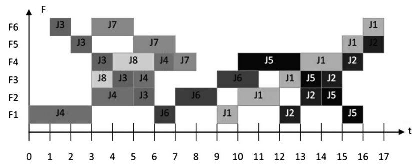


Fig. 12. Schedule (Gantt chart) for machines corresponding to example P_4 for $C_1 = C_2 = C_3 = C_4 = 10$

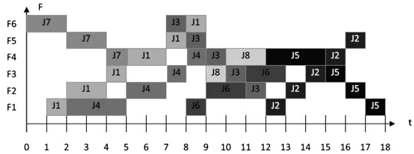


Fig. 13. Schedule (Gantt chart) for machines corresponding to example P_{8b} for $C_1 = C_3 = 4$ $C_1 = C_4 = 2$

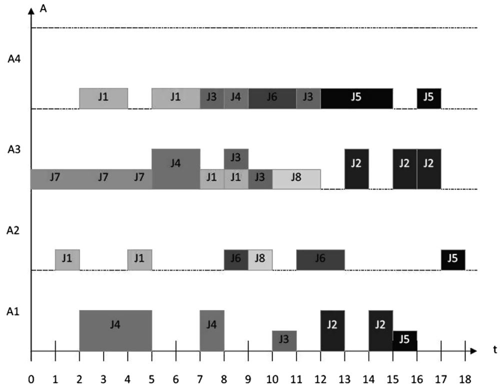


Fig. 14. Schedule (Gantt chart) for additional resource corresponding to example P_{8b}

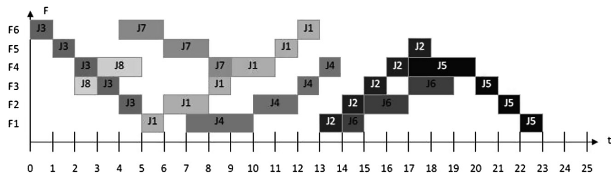


Fig. 15. Schedule (Gantt chart) for additional resource corresponding to example P_{9b}.

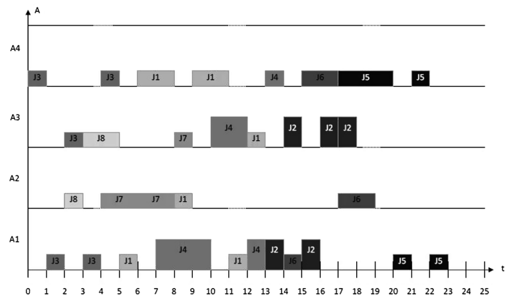


Fig. 16. Schedule (Gantt chart) for machines corresponding to example P_{9b} for $c_3 = 4$ $c_2 = c_4 = 2$ $T = 25$.

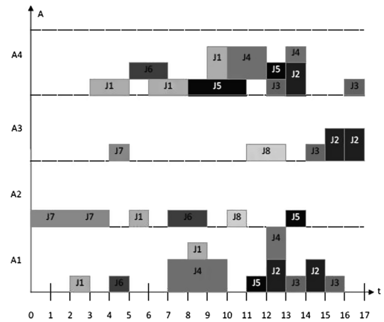


Fig. 17. Schedule (Gantt chart) for additional resource corresponding to example P_7

Table 3. Results for numerical experiments at different values of set C_a (P_1, P_2, P_3, P_8, P_9) and logic constraints (P_4..P_5)

Pn	Parameters	MILP (MP)				$MILP^T$ (Implementation framework)			
		V	Co	T	Fc	V	Co	T	Fc
P_1	$c_1 = c_2 = c_3 = c_4 = 10$	7346	4034	345	$C_{max} = 13$	1991	1377	29	$C_{max} = 13$
P_2	$c_1 = c_2 = c_3 = c_4 = 4$	7346	4034	456	$C_{max} = 14$	1991	1377	38	$C_{max} = 14$
P_3	$c_1 = c_3 = 4$ $c_2 = c_4 = 2$	7346	4034	462	$C_{max} = 15$	1991	1377	39	$C_{max} = 15$
P_4	Exclusion F1 & F2	—	—	—	—	2105	1511	42	$C_{max} = 17$
P_5	Exclusion F1 & F3	—	—	—	—	2105	1511	42	$C_{max} = 16$
P_6	Exclusion A1 & A2	—	—	—	—	2181	1607	52	$C_{max} = 16$
P_7	Exclusion A2 & A3	—	—	—	—	2181	1607	48	$C_{max} = 17$
P_{8a}	$c_1 = c_3 = 4$ $c_2 = c_4 = 2$ T = 15	7346	4034	345	Wp = 1040	1991	1377	28	Wp = 1040
P_{8b}	$c_1 = c_3 = 4$ $c_2 = c_4 = 2$ T = 20	7346	4034	452	Wp = 980	1991	1377	31	Wp = 980
P_{9a}	$c_3 = 4$ $c_2 = c_4 = 2$ T = 16	7346	4034	365	$U_3 = 24$	1991	1377	28	$U_3 = 24$
P_{9b}	$c_3 = 4$ $c_2 = c_4 = 2$ T = 25	7346	4034	354	$U_3 = 40$	1991	1377	25	$U_3 = 40$

V - the number of decision variables, Co - the number of constraints, T - calculation time, Fc - objective function

6 Conclusion

The proposed hybrid approach to the modeling and optimization RCSPs is characterized by high flexibility and efficiency. Efficiency is the result of integrating two environments (CLP and MP) and introducing methodology based on facts to transformation of the model. This transformation is using as a presolving method. The use of transformations reduced the solution space for modeled problem. The proposed approach implemented for the illustrative example resulted in more than ten-fold search time reduction compared to the MP-based approach (Table 3). Also, the number of variables and constraints were reduced more than three and four times the initial number. The size of the search space was reduced to consequently enable addressing larger size problems within acceptable time limits. Moreover, the proposed constraint-based approach allows extremely flexible modeling of various types of problems with different types of constraints. Therefore, the model may contain logical constraints. The introduction of logical constraints into the MP-based approach is virtually impossible or requires a lot of effort (e.g. it would require the introduction and modeling of many additional constraints). The optimization model proposed makes it possible to find optimal scheduling and to allocate the resources and the demand to them in many different situations (Table 3).

Further studies will include implementation of various types of other scheduling problems in the area of flexible manufacturing, transport systems, capacity vehicle routing problems [21], multi-project scheduling problems [4], resource capacity project scheduling, and assembly line balancing problem [22], scheduling of mobile robots [23] and multi-assortment repetitive production. The hybrid implementation framework will also be supplemented by the methods of Fuzzy Logic and Answer Set Programming [24]. It is also planned to expand the hybrid concept by introducing heuristics and metaheuristics instead of MP.

Appendix A. Sets of Facts for Illustrative Example

```
%Ma_F(#F). – machine/processor
 Ma_F(F1). Ma_F(F2). Ma_F(F3). Ma_F(F4). Ma_F(F5). Ma_F(F6).
%pr_F(#J). – product/job
 J_Fa(J1). J_Fa(J2). J_Fa(J3). J_Fa(J4).  J_Fa(J5).  J_Fa(J6).
 J_Fa(J7). J_Fa(J8). J_Fa(J9). J_Fa(J10). J_Fa(J11). J_Fa(J12).
%T_Fa(#J,#F,D). – technology
 T_Fa(J1,F1,1).  T_Fa(J1,F2,2).  T_Fa(J1,F3,1). T_Fa(J1,F4,2).
 T_Fa(J1,F5,1).  T_Fa(J1,F6,1).  T_Fa(J2,F1,1). T_Fa(J2,F2,1).
 T_Fa(J2,F3,1).  T_Fa(J2,F4,1).  T_Fa(J2,F5,1). T_Fa(J3,F2,1).
 T_Fa(J3,F3,1).  T_Fa(J3,F4,1).  T_Fa(J3,F5,1). T_Fa(J3,F6,1).
 T_Fa(J4,F1,3).  T_Fa(J4,F2,2).  T_Fa(J4,F3,1). T_Fa(J4,F4,1).
 T_Fa(J5,F1,1).  T_Fa(J5,F2,1).  T_Fa(J5,F3,1). T_Fa(J5,F4,3).
 T_Fa(J6,F1,1).  T_Fa(J6,F2,2).  T_Fa(J6,F3,2). T_Fa(J7,F4,1).
 T_Fa(J7,F5,2).  T_Fa(J7,F6,2).  T_Fa(J8,F3,1). T_Fa(J8,F4,2).
 T_Fa(J9,F1,1).  T_Fa(J9,F2,2).  T_Fa(J9,F3,1). T_Fa(J9,F4,2).
 T_Fa(J10,F4,1). T_Fa(J10,F5,1). T_Fa(J10,F6,1).
 T_Fa(J11,F3,1). T_Fa(J11,F4,1). T_Fa(J11,F5,1).
 T_Fa(J12,F1,1). T_Fa(J12,F2,1). T_Fa(J12,F3,1).
%AR_Fa (#A,C,R). – additional resources
 AR_Fa(A1,20,40). AR_Fa(A2,20,30). AR_Fa(A3,20,20).
 AR_Fa(A4,20,10).
%AL_Fa(#A,#F,#J,B) - allocation
 AL_Fa(A1,J1,F1,2). AL_Fa(A2,J1,F1,2). AL_Fa(A3,J1,F2,2).
 AL_Fa(A4,J1,F2,2). AL_Fa(A2,J1,F3,2). AL_Fa(A4,J1,F4,2).
 AL_Fa(A1,J1,F5,2). AL_Fa(A3,J1,F5,2). AL_Fa(A3,J1,F6,2).
 AL_Fa(A4,J1,F6,4). AL_Fa(A1,J2,F1,4). AL_Fa(A2,J2,F1,4).
 AL_Fa(A3,J2,F2,4). AL_Fa(A4,J2,F2,4). AL_Fa(A1,J2,F3,4).
 AL_Fa(A3,J2,F4,4). AL_Fa(A2,J2,F5,4). AL_Fa(A3,J2,F5,4).
 AL_Fa(A3,J3,F2,2). AL_Fa(A4,J3,F2,2). AL_Fa(A1,J3,F3,2).
 AL_Fa(A3,J3,F4,2). AL_Fa(A1,J3,F5,2). AL_Fa(A2,J3,F5,2).
```

```
  AL_Fa(A3,J3,F5,2). AL_Fa(A2,J3,F6,2). AL_Fa(A4,J3,F6,2).
  AL_Fa(A1,J4,F1,4). AL_Fa(A2,J4,F1,4). AL_Fa(A3,J4,F2,4).
  AL_Fa(A4,J4,F2,4). AL_Fa(A1,J4,F3,4). AL_Fa(A4,J4,F4,2).
  AL_Fa(A1,J5,F1,2). AL_Fa(A2,J5,F1,2). AL_Fa(A3,J5,F2,2).
  AL_Fa(A4,J5,F2,2). AL_Fa(A1,J5,F3,2). AL_Fa(A4,J5,F4,2).
  AL_Fa(A1,J6,F1,2). AL_Fa(A2,J6,F1,2). AL_Fa(A3,J6,F2,2).
  AL_Fa(A4,J6,F2,2). AL_Fa(A2,J6,F3,2). AL_Fa(A3,J7,F4,2).
  AL_Fa(A2,J7,F5,2). AL_Fa(A3,J7,F5,2). AL_Fa(A2,J7,F6,2).
  AL_Fa(A3,J7,F6,2). AL_Fa(A2,J8,F3,2). AL_Fa(A3,J8,F4,2).
  AL_Fa(A1,J9,F1,2). AL_Fa(A2,J9,F1,1). AL_Fa(A3,J9,F2,2).
  AL_Fa(A4,J9,F2,2). AL_Fa(A3,J9,F3,1). AL_Fa(A4,J9,F4,2).
  AL_Fa(A1,J10,F4,2). AL_Fa(A3,J10,F4,1). AL_Fa(A3,J10,F5,1).
  AL_Fa(A1,J10,F6,2). AL_Fa(A2,J10,F6,1). AL_Fa(A1,J11,F3,2).
  AL_Fa(A1,J11,F3,1). AL_Fa(A3,J11,F4,1). AL_Fa(A4,J11,F5,1).
  AL_Fa(A1,J12,F1,2). AL_Fa(A2,J12,F2,2). AL_Fa(A4,J12,F3,1).
%Fc_Fa(#J,#F1,#F2). - Frecedence
  Fc_Fa(J1,F1,F2). Fc_Fa(J1,F2,F3). Fc_Fa(J1,F3,F4).
  Fc_Fa(J1,F4,F5). Fc_Fa(J1,F5,F6). Fc_Fa(J2,F1,F2).
  Fc_Fa(J2,F2,F3). Fc_Fa(J2,F3,F4). Fc_Fa(J2,F4,F5).
  Fc_Fa(J3,F6,F5). Fc_Fa(J3,F5,F4). Fc_Fa(J3,F4,F3).
  Fc_Fa(J3,F3,F2). Fc_Fa(J4,F1,F2). Fc_Fa(J4,F2,F3).
  Fc_Fa(J4,F3,F4). Fc_Fa(J5,F4,F3). Fc_Fa(J5,F3,F2).
  Fc_Fa(J5,F2,F1). Fc_Fa(J6,F1,F2). Fc_Fa(J6,F2,F3).
  Fc_Fa(J7,F6,F5). Fc_Fa(J7,F5,F4). Fc_Fa(J8,F3,F4).
  Fc_Fa(J9,F1,F2). Fc_Fa(J9,F2,F3). Fc_Fa(J9,F3,F4).
  Fc_Fa(J10,F4,F5). Fc_Fa(J10,F5,F6). Fc_Fa(J11,F3,F4).
  Fc_Fa(J11,F4,F5). Fc_Fa(J12,F1,F2). Fc_Fa(J12,F2,F3).
Ord_Fa(#J,O).
  Ord_Fa(J1,1). Ord_Fa(J2,1). Ord_Fa(J3,1). Ord_Fa(J4,1).
  Ord_Fa(J5,1). Ord_Fa(J6,1). Ord_Fa(J7,1). Ord_Fa(J8,1).
```

Appendix B. Illustrative Example-Formal Model

Table B1. Indices, parameters and constraints for mathematical model of RCSP

Indices	
f	machine/workstation/processor/$F = 1..F$
j	product/job/item type $j = 1..J$
a	additional resource (i.e. worker/tool/memory) $a = 1..A$
t	period $t = 1..T$

(*continued*)

Table B1. (*continued*)

Indices	
Parameters	
o_j	order quantity for product/job/item *j*
$d_{j,f}$	the time required to make a product/job/item *j* on the machine/processor *f*
$d1_{j,f}$	if the product/job/item *j* is made on the machine/processor *f* than $d1_{j,f} = 1$, otherwise $d1_{j,f} = 0$
c_a	the number of additional resources *a*
r_a	the cost of use of additional resource *a* over a unit of time
$bu_{j,f,a}$	if additional resource *a* is required to make a product/job/item *j* on the machine/processor *f* $bu_{j,f,a} = 1$, otherwise $bu_{j,f,a} = 0$
$b_{j,f,a}$	the number of additional resources *a* need to make the product/job/item *j* on the machine/processor *f*
$ko_{j,f1,f2}$	If the product *j* is made on the machine/processor *f1* before on the machine/processor *f2* than $ko_{j,f1,f2} = 1$, otherwise $ko_{j,f1,f2} = 0$
e_t	parameters for conversion numbers of periods *t* for the variables $e_t = t$
Decision Variables	
$X_{j,f,a,t}$	if the worker/tool *a* in period *t* makes the product *j* on the machine/processor machine *f* than $X_{j,f,a,t} = 1$ otherwise $X_{j,f,a,t} = 0$
$G_{j,f}$	the number of last period in which the product/job/item *j* is made on the machine/processor *f*
$Z_{j,f,a,t}$	if the period *t* is the latest in which the worker/tool *a* makes the product *j* on the machine *f* than $Z_{j,f,a,t} = 1$ otherwise $Z_{j,f,a,t} = 0$
C_{max}	makespan

$$\mathrm{Min}\, C_{max} \tag{f1}$$

$$G_{j,f} \le C_{max} \forall j = 1..J, f = 1..F \tag{o2}$$

$$\sum_{a=1}^{A} \sum_{t=1}^{T} d1_{j,f} \cdot bu_{j,f,a} \cdot X_{j,f,a,t} = d_{j,f} \cdot o_j \forall j = 1..J, f = 1..F \tag{o3}$$

$$\sum_{j=1}^{J} \sum_{a=1}^{A} X_{j,f,a,t} \le 1 \forall f = 1..F, t = 1..T \tag{o4}$$

$$\sum_{j=1}^{J} \sum_{f=1}^{F} b_{j,f,a} \cdot X_{j,f,a,t} \le c_a \forall a = 1..A, t = 1..T \tag{o5}$$

$$X_{j,f,a,t-1} - X_{j,f,a,t} \le Z_{j,f,a,t-1} \forall j = 1..J, f = 1..F, a = 1..A, t = 2..T \tag{o6}$$

$$\sum_{t=1}^{T} Z_{j,f,a,t} = 1 \forall j = 1..J, f = 1..F, a = 1..A \tag{o7}$$

$$G_{j,f} \geq e_t \cdot X_{j,f,a,t} \forall j = 1..J, f = 1..F, a = 1..A, t = 1..T \tag{o8}$$

$$G_{j,f2} - b_{j,f2} \geq G_{j,f1} \forall j = 1..J, f1, f2 = 1..F : ko_{j,f1,f2} = 1 \tag{o9}$$

$$G_{j,f} \in C \forall j = 1..J, f = 1..F \tag{o10}$$

$$X_{j,f,a,t} \in \{0, 1\} \forall u = j..J, f = 1..F, a = 1..A, t = 1..T \tag{o11}$$

$$Z_{j,f,a,t} \in \{0, 1\} \forall j = 1..J, f = 1..F, a = 1..A, t = 1..T \tag{o12}$$

$$\text{Exclusion\_M}(f1, f2) \exists f1, f2 = 1..F : f1 \neq f2 \tag{o13}$$

$$\text{Exclusion\_R}(a1, a2) \exists a1, a2 = 1..A : a1 \neq a2 \tag{o14}$$

$$\min \sum_{j=1}^{J} \sum_{f=1}^{F} \sum_{a=1}^{A} \sum_{t=1}^{T} (X_{j,f,a,t} \cdot r_a) \tag{f2}$$

$$\min U_a = \sum_{j=1}^{J} \sum_{f=1}^{F} \sum_{t=1}^{T} (X_{j,f,a,t}) \forall a = 1..A \tag{f3}$$

References

1. Schrijver, A.: Theory of Linear and Integer Programming. Wiley, New York (1998)
2. Leung, J.Y.-T., Anderson, J.H.: Handbook of Scheduling: Algorithms, Models, and Performance Analysis. Chapman & Hall/CRC, Boca Raton (2004). ISBN 1584883979
3. Błażewicz, J., Ecker, K.H., Pesch, E., Schmidt, G., Węglarz J.: Handbook on Scheduling. From Theory to Applications. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-32220-7. ISBN 978-3-540-28046-0
4. Coelho, J., Vanhoucke, M.: Multi-mode resource-constrained project scheduling using RCPSP and SAT solvers. Eur. J. Oper. Res. **213**, 73–82 (2011)
5. Rossi, F., Van Beek, P., Walsh, T.: Handbook of Constraint Programming (Foundations of Artificial Intelligence). Elsevier Science Inc., New York (2006)
6. Apt, K., Wallace, M.: Constraint Logic Programming Using Eclipse. Cambridge University Press, Cambridge (2006)
7. Milano, M., Wallace, M.: Integrating operations research. Constraint Program. Ann. Oper. Res. **175**(1), 37–76 (2010)
8. Achterberg, T., Berthold, T., Koch, T., Wolter, K.: Constraint integer programming: a new approach to integrate CP and MIP. In: Perron, L., Trick, Michael A. (eds.) CPAIOR 2008. LNCS, vol. 5015, pp. 6–20. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-68155-7_4

9. Bocewicz, G., Banaszak, Z.: Declarative approach to cyclic steady states space refinement: periodic processes scheduling. Int. J. Adv. Manuf. Technol. **67**(1–4), 137–155 (2013)
10. Sitek, P., Wikarek, J.: A hybrid approach to the optimization of multiechelon systems. Math. Probl. Eng. (2015). Article ID 925675. Hindawi Publishing Corporation. https://doi.org/10.1155/2015/925675
11. Sitek, P., Nielsen I.E., Wikarek, J.: A hybrid multi-agent approach to the solving supply chain problems. Procedia Comput. Sci. KES, 1557–1566 (2014). https://doi.org/10.1016/j.procs.2014.08.239
12. Sitek, P., Wikarek J.: A hybrid framework for the modelling and optimisation of decision problems in sustainable supply chain management. Int. J. Prod. Res. 1–18 (2015). https://doi.org/10.1080/00207543.2015.1005762
13. Sitek, P.: A hybrid approach to the two-echelon capacitated vehicle routing problem (2E-CVRP). Adv. Intell. Syst. Comput. **267**, 251–263 (2014). https://doi.org/10.1007/978-3-319-05353-0_25
14. Sitek, P., Wikarek, J.: A constraint-based approach to modeling and solving resource-constrained scheduling problems. In: Nguyen, N.-T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds.) ICCCI 2016. LNCS (LNAI), vol. 9875, pp. 423–433. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45243-2_39
15. Guyon, O., Lemaire, P., Pinson, Ă., Rivreau, D.: Solving an integrated job-shop problem with human resource constraints. Ann. Oper. Res. **213**(1), 147–171 (2014)
16. Blazewicz, J., Lenstra, J.K., Rinnooy Kan, A.H.G.: Scheduling subject to resource constraints: classification and complexity. Discrete Appl. Math. **5**, 11–24 (1983)
17. Lawrence, S.R., Morton, T.E.: Resource-constrained multi-project scheduling with tardy costs: comparing myopic bottleneck, and resource pricing heuristics. Eur. J. Oper. Res. **64** (2), 168–187 (1993)
18. Chandru, V., Rao, M.R.: Combinatorial optimization: an integer programming perspective. ACM Comput. Surv. **28**(1), 55–58 (1996)
19. Eclipse - The Eclipse Foundation open source community website. www.eclipse.org. Accessed 20 Apr 2016
20. SCIP. http://scip.zib.de/. Accessed 20 Apr 2016
21. Toth, P., Vigo, D.: Models, relaxations and exact approaches for the capacitated vehicle routing problem. Discrete Appl. Math. **123**(1–3), 487–512 (2002)
22. Li, Z., Janardhanan, M.N., Tang, Q., Nielsen, P.: Co-evolutionary particle swarm optimization algorithm for two-sided robotic assembly line balancing problem. Adv. Mech. Eng. **8**(9), 1–14 (2016). http://dx.doi.org/10.1177/1687814016667907
23. Nielsen, I., Dang, Q.-V., Nielsen, P., Pawlewski, P.: Scheduling of mobile robots with preemptive tasks. In: Omatu, S., Bersini, H., Corchado, Juan M., Rodríguez, S., Pawlewski, P., Bucciarelli, E. (eds.) Distributed Computing and Artificial Intelligence, 11th International Conference. AISC, vol. 290, pp. 19–27. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07593-8_3
24. Wang, J., Liu, C.: Fuzzy constraint logic programming with answer set semantics. In: Zhang, Z., Siekmann, J. (eds.) KSEM 2007. LNCS (LNAI), vol. 4798, pp. 52–60. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76719-0_9

Analysis of the Structured Information for Subjectivity Detection in Twitter

Juan Sixto(✉), Aitor Almeida, and Diego López-de-Ipiña

DeustoTech-Deusto Institute of Technology, Universidad de Deusto, Avenida de las Universidades 24, 48007 Bilbao, Spain
{jsixto,aitor.almeida,dipina}@deusto.es

Abstract. In this paper, we analyze the opportunities of the structured information of the social networks for the subjectivity detection on Twitter micro texts. The sentiment analysis on Twitter has been usually performed through the automatic processing of the texts. However, the established limit of 140 characters and the particular characteristics of the texts reduce drastically the accuracy of Natural Language Processing (NLP) techniques when compared with other domains. Under these circumstances, it becomes necessary to study new data sources that allow us to extract new useful knowledge to represent and classify the texts. The structured information, also called meta-information or meta-data, provide us with alternative features of the texts that can improve the classification tasks. In this paper we analyze the features of the structured information and their usefulness in the opinion mining sub-domain, specially in the subjectivity detection task. Also present a novel classification of these features according to their origin.

Keywords: Twitter · Text categorization
Data mining for social networks · Subjectivity detection
Social networks

1 Introduction

Since the Twitter social network was created in 2006, it has experienced a substantial growth, having more than 100 million of daily active users and 500 million tweets every day [43] nowadays. Currently Twitter is one of the largest textual data sources used in the data mining and knowledge extraction fields of research. As a part of these fields, sentiment analysis is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes [22]. Several research groups have used sentiment analysis techniques [27] over the Twitter micro-texts with an acceptable grade of success. However, the particular characteristics of Twitter (Hashtags, user references, inclusion of URLs, maximum of 140 characters) generate loosely formatted texts that are difficult to analyze. Addressing this

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXIX, LNCS 10840, pp. 163–181, 2018.
https://doi.org/10.1007/978-3-319-90287-6_9

challenge requires an adaptation of the classical techniques and tools to Twitter's unique requirements, that often results in a relevant decrease of their performance.

There are two possible approaches to this problem. The first one is improving the quality of the texts, in order to facilitate their automatic process. This text normalization task deals with several problems like the use of slang, word shortening, letter omissions and bad spelling [38]. The application of these techniques cleans the texts and improves the performance of the lexical analysis over them. The other approach is to improve the sentiment analysis process using the structured information [7] in addition to the tweeted text. Several researches have studied [2,32] how the external information can improve the sentiment analysis task. The obtained results show that the external information is a reliable source of knowledge about sentiment and opinion of texts.

This paper is an extended version of the paper published by Sixto et al. [41] in the ICCCI about the structured information used in opinion mining field, and its incorporation to the text analysis classical techniques. The Related Work section has been extended to incorporate more recent relevant works and a new subsection explaining in detail the ensemble methods. The Sect. 3 has been enriched through a series of new concepts and clarifications with regard to the original paper. The experiments presented in Sect. 4 are an advanced version of the previously presented works, including additional analysis of the studied features and more detailed specifications of the classification algorithms. Accordingly, the conclusions of the paper have been updated.

The rest of the paper is organized as follows: In Sect. 2, the context of this work is presented, according to the different related research fields. In Sect. 3, a novel analysis and classification of the most relevant elements of meta-data is presented. Section 4 covers a use case analysis based on the presented knowledge, and conclusions and future work are introduced in Sect. 5.

2 Related Work

Our study is focused on the contextual information of the user-generated opinions and their applications in the polarity classification tasks. Additionally, the case of study presented in Sect. 4 belongs to the Spanish language sub domain of the sentiment analysis in Twitter task. In this section, we will review the papers which our work is based on. To facilitate their comprehension, the relevant concepts and publications are presented divided into three different subsections according to the issue are referenced.

2.1 Contextual Applications in Sentiment Analysis

The primary objective of the Sentiment Analysis field is the automatic retrieval of subjectivity and opinion polarity. However, determining their scope is a very complex task and their areas of application are extensive. There are several surveys that summarize the main applications and the most common techniques in sentiment analysis [1, 24, 35].

In the field of the application of contextual information, there are several researchers who use the additional information available in social networks in the classification tasks. In 2011, Pennacchiotti and Popescu [37] presented a generic model for user classification in social media that combines linguistic features and explicit social network features. They also emphasize one of the main problems of contextual information, the difficulty of collecting the social network features of a dataset. Mislove et al. [30] analyzed data on a set of Twitter users in order to compare them with the U.S. population. To this end, they developed several techniques to enrich the information available of each user, detecting the gender, the ethnicity and geographic distribution of the users. This was one of the first studies that addressed the idea of the sampling bias and the study of the dataset population as an approach to improve predictions or measurements. Bermingham and Smeaton [4] modeled the political sentiment in order to predict electoral results in Twitter, including socio-linguistic features and unconventional punctuation. In the psychiatry sphere, De Choudhury et al. [10] developed a SVM classifier that can predict the likelihood of an individual to be depressed using Twitter. This work demonstrates the potential of the social networks as a tool for measuring and predicting emotional states of the users and gives new insights about the feature measures. Some of these features, used in their research, are the diurnal trends of the users, the volume of replies and the ego networks. Jiang et al. [19] present a target-depend sentiment classifier using the relations between tweets.

2.2 Spanish Polarity Classification

During the last years, research groups have published a large amount of approaches and methods in the sentiment analysis sphere, and have generated lexicons and polarity dictionaries that facilitate the tasks. Nevertheless, these tools are language dependent. Usually these are generated in one language and, at times, are translated to some other languages. This, combined with the difficulty of establishing standard linguistic rules between languages, causes a performance decrease when adapting the tools to other languages. Some datasets published that contain tweets in Spanish language are; the TASS Dataset published by Villena-Román et al. [45], the COST (Corpus of Spanish Tweets) dataset published by Martínez-Cámara et al. [26], or the multi-language dataset published by Cui et al. [8].

Several papers have been published on the field of sentiment analysis in social media, specifically focused on the Spanish language. Vilares and Alonso [1], reviewed a large quantity of bibliographic references in the Spanish scope. Also the TASS workshop [46], a satellite event of the SEPLN Conference[1], presents a huge amount of algorithms and techniques based on opinion extraction in Twitter.

[1] http://www.sepln.org/.

Analysis of the most used algorithms in the Spanish domain reveals that the most used machine learning algorithms are the Logistic Regression and the Support Vector Machines (SVM). Some examples of published classifiers that use these algorithms are Cerón-Guzmán [6], Sixto et al. [42] or Hurtado et al. [17].

Also in the Spanish domain, exist a few research about contextual information and its applications. For example, Park [36] introduces an interesting approach about the similarity of language use among people with similar social state. Siordia et al. [40] use emoticons and punctuation signs as polarity features in combination with linguistic features. It is particularly relevant the political affinity classification system published by Cotelo et al. [7], that uses vectorial models integrating the structured and unstructured of political tweets.

2.3 Ensemble Methods in Polarity Classification

As noted above, a large proportion of sentiment analysis approaches are based on learning algorithms and statistical mechanics. These classification systems usually use an array of linguistic features in order to train a probabilistic classifier that is able to predict the sentiment polarity of each tweet. In this context, the ensemble methods play an important role to obtain better predictive performance of standard learning algorithms. The ensemble methods consist on combining the results of multiple classifiers to produce a single hypothesis, trying to take advance of the combined accuracy, flexibility and the reduction of problems related to over-fitting [34].

At present, several ensemble methods are proposed, different techniques to address the combination issue that help to improve various aspects of the classification tasks. While some ensemble methods improve the accuracy of a single model constructing a multitude of instances of the same model through permutation, e.g., random decision forests, other ensemble methods combine the predictions of several other classification systems, e.g., the stacked generalization systems. This approach deals with the incorporation of multiple data sources in a single classifier, extracting knowledge from more than one aspect of the same item.

Ensemble methods also address the problems related with imbalanced datasets. Weiss [48] consider that ensemble methods, like AdaBoost [39] or the PNRule [20], are useful procedures to dealing with data rarity and imbalanced train corpora. By the same token, Liu et al. [23] present two algorithms to overcome the class-imbalance problems, called EasyEnsemble and BalanceCascade, ensembling classifiers using a particular way of stacking. An in-depth study about the main issues of the stacked generalization, a subset of the ensemble methods, was presented by Ting and Witten [44].

In the Spanish polarity classification domain, several ensemble methods have been presented, providing considerable improvements respecting their base models. Martínez-Cámara et al. [25] present a combination of three different polarity classifiers, using two strategies to break the ties; considering that a tie result must have a neutral semantic orientation and considering that a tie result implies that the tweet has no polarity. In similar way, Hurtado et al. [17] select the most frequent polarity in the training set as method to resolve ties between combined classifiers.

3 Structured Information: A Novel Analysis and Classification

Twitter contains a large amount of information about each tweet in addition to the user-generated text. Hashtags, retweets, replies, mentions, followers and many other relations bring us a considerable volume of information about the user network and all its components. Also, the user-generated texts contain some relevant elements despite being part of the text, are hard to model by the standard text representation models, like Bag-of-words [16] or Word2Vec [29]. There are many examples of different aspects of the structured information used as part of classification models, most of the time as reinforcement of other basic classification techniques. All of these information can be a knowledge source about users and their opinions, as we have seen in previous researches, and bring an improvement to the sentiment analysis tasks. Assuming that the use of structured information in sentiment analysis tasks has been proved, the aim of this study is to check the efficiency of the structured information in the sub domain of supervised sentiment analysis at global level. This sub domain consist in performing an automatic sentiment analysis to determine the global polarity of each message about any topic, without any previous topic discrimination.

According to our research proposal, we pretend to study the new possibilities of the structured information in the global level of sentiment analysis, adapting features used in concrete domains as politics [7] or psychiatry [15], and other features not used yet. Also, we establish a novel classification between the different aspects of the structured information in order to define their characteristics and performance.

3.1 Structured vs Unstructured Information

Currently, does not exist a unequivocal terminology to refer to the contextual information of the tweets. Barbosa and Feng [2] name them as "Tweet Syntax Features", referring to the features inferred from meta-information with higher relevance, and Liu [21] name them as "Twitter specific clues", in reference to the inclusion of meta-information based features in a subjective classifier of tweets. In this paper we use a terminology based on the Structured and Unstructured information concepts, described in Cotelo et al. [7], to difference between the "standard" and the contextual information. As Cotelo et al. [7] defined, "Twitter, in addition to its defects in the quality of the texts, has a great virtue: the texts are accompanied by much structured information that is related to multiple entities of different nature."

In order to achieve the proposed task, it's necessary a full understanding of the Twitter structure and of how their components (Users, Texts, Communities) relate among them. The Fig. 1 represents the most frequent components of the social network and their relations, considered relevant to this work. Also, as part of the study of the data, the structural information of the social network has been classified according to its origin, that is, the component where the information origins. The basic components described in the Fig. 1 and the relations between them are briefly described below.

- **Users:** The users are the content generator elements of the network, posting tweets and linking to them as authors. Users may also be related between them in two different forms; following another user (direct relation) or mentioning another user or users in a tweet (indirect relation).
- **Tweets:** Tweets are the main object of the social network and contains several elements in addition to the text. Each tweet is always directly related with their author (user), and may be related with any other element of the network in different ways, like sharing Hashtags with other tweets or referring to other tweet.
- **Hashtags:** These meta-data tags are used to specify a topic or a idea in relation with the content of the text. Because is strongly related with the subject of the tweet, these tags may serve as reference to know the author's specific opinions.
- **URLs:** The hyper-links included in some tweets are sources of additional information that can be extracted and analyzed. Additionally, these hyper-links may be used in order to categorize tweets.

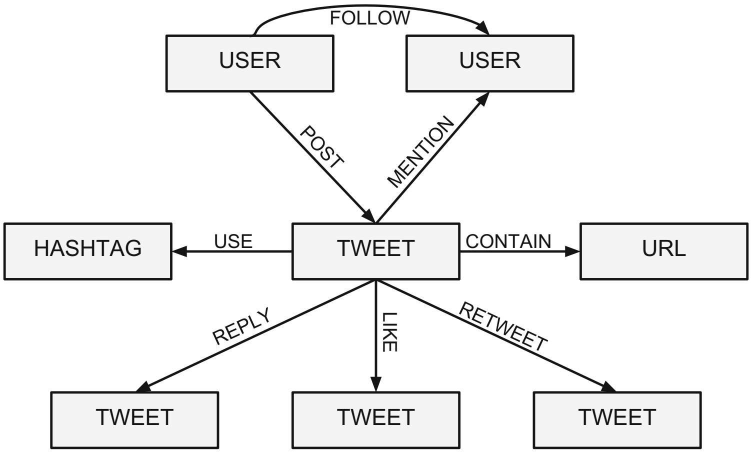


Fig. 1. Twitter structure representation.

As seen above, a wide range of existing data can be used to extract information and generate new knowledge analyzing the structured relations of these elements. To this end, as part of our research, we have classified the available tweet data in four categories (explained in the next subsections), based on their origin within the social network structure.

3.2 Text Attributes

This category comprehends the attributes that appear in the text, but that does not depend on the words representation. These attributes emphasize on how the tweets are written, including the characteristic properties of Twitter that are part of the text, but not encompass semantic or syntactic analysis. Technically, this category could be considered as a subcategory of Tweet attributes, as long as the text is a part of each tweet, even so we consider this features clearly different from the Tweet attributes. Examples include: *hathtags, links, emoticons, punctuation, retweet, used language.* These attributes can be used by the standard text representation models, although often are used as independent features by classification models. Text attributes may be categorical (usually boolean) as the presence of uppercases, or continuous as the percent of uppercases. Also may be simple as presence/no presence flags or complex as the presence of irony.

3.3 Tweet Attributes

This category comprehends all the characteristics that define a only tweet but not are part of the text. These usually refer to the living process of the tweet within the network, the relations of the users with them or the way it had been posted. Examples include: *quantity of retweets, quantity of favourites, creation date/time, the application that sent the tweet, associated place.* In some cases, tweet attributes involve an extra difficulty due to their evolution over time, requiring a delay to observe them.

3.4 User Attributes

This category comprehends the attributes relative to the authors of the tweets. These attributes represent several aspects of the users that may be relevant to understand the context of a tweet. Occasionally, these aspects are compiled to generate "user profiles" that simplify the user representation in the system. Examples include: *location, political affiliation, post habits.* Seems reasonable to expect that the ideological affiliation of a user provides information about their subjective opinion about a particular topic. Likewise, the behavior toward the user from other users could provide important information about subjectivity. Cotelo et al. [7] provide a good example of use users behavior to infer the political affiliation.

3.5 Topographic Attributes

This category comprehends all the measures about the network topology. As we have seen before, the structure of the social network and the relations between users generate a complex web of elements. The topographic attributes often require some complex calculations and help us to know the role of a user or tweet in the network. Examples include: *Modularity class of user, In-degree and Out-degree of users, Network Communities of the mentioned users.* As we seen before, the structure of the social network and the relations between users generate a complex web of elements that also provide information of themselves.

4 Experiments

In order to demonstrate the usefulness of the contextual information in sentiment classification tasks, we present a research case study based on the subjectivity detection task. Our purpose is to predict the subjectivity of tweets using the structured components. The main characteristics of the subjectivity classification task are reviewed below. In order to detail the realized experiments, we also describe the chosen corpus, the studied features and the classification algorithms.

4.1 Use Case: Subjectivity Detection in Spanish Language

Liu [21] defines the subjectivity classification as follows: *"Subjectivity classification classifies sentences into two classes: subjective and objective. An objective sentence expresses some factual information, while a subjective sentence usually gives personal views and opinions"*. Figures 2 and 3 show two examples of tweets obtained from the TASS dataset, described in Sect. 4.2. The example 1 was labeled as "Negative" polarity, and the example 2 was labeled as "NONE" polarity, and therefore, the examples are a subjective and an objective text respectively. However, neither of these examples include adjectives that explicits subjectivity or opinion. In fact, the only one element that difference the first example of an objective fact is the irony aspect of the last question. Had it not been for the ironic intention of the author, suggesting a potential criminal activity, the text only would express a past event without any value judgment, in similar way to the example 2. In contrast, an informed human notices that the

Fig. 2. Example Tweet 1 (Negative polarity). Translation: "*So Miss Transparency @Cospedal breaches the law of publishing her income and patrimony ... why will it be?*"

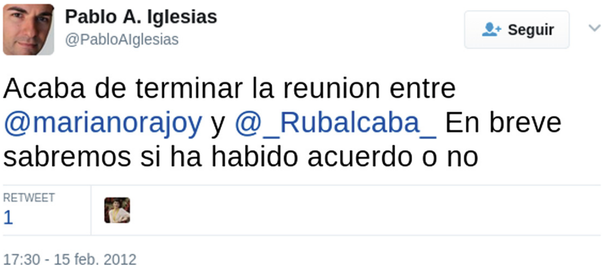


Fig. 3. Example Tweet 2 (None polarity). Translation: "*The meeting between @marianorajoy and @_Rubalcaba_ has just ended. Soon we will know if there has been agreement or not.*"

author (@sanchezcastejon) and the mentioned user (@cospedal) own to antagonistic political parties[2] and therefore the aim of the tweet is to accuse a political opponent. Some specific cases of opinion, as irony, involves an extra complexity in the task that requires new techniques to resolve them (Bosco et al. [5]). That is why structured information may connect the human knowledge and the standard text analysis techniques, providing new information to the model.

The subjectivity detection problem has been studied for several years in different areas, especially the approaches based on supervised learning. Since the beginning of the sentiment classification researches, subjectivity has also been explored as part of the global sentiment classification area. The sentiment classification can be expressed as a classification problem of three or more classes: Positive opinion, Negative opinion and no opinion, although these three classes are expanded in several cases. Furthermore, the problem can be divided into two classification subproblems; the opinion detection task first and the distribution between positive and negative opinions later on.

In this paper we address this first problem that is usually named Subjectivity Detection problem. In the Twitter research area, several authors have worked on the sentiment classification in which the subjectivity detection plays an important role. Barbosa and Feng [2] use some twitter features to implement a subjectivity classification. Davidov et al. [9] propose a sentiment classifier that uses punctuation-based features in posted texts. Due to their short length, each tweet is considered as a single sentence and accordingly, each tweet has only a single sentiment polarity. Also Giorgis et al. [13] demonstrate that the use of a subjectivity detection classifier as feature in a general polarity classifier can lead to an improved performance.

4.2 Selected Corpus

The scope of our case study is focused on the detection of subjectivity in Tweet texts in Spanish language. At present, only a few number of datasets that satisfy

[2] @sanchezcastejon is a Spanish politician of the Spanish Socialist Workers' Party (PSOE) and @cospedal is a Spanish politician of the People's Party (PP).

these conditions can be found in the state of art. Only the multi-language dataset presented in Volkova et al. [47] and the TASS dataset [46] includes texts with no sentiment in Spanish language.

Finally we decided to use the datasets of the sentiment analysis task at TASS'15[3] workshop [46]. This is an evaluation workshop for sentiment analysis focused on Spanish language, organized as a satellite event of the annual conference of the Spanish Society for Natural Language Processing (SEPLN). This paper is focused on the first task of the workshop, that consists in performing an automatic sentiment analysis to determine the global polarity of each message in the provided corpus. Tweets are divided into six different polarity labels: strong positive (P+), positive (P), neutral (NEU), negative (N), strong negative (N+) and one additional no sentiment tag (NONE). Tweets are also divided into two sets, the training dataset with 7.219 (11%) items and the test dataset with 60.798 (89%) items. Additionally, the task includes a 1.000 items dataset, a balanced and hand-labeled subset of the test dataset, that we use as evaluation of the performance of our systems.

The dataset contains a 20.54% (Train) and 12.30% (Test) of tweets tagged as NONE, that are considered as True results by our classifier, whereas the rest of labels are considered as False results. As our aim is the detection of objective and subjective texts as a first step of the polarity detection task, the use of a balanced or binary corpus was discarded. The significance of obtaining not only a high accuracy but also a high F1 measures have been taken into account and are explained in the Experimental Results section. As shown in Fig. 4, the dataset presents a imbalanced class distribution, including rare classes like NEU (2.90%). This involves a critical role of the evaluation metrics and issues associated with data rarity, according to Weiss [48].

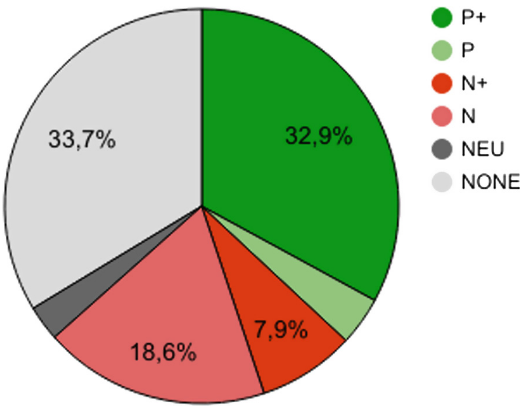


Fig. 4. Distribution of polarity labels in TASS dataset.

[3] Workshop on Sentiment Analysis at SEPLN Conference.

4.3 Extracting the Structured Information

The TASS Dataset is distributed in XML[4] format and includes the text, the date, the author and the unique ID of each tweet. Using these unique IDs, we implemented a data mining wrapper in order to extract the full information of each tweet through the Twitter REST API[5]. In similar way, we implemented a second wrapper to extract the relations between the different authors of the dataset. This tool allowed us to establish a map of relations, based on the *Follow* option of the social network, to deepen knowledge of it and to know better the behavior of the dataset authors. These directional relation map is presented in Fig. 5, showing the three different communities detected in the graph using the modularity classes of the users. According to the information provided by their creators, all of the authors of the dataset are well-known Spanish-speaking Journalists, Celebrities or Politicians. Using this information, during our experiments authors and tweets were classified according to their social category, in order to extract knowledge about their behavior. This classification has served us to analyze the performance of the quantity of re-tweets and likes in the tweets, expressing several differences between user categories, as presented in Fig. 6. Figure 6 shows the statistical distribution of Retweets and Favorites by polarity and by user category. The left-hand column shows the statistical distribution the quantity of Retweets of each Tweet, and the right-hand column shows the statistical distribution of the quantity of Favorites obtained of each Tweet. Results are summarized in four subdivisions; all the tweets of the dataset (line 1), tweets published by Politicians (line 2), tweets published by Journalists (line 3) and tweets published by Celebrities (line 4). Significative differences can be observed between the median of both Tweet attributes of each user category, for example, the quantity of Favorites obtained by Celebrities are significantly higher than the obtained by other users. Also the differences between sentiment polarities are clearly visible, for example, the contrast between Retweets obtained by Negative and Neutral tweets compared with Neutral and Positive polarities.

The meta-information extraction process has involved some problems that required different methods to resolve them. First, the tweets of the TASS Dataset were posted during the year 2012, that is, tweets with years of lifespan in the moment of extraction. Obviously, the more time passed since the posting time, the more probable is that the tweet or the author have it been deleted, losing their information. These problem not only includes the tweets deleted individually by their authors but also the accounts deleted, that includes the complete deletion of all his texts and the information relative of the user. Table 1 show the number and percent of tweets recovered during the experiments according to their user type.

[4] Extensible Markup Language.

[5] https://dev.twitter.com/rest/public.

Table 1. TASS Dataset distribution and recovery percentages.

	Total	Journalists	Celebrities	Politicians
Tweets	68.017	33.228	10.318	24.471
Recovered	60.687	30.250	9.516	20.921
Recovered (%)	89.22%	90.96%	92.23%	85.60%

4.4 Features

Several features have been used to address the subjectivity classification problem [21], the vast majority based in the text: bag-of-words, vectorial representations of words, n-grams, etc. In the structural features area, some authors have studied the use of several features in the subjectivity classification of English tweets. Barbosa and Feng [2] exploit the use of retweets, hashtags, replies, links, punctuation, emoticons and the number of upper cases in a subjectivity classifier. Davidov et al. [9] use the length of the words and the punctuation signs as features.

Based on the approach to structured and unstructured information presented in by Cotelo et al. [7], we implement a single sentiment classifier for subjectivity detection, combining two classifiers, each one trained with a different type of information. As shown above, the corpus contains six different labels and as consequence it is not balanced in respect to the objectivity-subjectivity axis. This has been taken into account during the training and evaluation process. In order to train the classifier based on structural information, we have composed a feature list:

URL, Exclamation marks, Emoticons and Uppercase words. According to the work of Barbosa and Feng [2], these features have been used in our work.

Uppercase Percent. In addition to the number of uppercase words, we used the percentage of uppercase characters of the total characters. This technique differs from the feature proposed by Barbosa and Feng [2], not only counting the words that starts with upper case, but counting all the characters. This ensures that the tweets with all capitals texts, typically used for emphasis or "shouting", are taking into account.

Favorites. Twitter includes an option called "favorite", that allows the users to like individual tweets. Our study detected a relation between the average of "favorites" by tweet and their sentiment polarity.

Modularity Class. In Twitter, users may subscribe to other users tweets. This is known as "following" and establishes a directed graph of relations between users. During the conduct of our study, we proposed to extract the relations between the users (authors of the tweets) through their "following" relations and generated their relation graph. This relation graph is represented in Fig. 5. Their modularity class revealed the existence of only three communities. A preliminary research shows that this communities are formed by associated individuals related to the left/right political parties or ideologies, and a third group of neutral celebrities. Used as feature, the modularity of each user generates a increase of accuracy and F1-measure in the classifier.

Graph Degrees. Some other attributes of the relation graph have been proposed and tested as features. In detail, the In-Degree and Out-Degree punctuations of the authors have proven to be useful to classification task.

RT. Twitter includes an option called"Retweet", that allows to share a message from another user. This boolean feature expresses if the analyzed tweet is a "retweet" of an original tweet.

Ellipsis. During our study, it was noted that some objective tweets includes an ellipsis. In Twitter, ellipsis is often used to make observations about external information, like headlines, urls, or quotes from other users.

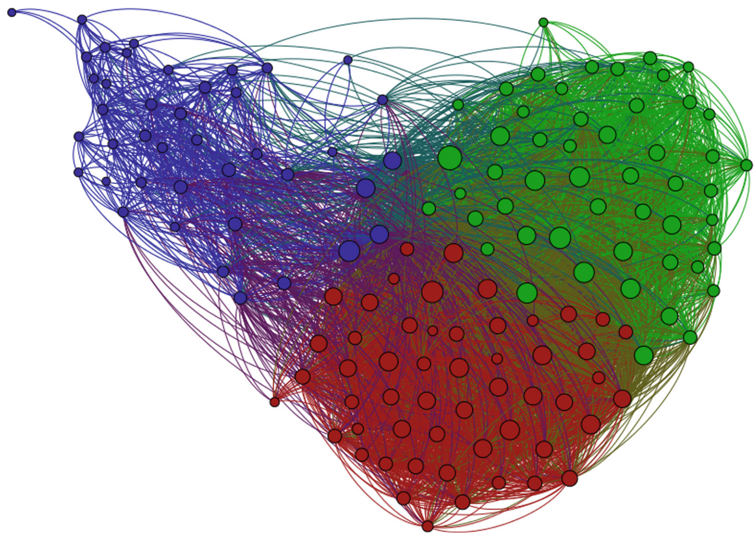


Fig. 5. User relations graph.

The second classifier was based on nonstructural information. The selected model to represent the texts was the commonly used the bag-of-words [16]. This model represents each tweet as a matrix of token counts of its words.

4.5 Experimental Results

Considering that the dataset is unbalanced, as shown in Sect. 4.2, we decided to extract three different measures to evaluate the performance of the tested systems; Accuracy, Macro-F1 and NONE-F1. The Accuracy is the proportion of *true positive* and *true negative* results among the total of the dataset, expressed as a percentage. However, when the prior probabilities of the classes are unbalanced, this metric can be misleading [18]. In binary classifiers, accuracy ranks

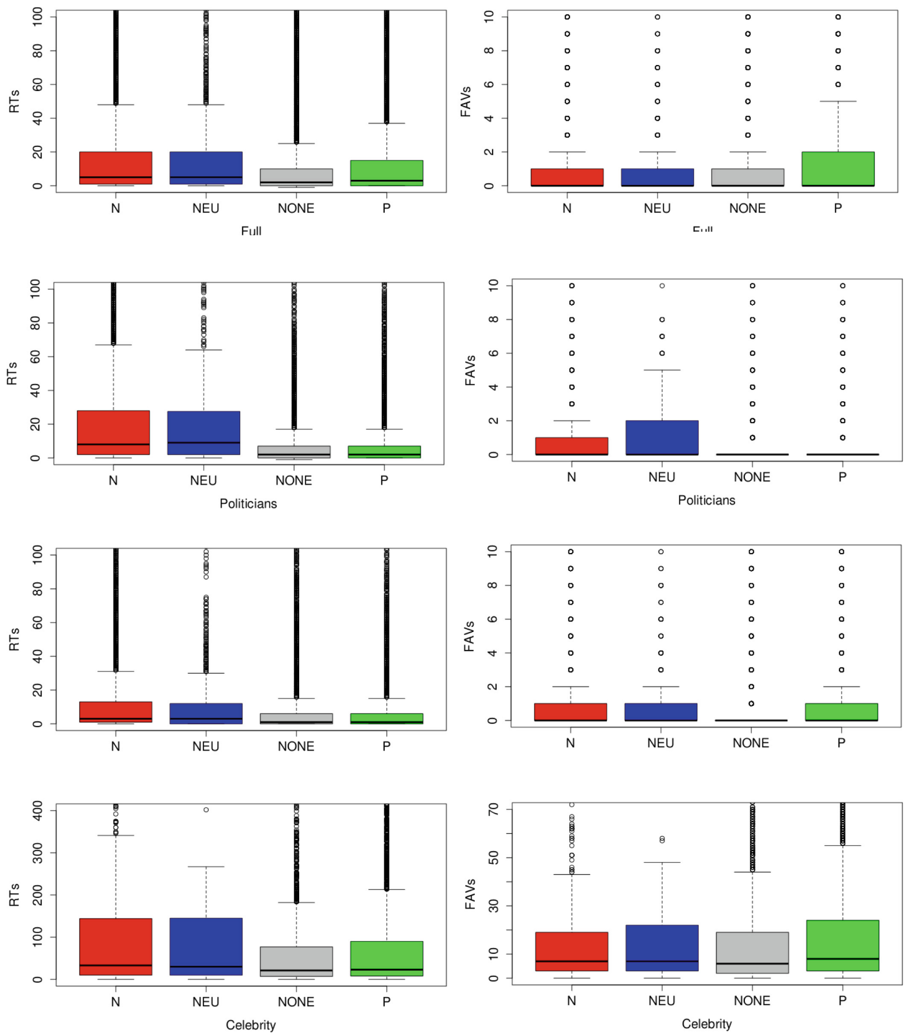


Fig. 6. Statistical distribution of retweets and likes by polarity and user type.

more weight on the majority class than on minority class, obtaining a low perform on the minority class [48].

The F1-measure, also F-score, is a widely used metric in the natural language processing context [51]. The F1-measure is the harmonic average of precision and recall, covering the positive predicted values and the true positive rate of the systems. While the macro-averaged F1-measure provides information of the full system performance overall the dataset, the so-called NONE-F1 are the specific measure of the NONE (or True) labels of the system. The F1-measure is considered as a relevant score for evaluating the accuracy of a test with a unbalanced dataset. Also we consider relevant to evaluate the specific F1-measure of the NONE labels, in order to rank the contribution of the classifier to a polarity detection task. Obtained results are reviewed in the Table 2.

Table 2. Results for subjectivity detection.

System	Accuracy	Macro F1	NONE-F1
Meta-information	89.50%	70.80%	43.20%
Bag-of-words	79.30%	70.40%	46.80%
MI+BoW	88.30%	69.20%	43.50%
Stacking MI+BoW	**89.80**%	**90.20**%	**55.65**%

The first approach to the task consists on a single classifier with the structural features described in Sect. 4.3. Multiple classifier models were tested, like LogisticRegression, Naive Bayes, and Random Forest, obtaining the best results with a GradientBoosting classifier [12], reaching a 70.8% in Macro-F1 and a 43.0% in NONE-F1. Then, a new test has been realized using the bag-of-words model, in order to contrast the performance of structured and unstructured approaches. The best results have been obtained with a LogisticRegression model with a balanced mode to automatically adjust weights, obtaining a 70.4% in Macro-F1 and a 46.8% in NONE-F1. The similarity of the results shows that both approaches have a relevance for the task, although the poor results involve that the task is complex.

We also investigated the chance of combining both approaches in order to improve the results of the classification task. To do this, we selected two different approaches; using both feature lists in a single classifier and a stacked generalization [50]. In the first case, the structural features of each tweet were added to their matrix representation, generating a new features list. This is a simple way to merge both models and check if both feature lists are directly complementary or need more complex techniques to improve the results. The best results has been realized with a GradientBoosting classifier and obtained a 69.2% in Macro-F1 and a 43.5% in NONE-F1. This technique does not improve the results, and in several cases decreases them, proving that is necessary the use of other techniques to merge the heterogeneous features effectively.

At least, we realized a stacked generalization work for combining both models. For the level-0 generalizer we use five different classifiers; Logistic Regression, GradientBoosting, Multinomial Naive Bayes, Random Forest and Calibrated with Isotonic Regression. Each of the classifiers were trained with both models, generating a ten classifiers array that formed the level-0 models. Then we used a Logistic Regression model for the level-1 classification model. We found that the use of regression models obtain best results, according to the presented by Ting and Witten in [44]. This approach obtained a 90.22% in Macro-F1 and a 55.66% in NONE-F1, being the best obtained results. This improvement implies that a complex technique, like the stacking, benefits from the heterogeneous features in relation to the other approaches.

5 Conclusions and Future Work

Our objective in this study was to learn about the contextual information, their uses at the subjectivity detection task and their application improving the text based models, as noted above. We have attempted to adapt previous knowledge to the contextual data to the global polarity detection task in Spanish language. Our study has verified a hypothesis already applied in other social areas and expanded the knowledge relative to the contextual information, adding new ways to use the contextual information to the previous approaches of the state of art. Also we presented a contextual data classification for a better understanding of their nature and characteristics, in addition to providing practical examples. We presented a first interaction of a subjectivity detection approach which uses some contextual elements to build its features. This approach overtakes the basic classifiers and achieves to combine the structured and unstructured information, establishing a method to complement the standard classification techniques. Although the accuracy and F1-measure are around 90%, the poor values in the micro average reveal that exists an huge margin for improvement in the task. As future work, we want to connect our investigation with a complete polarity detection task, applying the extracted knowledge in other sentiment categories, exploring new contextual features. We want to perform a more extensive analysis to check more Twitter components and their relation with the different polarities, considering that distinct features could be related with only a particular sentiment category. Also, we seek to apply the contextual features with more complex models that include lexicons of polarity and semantic resources to really see the impact of them.

5.1 Knowledge and Human Interaction

We consider that reach a more deep knowledge about the classification tasks is a relevant part of the future work in our domain. Our intention is, together with the improvements of the most recent text classification techniques, make progress in the integration of structured and unstructured information in order to achieve a much most complex and variated models capable of detecting the rare classes in imbalanced data. In the sentiment analysis task in Twitter, we consider that exist two relevant problems that could be addressed with meta-information in the future; the low amount of information provided by the short longitude of the texts in Twitter, and the imbalanced datasets. The problems presented by Weiss [48] and Weiss and Provost [49] are strongly linked with the difficulties of the sentiment analysis tasks.

Similarly, a deep-knowledge of the structured information, the users behavior, and their performance in the sentiment mining tasks could be useful for develop new more comprehensive datasets and evaluation workshops. Also the use of meta-information features could be useful to improve the criteria used to generate silver corpus to train the general embedding, like Nabil et al. [33]. These techniques are very extended in the state-of-art, as Montejo-Ráez and Díaz-Galiano [31] or Go et al. [14], and consist in collecting high quantities of tweets to train language models. This tweets are labeled automatically according to the emoticons present in the texts.

Given the high cost of a hand-labeling process, researchers usually use human-knowledge based approaches, like emoticons. This criteria are not infallible but their mistakes are not statistically relevant, specially in positive-negative classification tasks. However, these criteria could be improved using additional knowledge based on human interaction over the structured data.

Acknowledgements. This work has been partially supported by the Spanish Ministry of Economy and Competitiveness under the project E-RMP (CSO2015-64495-R).

References

1. Alonso, M.A., Vilares, D.: A review on political analysis and social media. Procesamiento del Lenguaje Natural **56**, 13–24 (2016)
2. Barbosa, L., Feng, J.: Robust sentiment detection on Twitter from biased and noisy data. In: Proceedings of 23rd International Conference on Computational Linguistics: Posters, pp. 36–44 (2010)
3. Belkaroui, R., Faiz, R.: Towards events tweet contextualization using social influence model and users conversations. In: Proceedings of 5th International Conference on Web Intelligence, Mining and Semantics, p. 3. ACM (2015)
4. Bermingham, A., Smeaton, A. F.: On using Twitter to monitor political sentiment and predict election results (2011)
5. Bosco, C., Patti, V., Bolioli, A.: Developing corpora for sentiment analysis: the case of irony and senti-TUT. IEEE Intell. Syst. **28**(2), 55–63 (2013)
6. Cerón-Guzmán, J.A.: JACERONG at TASS 2016: an ensemble classifier for sentiment analysis of Spanish tweets at global level. In: Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 32nd SEPLN Conference (SEPLN 2016), pp. 35–39 (2016)
7. Cotelo, J.M., Cruz, F., Ortega, F.J., Troyano, J.A.: Explorando Twitter mediante la integración de información estructurada y no estructurada. Procesamiento del Lenguaje Natural **55**, 75–82 (2015)
8. Cui, A., Zhang, M., Liu, Y., Ma, S.: Emotion tokens: bridging the gap among multilingual Twitter sentiment analysis. In: Salem, M.V.M., Shaalan, K., Oroumchian, F., Shakery, A., Khelalfa, H. (eds.) AIRS 2011. LNCS, vol. 7097, pp. 238–249. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25631-8_22
9. Davidov, D., Tsur, O., Rappoport, A.: Enhanced sentiment learning using twitter hashtags and smileys. In: Proceedings of 23rd International Conference on Computational Linguistics: Posters (2010)
10. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. In: ICWSM, p. 2 (2013)
11. Esparza, S.G., O'Mahony, M.P., Smyth, B.: Mining the real-time web: a novel approach to product recommendation. Knowl.-Based Syst. **29**, 3–11 (2012)

12. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232 (2001)
13. Giorgis, S., Rousas, A., Pavlopoulos, J., Malakasiotis, P., Androutsopoulos, I.: aueb.twitter.sentiment at SemEval-2016 task 4: a weighted ensemble of SVMs for Twitter sentiment analysis. In: Proceedings of SemEval, pp. 96–99 (2016)
14. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, vol. 1, no. 12 (2009)
15. Han, B., Cook, P., Baldwin, T.: Unimelb: Spanish text normalisation. In: Tweet-Norm@SEPLN, pp. 32–36 (2013)
16. Harris, Z.S.: Distributional structure. Word **10**(2–3), 146–162 (1954)
17. Hurtado, L.F., Pla, F., Buscaldi, D.: ELiRF-UPV en TASS 2015: Análisis de Sentimientos en Twitter. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN Co-located with 31st SEPLN Conference (SEPLN 2015) (2015)
18. Jeni, L.A., Cohn, J.F., De La Torre, F.: Facing imbalanced data-recommendations for the use of performance metrics. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), pp. 245–251. IEEE (2013)
19. Jiang, L., Yu, M., Zhou, M., Liu, X., Zhao, T.: Target-dependent twitter sentiment classification. In: Proceedings of 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 151–160 (2011)
20. Joshi, M.V., Agarwal, M.C., Kumar, V.: Mining needle in a haystack: classifying rare classes via two-phase rule induction. ACM SIGMOD Rec. **30**(2), 91–102 (2001)
21. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. **5**(1), 1–167 (2012)
22. Liu, B., Zhang, L.: A survey of opinion mining and sentiment analysis. In: Aggarwal, C., Zhai, C. (eds.) Mining Text Data, pp. 415–463. Springer, Boston (2012). https://doi.org/10.1007/978-1-4614-3223-4_13
23. Liu, X.Y., Wu, J., Zhou, Z.H.: Exploratory undersampling for class-imbalance learning. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **39**(2), 539–550 (2009)
24. Martínez-Cámara, E., Martín-Valdivia, M.T., Ureña-López, L.A., Montejo-Ráez, A.R.: Sentiment analysis in Twitter. Nat. Lang. Eng. **20**(01), 1–28 (2014)
25. Martínez-Cámara, E., Gutiérrez-Vázquez, Y., Fernández, J., Montejo-Ráez, A., Muñoz-Guillena, R.: Ensemble classifier for Twitter sentiment analysis (2015)
26. Martínez-Cámara, E., Martín-Valdivia, M.T., Ureña López, L.A., Mitkov, R.: Polarity classification for Spanish tweets using the COST corpus. J. Inf. Sci. **41**(3), 263–272 (1015)
27. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. J. **5**(4), 1093–1113 (2014)
28. Mejova, Y., Srinivasan, P., Boynton, B.: GOP primary season on Twitter: popular political sentiment in social media. In: Proceedings of 6th ACM International Conference on Web Search and Data Mining. ACM (2013)
29. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
30. Mislove, A., Lehmann, S., Ahn, Y.Y., Onnela, J.P., Rosenquist, J.N.: Understanding the demographics of Twitter users. In: ICWSM, vol. 11, no. 5 (2011)
31. Montejo-Ráez, A., Díaz-Galiano, M.C.: Participación de SINAI en TASS 2016. In: Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN (2016)
32. Monti, C., Rozza, A., Zapella, G., Zignani, M., Arvidsson, A., Colleoni, E.: Modelling political disaffection from Twitter data. In: Proceedings of 2nd International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM 2013) (2013)

33. Nabil, M., Atyia, A., Aly, M.: CUFE at SemEval-2016 task 4: a gated recurrent model for sentiment classification. In: Proceedings of 10th International Workshop on Semantic Evaluation (SemEval-2016) (2016)
34. Opitz, D., Maclin, R.: Popular ensemble methods: an empirical study. J. Artif. Intell. Res. **11**, 169–198 (1999)
35. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1–2), 1–135 (2008)
36. Park, S.: Sentiment classification using sociolinguistic clusters. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN Co-located with 31st SEPLN Conference (SEPLN 2015), pp. 99–104 (2015)
37. Pennacchiotti, M., Popescu, A.M.: A machine learning approach to Twitter user classification. ICWSM **11**(1), 281–288 (2011)
38. Porta, J., Sancho, J.L.: Word normalization in Twitter using finite-state transducers. In: Tweet-Norm@SEPLN, vol. 1086, pp. 49–53 (2013)
39. Schapire, R.E.: A brief introduction to boosting. IJCAI **99**, 1401–1406 (1999)
40. Siordia, O.S., Guerrero, M.G., Avila, E.S.T., Jimenez, S.M., Moctezuma, D., García, E.A.V.: Sentiment analysis for Twitter: TASS 2015. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN Co-located with 31st SEPLN Conference (SEPLN 2015) (2015)
41. Sixto, J., Almeida, A., López-de-Ipiña, D.: An approach to subjectivity detection on Twitter using the structured information. In: International Conference on Computational Collective Intelligence, Part 1, pp. 121–130 (2016)
42. Sixto, J., Almeida, A., López-de-Ipiña, D.: Improving the sentiment analysis process of Spanish tweets with BM25. In: Métais, E., Meziane, F., Saraee, M., Sugumaran, V., Vadera, S. (eds.) NLDB 2016. LNCS, vol. 9612, pp. 285–291. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41754-7_26
43. Smith, C.: DMR Twitter statistic report, Last Modified 26 Feb 2016. http://expandedramblings.com/index.php/downloads/twitter-statistic-report/. Accessed 28 Mar 2016
44. Ting, K.M., Witten, I.H.: Issues in stacked generalization. J. Artif. Intell. Res. (JAIR) **10**, 271–289 (1999)
45. Villena-Román, J., Lana-Serrano, S., Martínez-Cámara, E., González-Cristóbal, J.C.: TASS - workshop on sentiment analysis at SEPLN. Procesamiento del Lenguaje Natural **50**, 37–44 (2013)
46. Villena-Román, J., García-Morera, J., García-Cumbreras, M.A., Martínez-Cámara, E., Martín-Valdivia, M.T., Ureã-López, L.A.: Overview of TASS 2015. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN, vol. 1397. CEUR-WS.org (2015)
47. Volkova, S., Wilson, T., Yarowsky, D.: Exploring demographic language variations to improve multilingual sentiment analysis in social media. In: EMNLP, pp. 1815–1827 (2013)
48. Weiss, G.M.: Mining with rarity: a unifying framework. ACM SIGKDD Explor. Newsl. **6**(1), 7–19 (2004)
49. Weiss, G.M., Provost, F.: Learning when training data are costly: the effect of class distribution on tree induction. J. Artif. Intell. Res. **19**, 315–354 (2003)
50. Wolpert, D.H.: Stacked generalization. Neural Netw. **5**(2), 241–259 (1992)
51. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Hoste, V.: SemEval-2016 task 5: aspect based sentiment analysis. In: ProWorkshop on Semantic Evaluation (SemEval-2016) (2016)

An Efficient Parallel Method for Optimizing Concurrent Operations on Social Networks

Phuong-Hanh Du(✉), Hai-Dang Pham(✉), and Ngoc-Hoa Nguyen(✉)

Department of Information Systems, VNU University of Engineering and Technology, Hanoi, Vietnam
{hanhdp,dangph,hoa.nguyen}@vnu.edu.vn

Abstract. This paper presents our approach to optimize the performance of both reading and writing concurrent operations on large-scale social networks. Here, we focus on the directed and unweighted relationships among members in a social network. It can then be illustrated as a directed, unweighted graph. Moreover, determining the relationship between any two members is similar to finding the shortest path between two vertices. With such a large-scale dynamic social network, we face the problem of having concurrent operations from adding or removing edges dynamically while one may ask to determine the relationship between two members. To solve this issue, we propose an efficient parallel method based on (i) utilizing an appropriate data structure, (ii) parallelizing the updating actions and (iii) improving the performance of query processing by both reducing the searching space and computing in multi-threaded parallel. Our method was validated by the datasets from SigMod Contest 2016 and SNAP DataSet Collections with fine positive experimental results compared to other solutions.

Keywords: Bidirectional BFS search
Concurrent operations on social networks
Multi-threaded parallel computing

1 Introduction

Today, social networks play a significant role in our networked society. Facebook, Twitter, WhatsApp, etc. are popularly used in everyday life. To model social networks, graph theory has been considered as a proper methodology. Furthermore, a member is generally modeled by a vertex, and the direct relationship between two members is represented by an edge. In the graph, the optimal (shortest) path problem means finding a path between two vertices which the cost of its constituent edges is minimized. This is a fundamental and well-studied combinatorial optimization problem with many practical uses: from GPS navigation to routing schemes in computer networks; search engines apply solutions to

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXIX, LNCS 10840, pp. 182–199, 2018.
https://doi.org/10.1007/978-3-319-90287-6_10

this problem on website interconnectivity graphs and social networks implement them on graphs of peoples' relationships [16].

On social networks, there are three aspects regarding this kind of graph that we should consider: (i) the number of vertices and edges are enormous; (ii) the graph is dynamic due to the relationship changes among members and new members registered; and (iii) the shortest distance query is mostly performed in order to find the way to establish the connection between any two members. Thus, the shortest distance (SD) query allows analyzing the influence of a user to the community [2]; to identify the closeness between two users; to find more related users or contents by using the socially-sensitive search [4]. Although the SD problem is frequently trivial, the task to answer optimal path queries is a considerable challenge in the context of having large-scale and quickly changing/elastic social network in reality [7].

In this paper, we demonstrate a method to enhance the performance of parallelizing the concurrent operations on large-scale and elastic social networks. To gain this purpose, we propose an appropriate data structure for modeling the network, following a strategy to parallelize the updating operations and finally the sufficient approach to perform the computing queries (such SD) in parallel. This is an extended method from our previous work [17] where the parallelization technique of concurrent updating operations is not yet proceeded.

The rest of this paper is organized as follows. Section 2 presents preliminaries and related works. Section 3 details our efficient method for improving the performance of both updating and computing operations. In Sect. 4, we summarize our experiments to verify and benchmark our approach. Finally, the last section provides some conclusions and future works.

2 Problem Formulation and Related Works

Here, we concentrate on the directed and no-weighted relationship among members of a social network. Hence, it can be described by a directed, unweighted graph: a vertex represents a member, and an edge represents the relationship between two corresponding members. Determining the relationship between any two members is similar to finding the shortest path between two vertices.

2.1 Data and Operation Model

As mentioned above, in this article, we focus only on the directed/unweighted social networks. The latter can then be represented as a graph $G(V, E)$ where V is the set of all members (and called vertices) and $E = (v_i, v_j)|v_i, v_j \in V$ represents the set of all directed relationships (called edges) (v_i and v_j are connected with a single unweighted link). The total number of edges to (incoming) and from (outgoing) a vertex v_i is called the degree of v_i and is represented as $deg(v_i)$.

For the vertex, it is conveniently represented and identified by a number. That leads to encode the $|V|$ vertices from 0 to $|V| - 1$. For the graph edges, there are three main structure types: (i) edge lists, (ii) adjacency matrices and (iii) adjacency lists. In the large scale graph, the adjacency matrices representation cannot be used because of the limit of main memory size. The edge list structure is simple, but the operations on the graph, such as insertion and deletion, are difficult. The appropriate way to represent the large-scale edges of the graph is the adjacency list structure.

For the directed and unweighted graph, there are two ways to organize its adjacency lists [16]:

– List of incoming edges from a given vertex: Graph edges will be represented by a list of consecutive incoming vertices (*incoming_edges*),
– List of outgoing edges from a given vertex: Similarly, the edges will be represented by a list of consecutive outgoing vertices (*outgoing_edges*).

For the social networks modeled by the directed graph, the fundamental relationship operations can be "*Read*" and "*Write*". Once writing on such a network, an edge is simply added or deleted, whereas a *Read* on it is traversing its vertices. A graph traversal is also considered as a query on the graph. In a traversal, we are performing whether a Depth-First Search or a Breadth-First Search (BFS) method [3]. From that, three operation types will be specified and described in more details as follows:

Definition 1. *To add an edge ['A' u v]: Modifying the current network by adding another relation (edge) from the member u to v. In the case of the input of the original graph, if the edge already exists, the graph remains unchanged. If one (or both) of the specified endpoints of the new edge does not exist in the graph, it should be added.*

Definition 2. *To delete an edge ['D' u v]: Removing the relation between (u v) from the current network. If the specified relation does not exist in the network, the latter should remain unchanged.*

Definition 3. *To compute the SD ['Q' u v]: Performing and returning the shortest distance (SD) from u to v in the current network. If there is no relation between these members or if neither of them exists in the network, the answer should be* -1*. The distance between any member and itself is always 0.*

2.2 Problem Statement

From the above data and operation model, let $S = \{op_1, op_2, ..., op_n\}$ be the liste of n concurrent operations on the graph G. Each $op_i = (a_i, u_i, v_i)$ of S is an operation having one of above three operation types (A, D or Q).

Similar to performing with concurrent transactions (in the database management system) or concurrent processes (in the operating system), we will model the concurrent operations S in G as well as a schedule for updating and SD operations. In S, the order in which each op_i is performed is critical, especially with update operations and directly affecting the SD queries in S. Therefore, the processing of concurrent operations in S must ensure the correct order of each update operation in S. From that, it is clear that the consistency of graph will be ensured when performing each action in this schedule.

To facilitate the experiment, a list S of concurrent operations will be represented by a set of lines in a text file. Each line represents an operation op_i. To denote the end of S, we add the last line containing the single character F (as Finish). Thus, the schedule is organized similar to a *batch* concept in the ACM SigMod Programming Contest 2016 [18].

The following figure indicates some example schedules of concurrent operations on a graph:

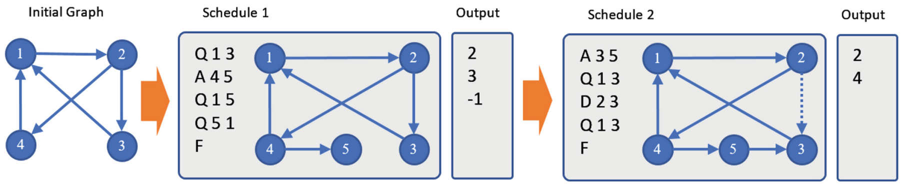


Fig. 1. Concurrent operations on graph

From that point on, what interests the researchers is how to parallelize the concurrent operations S on G so that performance is as optimal as possible. Before discussing our approach, some of related works will be covered in the next section.

2.3 Related Works

To perform the actions on the graph, a substantial number of tools and libraries can feasibly be used to address this problem. NetworkX, for instance, is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks and graph [15]. SNAP C++ library [5] is very popular for a general purpose, high-performance system for analysis and manipulation of large graphs. These libraries use the bidirectional BFS (bBFS) algorithm to compute the shortest distance between two vertices. However, their implementation is not optimal due to general purpose requirements: the shortest distance is calculated only in sequence (performed by only one CPU core) and the directional selection is based on the number of enqueued vertices only. The latter leads to the situation that there are a lot of enqueued vertices in the next traversal loop.

For the large-scale graphs, the recent study of GraphLab in [1], PowerGraph in [13], GraphX in [14] are dedicated for processing the large graph in both distributed and parallel computation. These systems are efficient for general purposes in case of having a dominant computing platform such as clusters and supercomputers [1]. Nevertheless, they are not adequate for the shortest distance computation for the dynamic graph in the context of medium computing platforms, similar to NetworkX and SNAP C++.

For the dynamic graph, the efficient shortest path grabs the attention of researchers. [6–8] show the works on processing of shortest path traversal queries in online and dynamic transportation graphs. In order to harness the multi-core architecture, [9–12] present the work-efficient parallel breadth-first search algorithms on the large graph. Notwithstanding, the graph update actions did not get the attention in these works.

The similar works for managing large-scale directed, dynamic graphs should describe five final teams participated the SigMod Programming Contest 2016 [18]. We will summarize their main ideas here:

1. **H_minor_free:** this team was the Overall Winner in the SigMod Programming Contest 2016. Their main ideas are based on considering the state of each edge: ALIVE for passable edges; DEAD for impassable edges; and UNKNOWN for the edges modified (insert/delete) in the batch, checked by binary search for SD queries. For each batch, they experience 3-step process: (i) updating operations add UNKNOWN to the edges; (ii) utilizing bBFS processes SD queries using one of OpenMP threads; and (iii) updating edge states to ALIVE or DEAD after bBFS. Additionally, this team also implemented a strategy to select an appropriate algorithm depending on graph parameter of average distance [19].
2. **uoa_team:** this group used the approach of multiversioning data structures for updating operations and using the heuristics to optimize multi-threading bBFS. They also used the *threadpool*11 and *concurrentqueue* libraries in order to enhance the performance of bBFS. The team got the Runner Up prize in that Contest.
3. **gStreamPKU:** the solution of the third placed team is based on (i) reducing the number of basic operations per query with Bit Compression and Optimizing program's spatial locality; and (ii) building parallelism Delta Graph (updating operations) to fully support concurrent bBFS query execution within a batch.
4. **while1:** this team used the idea of "transaction edge list" for performing the updating operations. All node and edge lists are duplicated for both NUMA nodes to ensure memory locality for processing parallel bBFS.
5. **akgroup:** the strategic solution of our team is based on the adjacency lists for both incoming and outgoing vertices. The updating operations have to be committed before performing the parallel heuristic bBFS [16].

The last three teams received the third prize in that Contest.

3 Method for Performing Concurrent Operations

To process the concurrent operations on such a social network represented by a directed, unweighted graph G, we suppose that all n concurrent operations have been executed by order of each operation $Op[i]$. This approach is strictly a serial schedule for concurrent operations. Thus, it is clear that the social network will be consistent after performing all concurrent operations.

3.1 Organizing the Data Structure

To make in profit of the multi-core, multi-chip computational power, we use the idea of H_minor_free team [19] to model an edge: each edge can be one of the three state: ALIVE, DEAD and UNKNOWN for executing concurrent operations. Therefore, the concurrent operations in the Fig. 1 can be illustrated by the following figure (Fig. 2):

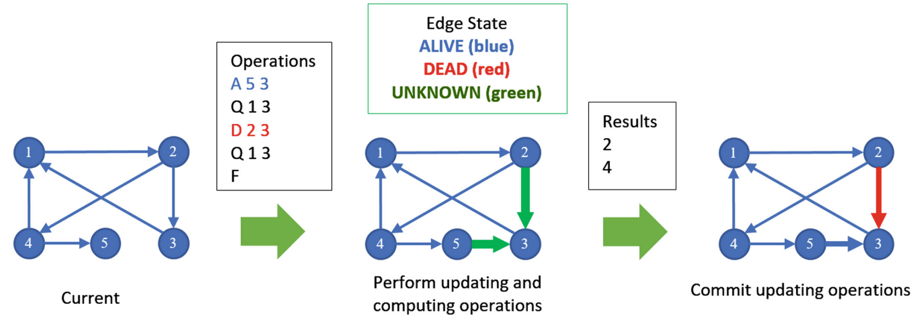


Fig. 2. Execution of updating operations

For managing big data, the consecutive item list seems to be the best way to allow having the highest cache hit rate [16]. Moreover, in this research, we mainly examine large social networks which have no more than one billion members, i.e. that are not Facebook-sized. Therefore, the member identification is represented by a 30-bit integer, and if we allocate 4-byte integer for a vertex, the two last bits for the state of edge can then be used.

From the above idea, the graph data is represented by the adjacency lists described as following: (i) Each vertex is identified by a 4-byte number; (ii) All incoming/outgoing nodes of a node u are stored in a sorted vector $incomingEdges[u]$/ $outgoingEdges[u]$. Each item v of these vectors has the first 30-bits for the vertex; the last 2-bits for the state of edge (u, v).

By using both incoming and outgoing edge lists, we can quickly explore the graph in both directions. That allows computing the SD queries by the bBFS algorithm [9].

3.2 Performing Concurrent Operations

Based on the above data structure, all concurrent operations of a batch can be executed by the serial schedule to assume the consistency of graph. This schedule will be ingested by the Algorithm 1.

Algorithm 1. Perform the concurrent operations

Input: Graph G and Op is the schedule of n operations (a,u,v) on the graph//'a' means 'action'
Output: G committed all updates and list of shortest distances for all queries

```
for t=0; t < n; t++ do
    (a,u,v) = Op[t];
    if a =' Q' then
        Queries.push_back(t,u,v);//push back the tuple (t,u,v) into the Queries vector
    else
        Updates.push_back(t,a,u,v);//push back the tuple (t,a,u,v) into the Updates vector
    end if
end for
Make edges in parallel for the updating operations; //Algorithm 2
Perform in parallel all queries;//Algorithm 4
Commit the updating operations in parallel;//Algorithm 3
```

In this algorithm, each tuple in both the Queries and Updates vector has the timestamp t specified by order of operation in the concurrent list. We will use this parameter for checking whether an edge can be used or not when computing the shortest distance queries.

3.3 Optimizing the Updating Operations

All updating operations will be firstly performed by adding/modifying an edge having the UNKNOWN state. In order to improve the performance of this task, we propose the strategy for parallelizing updating operations: the updating operations on each $outgoingEdges[v_t]/incomingEdges[v_t]$ have to be executed by sequence only on a thread t. Therefore, if we run in parallel $threadNum$ threads, we can perform in parallel $threadNum$ updating operations without implementing any locking techniques. This strategy is presented as the Algorithm 2.

In this algorithm, the operation $InsertNode$ assumes to insert a vertex v in the right position in the sorted vector V_n representing all relationship members of n. The edge (n, v) should have the state UNKNOWN if v does not exist on the V_n.

Meanwhile, the operation $RemoveNode$ of the vertex v in the sorted vector V_n will modify the state of the edge (n, v) to UNKNOWN if v exists on the V_n and the current state is not DEAD.

Algorithm 2. Make edges for the updating operations

Input: *Updates*: schedule of m updating operations (t,a,u,v); Graph G;
Output: G, *incomingSum*, *outgoingSum* modified by *Updates*

```
//Perform in parallel for updating outgoing edges
for w=0; w < maxThread; w++ do
    for each (t,a,u,v) in Updates do
        if u mod maxThread == w then
            if a == 'A' then
                InsertNode(outgoingEdges[u],v);
            else
                RemoveNode(outgoingEdges[u],v);
            end if
        end if
    end for
end for
//Perform in parallel for updating incoming edges
for w=0; w < maxThread; w++ do
    for each (t,a,u,v) in Updates do
        if v mod maxThread == w then
            if a == 'A' then
                InsertNode(incomingEdges[v],u);
            else
                RemoveNode(incomingEdges[v],u);
            end if
        end if
    end for
end for
```

The two global lists *incomingSum*[*v*] and *outgoingSum*[*v*] are used to store all the incoming and outgoing vertices from every vertex of *incomingEdges*[*v*] and *outgoingEdges*[*v*] respectively. Theses parameters will be used for computing the SD queries.

Algorithm 3. Commit the updating operations

Input: *Updates*: schedule of m updating operations (t,a,u,v); Graph G;
Output: G committed by all m updating operations

```
//Perform in parallel for all Updates
for each (t,a,u,v) in Updates do
    if a == 'A' then
        CommitAdd(outgoingEdges[u],v);
        outgoingSum[u] += outgoingEdge[v].size();
        CommitAdd(incomingEdges[v],u);
        incomingSum[v] += incomingEdges[u].size();
    else
        CommitDelete(outgoingEdges[u],v);
        outgoingSum[u] -= outgoingEdge[v].size();
        CommitDelete(incomingEdges[v],u);
        incomingSum[v] -= incomingEdges[u].size();
    end if
end for
```

Once all queries have been executed, we will commit all updating operations by modifying the state of updating edges to the right state by the Algorithm 3. Note that the operation *CommitAdd* of a vertex v into the sorted vector V_n will change the state of v (for the edge (n,v)) to ALIVE. Similarly, applying for the operation *CommitDelete*, the state of v for the edge (n,v) will be DEAD.

3.4 Optimizing the Query Processing

For the query processing, we are first interested in determining the shortest relationship distance between two members of the social network. Since this is modeled by a directed and unweighted graph, the parallel BFS algorithm is adequate to answer that distance [9,11], especially for large-scale and dynamic directed/unweighted graph. Another point should be emphasized here that all the capability of the modern multi-core and multi-thread CPU have to be exploited. To do this, our strategy has to offer in parallel processing for all consecutive queries, which will be discussed more details as following.

i. Algorithm of Shortest Distance Computation:

For bidirectional BFS, the incoming edges and outgoing edges should be used. Two bitmap arrays (called *incomingMaps*/*outgoingMaps*) are also used for remarking traveled incoming/outgoing nodes. To reduce the search space, the strategy of predicting deeply on the graph has been proposed [16]. In the standard bBFS, we follow the branch that has the smaller number of traversing vertices in the incoming/outgoing queue [9]. However, this information is not enough. As shown in the Fig. 3 for the query of finding the shortest distance from 1 (src) to 9 (dst), by using the normal bBFS, we follow the outgoing queue of source vertex 1 (due to possessing only one child) and outgoing vertex 3 will be traversed at first. It means that $(1+6+2=9)$ vertices must be enqueued.

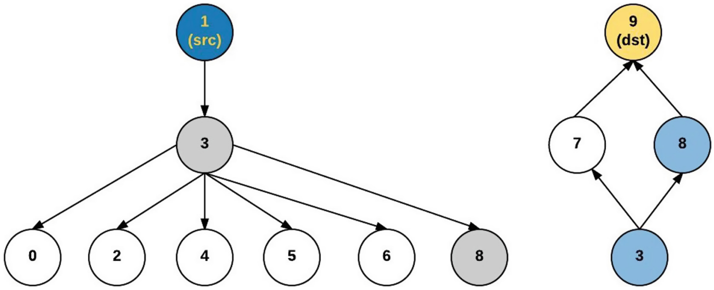


Fig. 3. Bidirectional BFS shortest distance traverse

In order to reduce the search space (number of the traversal vertices), we modify the bBFS algorithm by the strategy: only following the direction that has the smaller sum of next level enqueued vertices and their children. For example, in the Fig. 3, in the beginning, we can realize that the sum of the outgoing queue (only one vertex) and their children (six vertices can be enqueued in the next level) is 7. Meanwhile, the sum of incoming queue (two vertices) and their children (only one vertex can be enqueued in the next level) is 3. Thus, following the incoming direction can be more effective and efficient than the outgoing one.

Our strategy tuning the bBFS algorithm can be explained in the Algorithm 4. The *incomingSum*, *outgoingSum* lists are used in this algorithm in order to implement the above strategy.

Algorithm 4. Compute SD (u,v) by the bidirectional BFS

Input: (t, u, v): containing the timestamp t and two vertices u and v; $incomingMaps, outgoingMaps$: global arrays for marking the traveled vertices of each worker; $incomingQueues, outgoingQueues$: global arrays for enqueueing the potential vertices to travel of each worker; $incomingSum$, $outgoingSum$ and G

Output: the shortest distance from u to v

```
if u == v then
    Return 0;
end if
if (incomingEdges[v].size() == 0) || (outgoingEdges[u].size() == 0) then
    Return -1; //No path between u and v
end if
Get inMap and outMap from global incomingMaps and outgoingMaps for this thread
Get inQueue and outQueue from global incomingQueues and outgoingQueues for this thread
inQueue ← v; //initiate in-queue
setBit(v, inMap); //indicate v was visited
outQueue ← u; //initiate out-queue
setBit(u, outMap); // indicate u was visited
inCost = 0, outCost = 0; //distance to/from u/v
outSize = outgoingSum[u]; inSize = incomingSum[v];
while (outSize > 0)&&(inSize > 0) do
    if (outSize < inSize) then //following the outQueue
        outSize = 0; outCost+ = 1;
        for each vertex e in outQueue do
            for each vertex n in outgoingEdges[e] do
                if !testBit(n, outMap) then
                    state = GetState(n);
                    if (state==ALIVE)||(state&UNKNOWN && IsEdgeAlive(e,n,state,t)) then
                        if testBit(n, inMap) then
                            Clear all bits of inMap and outMap;
                            Return inCost + outCost;
                        end if
                        //ifnot, we push this vertex to outQueue
                        outQueue ← n; setBit(n, outMap);
                    end if
                end if
            end for
            outSize+ = outgoingSum[e];
        end for
    else//following the inQueue
        inSize = 0; inCost+ = 1;
        for each vertex e in inQueue do
            for each vertex n in incomingEdges[e] do
                if !testBit(n, inMap) then
                    state = GetState(n);
                    if (state==ALIVE)||(state&UNKNOWN && IsEdgeAlive(n,e,state,t)) then
                        if testBit(n, outMap) then
                            Clear all bits of inMap and outMap;
                            Return inCost + outCost;
                        end if
                        //ifnot, we push this vertex to inQueue
                        inQueue ← n; setBit(n, inMap);
                    end if
                end if
            end for
            inSize+ = incomingSum[e];
        end for
    end if
end while
Clear all bits of inMap and outMap;
Return -1;
```

In this algorithm, the $testBit(v, map)$ function returns the value of bit at the position v in the buffer map. Meanwhile, $setBit(v, map)$ function will set the bit at the position v in the buffer map to 1.

For an UNKNOWN edge, we use the vector of updating operations for deciding to use this edge or not. The idea of the $IsEdgeAlive(u, v, t, state)$ function has to answer this problem and illustrated as the following algorithm:

Algorithm 5. Verify if an edge (u, v) is ALIVE at the timestamp t

Input: (u,v,state,t): edge (u, v) with $state$ at the query moment t; $Updates$ is vector of all updating operations
Output: TRUE if (u, v) is ALIVE at t; FALSE if not

```
i = lower_bound(Updates, (u, v, t, 0));
if i == Updates.begin() then
    Return (state&1) == 0;
end if
(lu,lv,la,lt) = Updates[i-1];
if (u == lu && v == lv) then
    Return (la=='A');
end if
Return (state&1) == 0;
```

In this algorithm, *lower_bound* is a function implemented in the C++ libraries. It finds (by using the binary search) and returns an iterator pointing to the first element in $Updates$ that is not less than the tuple $(u, v, t, 0)$.

ii. Query Parallel Processing:

Our solution for processing the consecutive queries is:

– Global incoming queue and outgoing queue are employed for each call of searching the shortest length. Then, each searching thread will use only proper in/out queues determined by an interval of global in/out queues.
– Each searching thread will also own the appropriate in/out map slots calculated from the global in/out maps. Since the searching process finishes, these in/out slots will be cleared for the next search.
– Cilk Plus is used for performing queries in parallel[1]. By conducting a lot of proper experiments, our solution was implemented with some multi-threaded parallel computing methods like OpenMP[2], Pthread[3]. We note that the Cilk Plus method seems to be the most efficient one and achieve the outstanding performance in our solution.

Thus, the parallelization of consecutive queries is illustrated as the following algorithm:

[1] https://www.cilkplus.org/cilk-documentation-full.
[2] http://openmp.org/wp/.
[3] https://computing.llnl.gov/tutorials/pthreads/.

Algorithm 6. Execute all consecutive queries on the graph G

Input: *Queries* having all shortest distance queries; Graph G
Output: vector of results

```
//Perform in parallel the queries by Cilk Plus method
for i = 0; i < Queries.size(); i++ do
    (u, v, t) ← Queries[i];
    distances[i] = shortest_distance(u, v, t);
end for
```

4 Experiments and Evaluation

Based on the proposed method, we built and implemented our solution in C++ language. The experiments were performed in the evaluation machine having 2 x Intel(R) Xeon(R) CPU E5-2697 v4 @ 2.30 GHz (45 MB Cache, 18-cores per CPU), 128 GB for the main memory, CentOS Linux release 7.2.1511, gcc 6.3.0. This computing system was configured with maximum 36-threads in parallel (disable hyperthreading).

To validate our method for optimizing the concurrent operations on large-scale dynamic directed graph, two datasets from the Stanford Large Network Dataset Collection [20] and one from SigMod Programming Contest [18] are selected to evaluate the results. Some statistics from the graphs are showed down below (Table 1):

Table 1. Graph collection statistics

Parameter	SigMod dataset	Pokec dataset	LiveJournal dataset
Edges	1574074	68993773	30622564
Nodes	3232855	4847571	1632803
Diameter (longest shortest path)	9	16	11

Particularly, LiveJournal is a free on-line community with almost 10 million members. In addition, Pokec is the most popular on-line social network in Slovakia that connects more than 1.6 million people [20].

For testing workloads, we built a tool in order to generate the workloads by using the testing protocol of SigMod [18]. For each graph, we produced two experiment sets of workloads and results. The first one concentrates on the query operations. This experiment workload composes 1,000,000 operations involving queries/insertions/deletions ratios 80/10/10 respectively (hence there are approximately 800,000 queries, 100,000 insertions, 100,000 deletions). We denote *8-1-1* for this set. In the second experiment set, we are interested in the real situation which update operations play an integral part in social networks. The set contains 1,000,000 operations as well, but differs from the respective proportion of queries/insertions/deletions of 50/40/10 (equivalent to roughly 500,000 queries, 400,000 insertions, 100,000 deletions). *5-4-1* is denoted for this kind of experiment set.

To measure the time performance and the parallel efficiency, we run our solution (called **bigGraph**) and other solutions detailed in the Sect. 2 (**akgroup** - our old solution; **H_minor_free**; **uoa_team**; **gStreamPKU**; and **while1**) with distinct numbers of thread for each dataset. To be specific, all solution was tested with 1, 2, 4, 8, 16, 24, and 32 threads (since we have maximum 36-threads in our computing system). Every test was executed ten times for each dataset per each number of thread and gave the average result. Due to the *while1*'s solution produced the wrong result (mismatches found), we do not list the while1's results. We are also interested in the *SNAP* (C++) and *networkx* (Python) solutions [16], however, they are not formed on parallel processing approach, so their performance is not sufficient enough to compare with the five above solutions.

The final results are illustrated in the figures below. It should be noted that we also display the best execution time of each solution for all ten test times and seven values of thread numbers (from 1 to 32 as above).

From Figs. 4 and 5, the execution time of the uoa_team is the longest and its trend is obviously upward relative with the increase of the number of thread. Afterward, the gStreamPKU's solution also takes a large amount of running time at around 1 s and 1.5 s in SigMod 8-1-1 and SigMod 5-4-1 dataset respectively. Additionally, our method does not get the first rank when running on the only single thread, it, however, come on top from two threads onwards.

Concerning Pokec dataset, as it is demonstrated in Figs. 6 and 7, the running time of all solutions decrease considerably when the number of thread rises. In the Pokec 8-1-1, at one thread, gStreamPKU team took the first place at about 18.5 s compared to practically 23.5 s of our new solution. Moreover, uoa_team

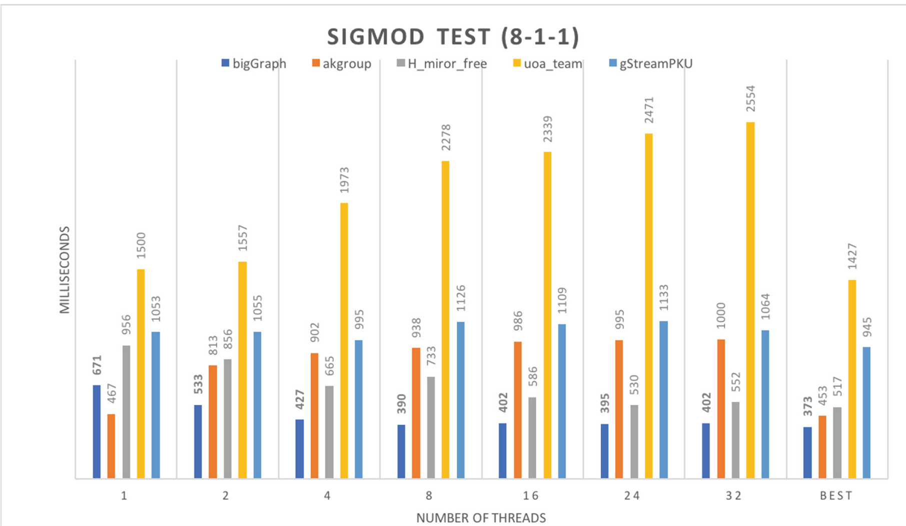


Fig. 4. 8-1-1 Sigmod dataset test results

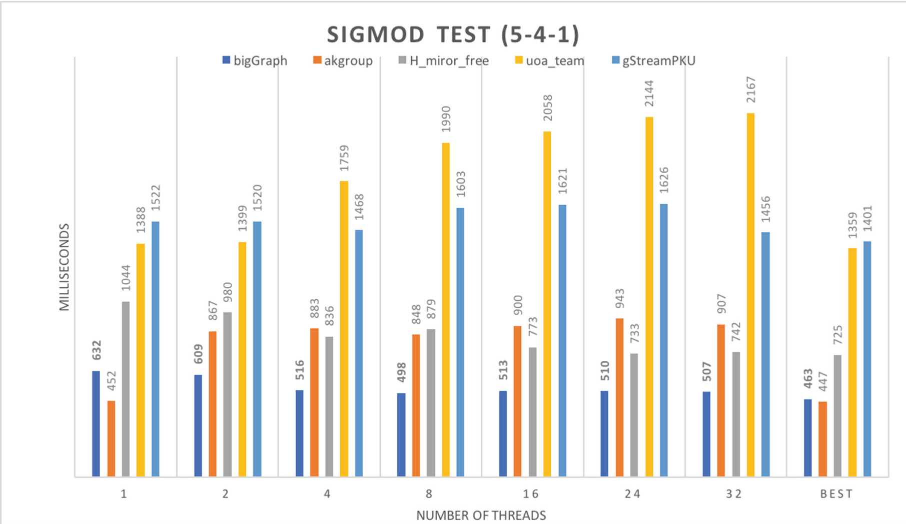


Fig. 5. 5-4-1 Sigmod dataset test results

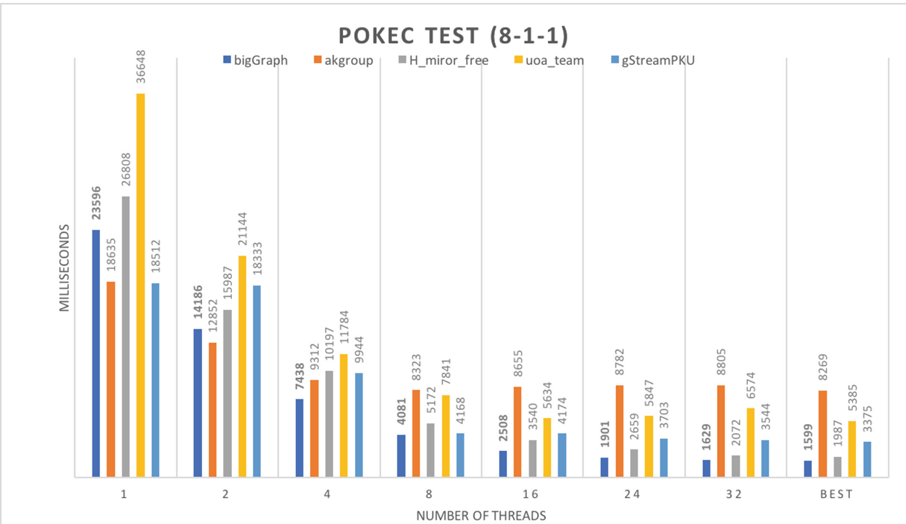


Fig. 6. 8-1-1 Pokec dataset test results

outstrip all other methods at approximately 12.8 s at two threads. Nevertheless, it can be noticeably seen that from four threads and more, our solution comes first. Likewise, the similar pattern is showed in all solutions with the exception of akgroup in the Pokec 5-4-1.

Regarding LiveJournal, similar to Pokec dataset, there is a gradual decline in the running time of five solutions with the rising number of threads. In the

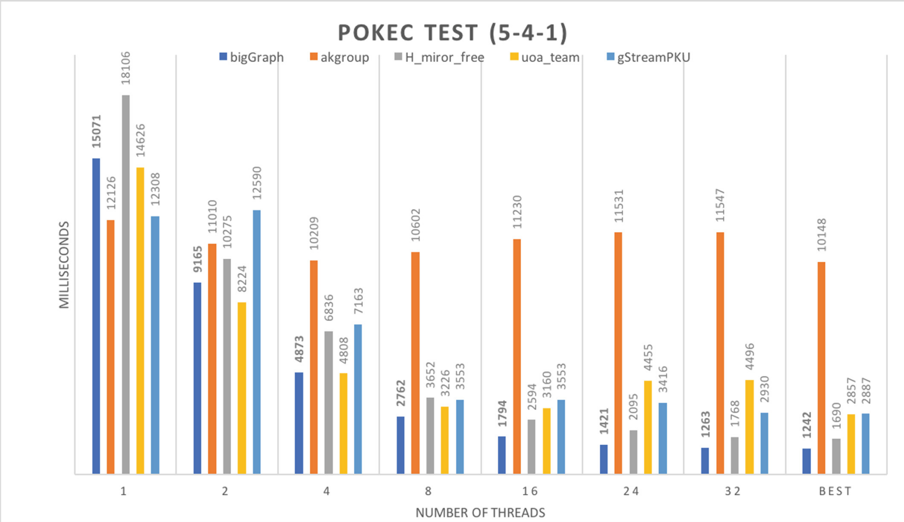


Fig. 7. 5-4-1 Pokec dataset test results

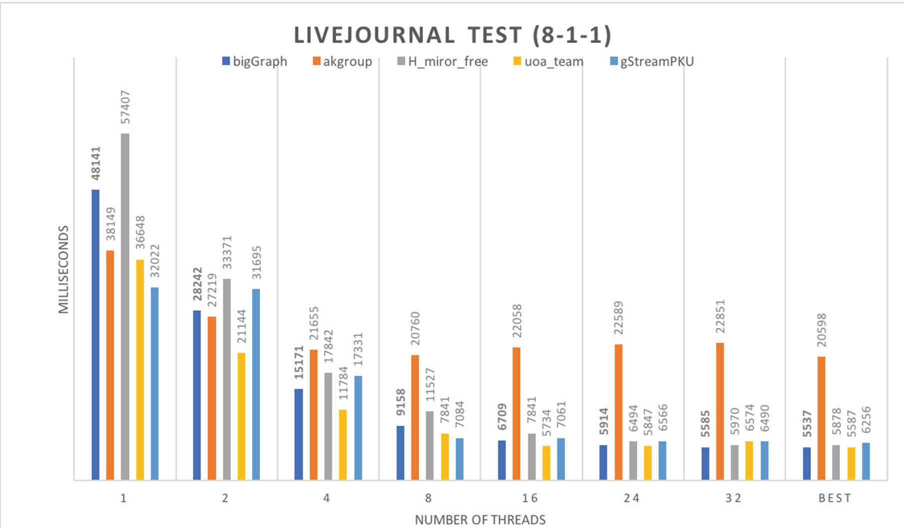


Fig. 8. 8-1-1 LiveJournal dataset test results

8-1-1 set, there are a few unique points such as gStreamPKU outdo all the other teams at the 1 and 8 threads and uoa_team still stand in first place at 2, 4, 16, 24 threads. Another point which should be emphasized is that uoa_team's method wins the race from 2-threads to 16-threads in the 5-4-1 set. At 24 and 32 threads, our solution overtakes all other methods as the fastest running time (Figs. 8 and 9).

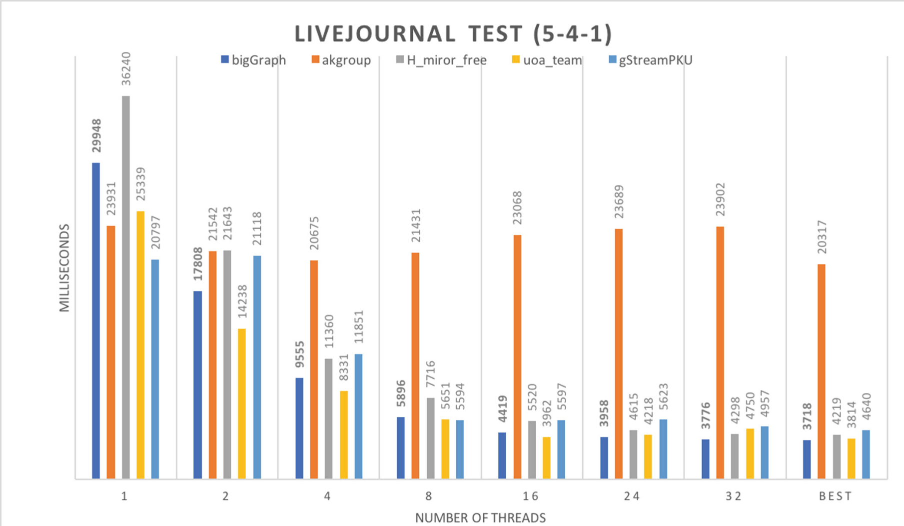


Fig. 9. 5-4-1 LiveJournal dataset test results

The best execution time for all of cases is illustrated as the following table (the corresponding thread number is also included) (Table 2):

Table 2. Best performance statistics

Solution	SigMod		Pokec		LiveJournal	
	8-1-1	5-4-1	8-1-1	5-4-1	8-1-1	5-4-1
akgroup	453 (1T)	**447** (1T)	8269 (8T)	10148 (4T)	20597 (8T)	20317 (2T)
H_minor_free	517 (24T)	725 (24T)	1987 (32T)	1690 (32T)	5878 (32T)	4219 (32T)
uoa_team	1427 (1T)	1359 (1T)	5385 (16T)	2857 (16T)	5587 (16T)	3814 (16T)
gSteamPKU	945 (4T)	1401 (32T)	3375 (32T)	2887 (32T)	6256 (32T)	4640 (32T)
bigGraph	**373** (8T)	463 (32T)	**1599** (32T)	**1442** (32T)	**5537** (32T)	**3718** (32T)

From the above results, it is clear that the bigGraph method is more prominent and efficient than other methods for performing the concurrent operations, even on graphs with high number of vertices and edges. This comes from our parallelization method for both read and update queries. The result also allows stating that concurrent operations on large-scale social networks can be processed in the normal and less costly computing systems.

All of codes and test results can be freely accessed from the GitHub link: https://github.com/nnhoa/bigGraph/.

5 Conclusion and Future Works

Concurrent operations on large-scale social networks have been a huge challenge today. We proposed in this research an efficient method with (i) the appropriate data structure (for reducing amount of time accessing the main memory for the graph data by increasing the cache hit rate), (ii) the method for parallelizing the updating operations, and (iii) the best bBFS algorithm by selecting the smaller queue for traversing to reduce the execution time. Another advantage is in the Cilk Plus multi-threaded parallel computing method, which offers a quick and easy way to harness the power of both multi-core and vector processing. Overall, the experiment results confirmed that bigGraph is the most efficient method. It obtained the outstanding performance in comparison with other approaches for executing concurrent operations in parallel. The execution time is reduced proportionally with the number of real parallel threads.

For future works, we aim to extend our method for performing more complex operations on social networks such as computing the relationship distance between two members involving the weight of each member relationship, the influence of a user to the community. The other works might be considered as the graph-based data model enabled Online Transaction Processing.

References

1. Wei, J., Chen, K., Zhou, Y., Zhou, Q., He, J.: Benchmarking of distributed computing engines spark and GraphLab for big data analytics. In: 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService), pp. 10–13 (2016)
2. Yahia, S.A., Benedikt, M., Lakshmanan, L.V.S., Stoyanovich, J.: Efficient network aware search in collaborative tagging sites. PVLDB **1**(1), 710–721 (2008)
3. Mondal, J., Deshpande, A.: Managing large dynamic graphs efficiently. In: Proceedings of the ACM SIGMOD 2012, pp. 145–156 (2012)
4. Gong, M., Li, G., Wang, Z., Ma, L., Tian, D.: An efficient shortest path approach for social networks based on community structure. CAAI Trans. Intell. Technol. **1**(1), 114–123 (2016)
5. Hallac, D., Leskovec, J., Boyd, S.: Network lasso: clustering and optimization in large graphs. In: ACM SIGKDD International Conference on KDD, pp. 387–396 (2015)
6. Nawaz, W., Khan, K.U., Lee, Y.K.: CORE analysis for efficient shortest path traversal queries. In: IEEE Fourth International Conference on Social Graphs, Big Data and Cloud Computing (BdCloud), Sydney, pp. 363–370 (2014)
7. Zhao, H.J., Yiu, M.L., Li, Y., Gong, Z.: Towards online shortest path computation. IEEE Trans. Knowl. Data Eng. **26**(4), 1012–1025 (2014)
8. Scano, G., Huguet, M.J., Ngueveu, S.U.: Adaptations of k-shortest path algorithms for transportation networks. In: 2015 International Conference on Industrial Engineering and Systems Management (IESM), Seville, pp. 663–669 (2015)
9. Chakaravarthy, V.T., Checconi, F., Petrini, F., Sabharwal, Y.: Scalable single source shortest path algorithms for massively parallel systems. In: IEEE 28th International Parallel and Distributed Processing Symposium, pp. 889–901 (2014)

10. Li, Q., Wei, W.: A parallel single-source shortest path algorithm based on bucket structure. In: 2013 25th Chinese Control and Decision Conference (CCDC), pp. 3445–3450 (2013)
11. Leiserson, C.E., Schardl, T.B.: A work-efficient parallel breadth-first search algorithm (or how to cope with the nondeterminism of reducers). In: Proceedings of the Twenty-Second Annual ACM Symposium on Parallelism in Algorithms and Architectures, pp. 303–314 (2010)
12. Banerjee, D.S., Sharma, S., Kothapalli, K.: Work efficient parallel algorithms for large graph exploration. In: 20th Annual International Conference on High Performance Computing, pp. 433–442 (2013)
13. Gonzalez, J.E., Low, Y., Gu, H., Bickson, D., Guestrin, C.: PowerGraph: distributed graph-parallel computation on natural graphs. In: OSDI, pp. 17–30 (2012)
14. Gonzalez, J.E., Xin, R.S., Dave, A., Crankshaw, D., Franklin, M.J., Stoica, I.: GraphX: graph processing in a distributed dataflow framework. In: Proceedings of the 11th USENIX Conference on Operating Systems Design and Implementation (OSDI), pp. 599–613 (2014)
15. Hagberg, A.A., Schult, D.A., Swar, P.J.: Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the 7th Python in Science Conference (SciPy), pp. 11–15 (2008)
16. Du, P.-H., Pham, H.-D., Nguyen, N.-H.: Optimizing the shortest path query on large-scale dynamic directed graph. In: Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2016, pp. 210–216 (2016)
17. Du, P.-H., Pham, H.-D., Nguyen, N.-H.: An efficient parallel method for performing concurrent operations on social networks. In: Nguyen, N.T., Papadopoulos, G.A., Jędrzejowicz, P., Trawiński, B., Vossen, G. (eds.) ICCCI 2017. LNCS (LNAI), vol. 10448, pp. 148–159. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67074-4_15
18. The ACM SIGMOD Programming Contest 2016. http://dsg.uwaterloo.ca/sigmod16contest/
19. H_minor_free: Engineering Distance Queries on Highly Evolving Networks (2016). http://dsg.uwaterloo.ca/sigmod16contest/downloads/H_minor_free-poster.pdf
20. Stanford Large Network Dataset Collection. https://snap.stanford.edu/data/index.html

Author Index