TECHNICAL TRANSACTIONS

CZASOPISMO TECHNICZNE

AUTOMATIC CONTROL

AUTOMATYKA

ISSUE 1-AC (2)

ZESZYT 1-AC (2)

YEAR 2013 (110)

ROK 2013 (110)

WYDAWNICTWO POLITECHNIKI KRAKOWSKIEJ

TECHNICAL TRANSACTIONS | CZASOPISMO TECHNICZNE

AUTOMATIC CONTROL | AUTOMATYKA

ISSUE 1-AC (2) | ZESZYT 1-AC (2)
YEAR 2013 (110) | ROK 2013 (110)

Pierwotną wersją każdego Czasopisma Technicznego jest wersja on-line
www.czasopismotechniczne.pl www.technicaltransactions.com

Automatic Control series

Editorial Board

TECHNICAL TRANSACTIONS | CZASOPISMO TECHNICZNE
AUTOMATIC CONTROL | AUTOMATYKA
1-AC/2013

SZYMON ŁUKASIK*, MARCIN HARĘZA**, MARCIN KACZOR**

DOCUMENT CONTENT MINING FOR AUTHORS' IDENTIFICATION TASK

EKSPLORACJA TREŚCI DOKUMENTÓW W PROBLEMIE IDENTYFIKACJI AUTORÓW

Abstract

This paper deals with automatic authorship attribution through documents content analysis. This approach is based on selecting sets of suitable features relying on specific use of grammar, punctuation or vocabulary and in the next step – executing given classification algorithm. The contribution first overviews various text characteristics which can be employed for that purpose, then presents the results of experiments involving feature selection and examines classifier performance for author identification problem. The paper concludes with discussion and proposals for further research.

Keywords: author identification, feature selection, classification

Streszczenie

Przedmiotem niniejszego artykułu jest problem identyfikacji autora na podstawie analizy treści dokumentów. Podejście to opiera się na wyborze odpowiednich cech związanych ze specyficznym użyciem struktur gramatycznych, interpunkcji oraz słownika, a następnie – użycie wybranego algorytmu klasyfikacji. W artykule przedstawiono najpierw różne charakterystyki tekstu, które mogą być użyte w omawianym zagadnieniu, a następnie załączono wyniki eksperymentów obliczeniowych obejmujących wybór cech i badanie skuteczności klasyfikacji w problemie identyfikacji autorów. Artykuł podsumowano wnioskami oraz propozycjami dalszych prac w rozważanej tematyce badawczej.

Słowa kluczowe: identyfikacja autora, wybór cech, klasyfikacja

* Szymon Łukasik, Ph.D., Department of Automatic Control and Information Technology, Faculty of Electrical and Computer Engineering, Cracow University of Technology; Systems Research Institute, Polish Academy of Sciences.

** Marcin Haręza, Marcin Kaczor, Department of Automatic Control and Information Technology, Faculty of Electrical and Computer Engineering, Cracow University of Technology.

1. Introduction

Author identification is a task commonly performed in historical research, archeology and criminology. Historically it was predominantly considered in the context of handwritten text i.e. taking into account author's writing style. Nowadays, as most of documents are being stored in their electronic versions, it is impossible to complete the identification procedure employing solely graphical features of the text. Moreover, novel types of textual content like computer programming source code, pose new problems, by excluding the possibility of using graphical text representation in the framework of automatic authorship attribution.

The task of selecting author of a given text from a known list of authors can be perceived as a classification or pattern recognition problem. It is a commonly known issue in data mining [23], with a broad range of areas where it transpires, e.g. biometrics, medical diagnostics or intrusion detection systems [6, 14]. Classification is a task of assigning elements from so called testing set, denoted by matrix Y:

$$Y = \left[Y_1 \; Y_2 \; ... \; Y_n \right] \tag{1}$$

which n columns represent features of m_{test} objects belonging to this dataset, to one of the known C classes. Usually a set of representative elements for those classes is additionally given, in the form of a training dataset:

$$X = \left[X_1 \; X_2 \; ... \; X_n \right] \tag{2}$$

with m_{train} elements having class labels explicitly defined. The task for a classifier is to learn how to predict class assignment for testing dataset using knowledge acquired from training set. In case of the authors identification tasks, class label corresponds to author's identifier, training set to a set of documents with known authors and testing set to a group of documents which authorship needs to be identified. To successfully perform such tasks one should select capable classification algorithm and, even more importantly, define suitable document representation, in the form of n distinctive features.

This paper investigates a possibility of employing various non-graphical sets of features for author identification task. Among others we take into account presence of distinctive grammatical structures, quantitative analysis of parts of speech, and diversity of words or punctuation. It is usability of aforementioned features that is experimentally assessed for selected authors identification task.

There exists a vast amount of studies in the area of this contribution, however usually not--involving such a broad range of characteristics being examined at the same time. Previous work in this area involve using: lexical features (e.g. functional words), character features, including alphabetic or digit characters count, uppercase and lowercase characters, letter frequencies, etc. [5], syntactic features [11], semantic features [1] and application specific characteristics, like the use of greetings, signatures, etc. [24]. The problem of language specific issues is also widely studied [4]. Interesting applications of authorship attribution include microblogging posts author identification [15], gender recognition [3] or combining author classification with opinion mining [18]. Accomplished surveys of techniques and strategies commonly employed for authorship attribution tasks can be found in [9, 13, 22].

The paper is organized as follows. First, we give a detailed description of documents characteristics, useful for their representation in authorship attribution task. It is followed by experimental setup, employed to assess the efficiency of classification algorithm used with various documents features sets. Finally, the discussion and proposals for further research are given.

2. Document content representation

Document content will be described here by five groups of features corresponding to the following characteristics of a text, given in its computer-written representation:

a) document grammatical composition, represented by features' group G_R, composed of $R + 9$ characteristics $g_1, g_2, ..., g_{R+9}$ with R being a parameter related to the size of analyzed set of grammatical structures;
b) used punctuation marks, represented by features group P which includes 16 characteristics $p_1, p_2, ..., p_{16}$;
c) words length, given by features group $L : l_1, l_2, ..., l_4$;
d) document formatting defined by features group $F : f_1, f_2, ..., f_5$;
e) words usage described by features group $V : v_1, v_2, ..., v_9$.

The same set of features can be used to characterize numerous documents – belonging to both, training and testing datasets. Albeit, a document D description might be for example given by $D = \{G_R, V\}$ – it would mean that it is characterized only by indicators related to grammatical composition and used words statistics. The study of applicability of different set of features for author's identification task will be conducted in the next Section of the paper. We will now provide a description of features which were assigned to groups listed above.

2.1. Document grammatical composition

Grammar is a linguistic concept concerning both the shape of words and how words (and phrases) can be combined [2]. Grammar, in essence, consists of two components: morphology, i.e. study of how words are formed out of smaller units (morphemes) and the syntax, which is a system of rules specifying how lexical items ought to be composed together [21]. Of those two, when considering writing style analysis, the syntax is of more importance.

To analyze syntactic structure of sentences one should define a notation to conveniently represent content of a given text. Tree structure is commonly used for that purpose. Its branch nodes represent non-terminal syntactic symbols (with sentence as a root) and leaves are equivalent to lexical tokens of the sentence. To obtain structure in this form text needs to be parsed. This language-dependent task is one of the most important in natural language processing. Here, we assume that analyzed text corpus is parsed with popular Stanford parser [12].

To identify a presence of selected tree components, we established 256 most commonly used syntactic structures in English, extracted from the Wall Street Journal corpora contained by Penn Treebank project [19]. It can be found on the website [16]. On that basis, a feature set:

$$g_1, g_2, \ldots, g_R \tag{3}$$

was created. Each characteristic from this set indicates how many times top R structures from the ranking were used in the analyzed document. For example g_1 corresponds to the number of "preposition + noun phrase" structures found in the text. In general, for ranking size R one can assume any value between 1 and 256.

Set of features mentioned above corresponds to the presence of selected grammatical constructs in analyzed document. Next attribute g_{R+1} describes concentration of syntactic elements listed in the ranking. Let $N_R = \sum_{i=1}^{R} g_i$ denote occurrence of constructs from the selected set of R grammatical structures in the given text and let N_G to indicate overall number of structures pointed out by syntactic parser for the analyzed document. Then g_{R+1} can be written as:

$$g_{R+1} = \frac{N_R}{N_G} \tag{4}$$

Subsequent attributes $g_{R+2}, g_{R+3}, ..., g_{R+9}$ describe the occurrence of the individual parts of speech in the text. Here g_{R+2} corresponds to the incidence of nouns, g_{R+3} – pronouns, g_{R+4} – adjectives, g_{R+5} – verbs, g_{R+6} – adverbs, g_{R+7} – prepositions, g_{R+8} – conjunctions and g_{R+9} – interjections.

2.2. Document punctuation

Punctuation is a practice of inserting standardized signs to clarify the meaning and separate language structural units [20]. Among fourteen most commonly used punctuation marks in English, at parsing stage the following symbols are identified here: full stop, exclamation mark, question mark, comma, semicolon, colon, apostrophe, quotation mark, ellipsis, dash, hyphen, slash plus additionally multiple exclamation marks and multiple question marks. Use of those signs is indicated by features group P formed of 16 characteristics.

The first attribute p_1, denoted by:

$$p_1 = \frac{N_P}{N_C} \tag{5}$$

represents number of punctuation marks N_p listed above found in the text, divided by total number of characters N_c.

Second feature from this set p_2 describes variety of punctuation marks, by employing the following formula:

$$p_2 = N_P^* \tag{6}$$

where N_P^* symbolizes overall number of different punctuation signs found in the analyzed document.

Finally, features $p_3, p_4, ..., p_{16}$ refer to the number of times each of punctuation marks listed above was used in the text, scaled by total number of punctuation marks N_p.

2.3. Document word length statistics

Features group L aims at capturing authors' tendency to use short words, long words or in general – words of approximately similar length. Let us denote by S a set of all sentences in the text, and by W_s a set of all words forming a given sentence $s \in S$. Consequently, by size(w) where $w \in W_s$ we understand length of a selected word w while using card(S) and card(W_s) at the same time to represent cardinalities of both sets defined above.

First feature l_1 introduced here is referring to the average word length in the sentences forming the text, and can be written as:

$$l_1 = \frac{\sum_{s \in S} \frac{\sum_{w \in W_s} \text{size}(w)}{\text{card}(W_s)}}{\text{card}(S)} \tag{7}$$

Second feature l_2 describes average word length in the document and is represented by:

$$l_2 = \frac{\sum_{s \in S} \sum_{w \in W_s} \text{size}(w)}{\sum_{s \in S} \text{card}(S)} \tag{8}$$

Two other features from this group refer to the number of short words (consisting of less than 4 characters) and long words (made up of more than 6 characters). Both attributes are relative to the sum of words used in the text, and can be written as follows:

$$l_3 = \frac{\sum_{s \in S} \text{card}(W_s^{\text{short}})}{\sum_{s \in S} \text{card}(W_s)} \tag{9}$$

with $W_s^{\text{short}} = \{w : w \in W_s \wedge \text{size}(w) < 4\}$ and:

$$l_4 = \frac{\sum_{s \in S} \text{card}(W_s^{\text{long}})}{\sum_{s \in S} \text{card}(W_s)} \tag{10}$$

where $W_s^{\text{short}} = \{w : w \in W_s \wedge \text{size}(w) < 4\}$.

2.4. Document formatting

Formatted text, as opposed to plain text, introduces additional styling information. In general it can include colors, font, characters size or other special elements e.g. hyperlinks.

However, most of text repositories comprise only carriage returns or additional soft returns. The following set of features captures writers' individual preferences to use those formatting elements in their documents.

The first feature f_1 considered here, refers directly to the number of paragraphs, i.e. sections of the text with first line being indented. Next attribute f_2 captures average number of sentences included in one paragraph, which can be written as:

$$f_2 = \frac{\text{card}(S)}{N_{\text{par}}} \tag{11}$$

with N_{par} representing total number of paragraphs in the analyzed document.

Features f_3 and f_4 characterize average number of words and characters per paragraph. They can be written as follows:

$$f_3 = \frac{\sum_{s \in S} \text{card}(W_s)}{N_{\text{par}}} \tag{12}$$

and

$$f_4 = \frac{N_C}{N_{\text{par}}} \tag{13}$$

Finally, last feature from this group, f_5 describes writer's tendency to format the text with empty lines. It is represented by relative number of blank lines:

$$f_5 = \frac{N_L{}^E}{N_L} \tag{14}$$

with $N_L{}^E$ being total amount of empty lines in the text and N_L – all lines in the analyzed document.

2.5. Words use statistics

Authors usually differ in the sizes and structures of their vocabularies. Therefore the analysis of vocabulary richness and its concentration could be useful in the context of authorship attribution [7]. Feature set *V* being introduced here aims to quantitatively express those characteristics.

First feature under consideration v_1 refers directly to the number of distinct words used in the text N_v. Five features which follow, employ the concept of *hapax legomenon* (gr. said once) that is words that only occur once in the text [10] and *dis legomenon* – words appearing twice. Let us denote by N_{hl}, N_{dl} number of *hapaxes* found in the document. First five of aforementioned features are now defined as follows:

$$v_2 = \frac{N_{hl}}{\sum_{s \in S} \text{card}(W_s)} \tag{15}$$

$$v_3 = \frac{N_{hl}}{N_v} \tag{16}$$

$$v_4 = \frac{N_{dl}}{\sum_{s \in S} \text{card}(W_s)} \tag{17}$$

$$v_5 = \frac{N_{dl}}{N_v} \tag{18}$$

$$v_6 = \frac{100 \log_{10} \sum_{s \in S} \text{card}(W_s)}{1 - \frac{N_{hl}}{N_v}} \tag{19}$$

with the latter two known from the literature under the names of Sichel (v_5) and Honore (v_6) measures [3].

Defining additional text characteristics – N_{il} representing number of words occurring in the text precisely *i*-times, allows us to formulate three last features in this set, recognized as Yule (v_7), Simpson (v_8) and entropy (v_9) measures [3]:

$$v_7 = 10^4 \left[-\frac{1}{\sum_{s \in S} \text{card}(W_s)} + \sum_{i=1}^{N_v} N_{il} \left(\frac{i}{\sum_{s \in S} \text{card}(W_s)} \right)^2 \right] \tag{20}$$

$$v_8 = \sum_{i=1}^{N_v} N_{il} \frac{i}{\sum_{s \in S} \text{card}(W_s)} \frac{i-1}{\sum_{s \in S} \text{card}(W_s) - 1} \tag{21}$$

$$v_9 = \sum_{i=1}^{N_v} N_{il} \left(-\log_{10} \frac{i}{\sum_{s \in S} \text{card}(W_s)} \right) \frac{i}{\sum_{s \in S} \text{card}(W_s)} \tag{22}$$

3. Experimental studies

Experiments performed were designed to evaluate usefulness of features sets listed above for authorship attribution tasks and to study performance of classifiers employing carefully selected groups of attributes.

Experimental studies were conducted using Thomson Reuters Text Research Collection (known as TRC2 dataset). The dataset includes almost 2 million news reports collected between January 2008 and February 2009. We have chosen randomly documents authored

by 20 writers (around 100 contributions each). The first part consisting of 10 writers contributions was used for first part of experiments – involving feature set evaluation, and will be referred to as the training dataset. The rest of experimental data (labeled as the testing dataset) was employed in the second part of numerical studies, more thoroughly investigating the performance of selected classification algorithms.

As classifiers Naïve Bayes, feed-forward Neural Network with back-propagation learning, classic *k*-Nearest Neighbor (with $k = 3$) and Random Forests were chosen, with the latter two only being employed in the last part of this study. Optimal values of parameters for investigated techniques were determined through a set of pilot runs. For more details on those algorithms one could refer to [6].

3.1. Feature set selection

First we examined usability of features listed in the previous Section for authorship attribution. For that purpose Sequential Forward Search (SFS) or Sequential Backward Search (SBS) algorithm were executed, with *k*-nearest neighbor classifier being used for evaluation. Both techniques constitute similar supervised feature selection paradigms [17]. First algorithm starts with empty candidate feature subset and at each iteration adds the feature which maximizes classifier accuracy. SBS is the opposite strategy – it starts with the entire set of available features, and then iteratively removes one feature at the time, as long as aforementioned measure of accuracy improves, where the feature removed maximizes this improvement [8]. Here, the discriminative power of a given feature set was determined through cross-validation.

Feature selection algorithms were executed with $R = 256$ and five different schemes:

- Basic Feature Selection (BFS) – feature selection is performed on all groups of characteristics listed in Section 2.
- Two Phase Feature Selection (TPFS) – feature selection algorithm is executed on non-grammatical characteristics (all but G_R). Obtained reduced feature set is merged with G_R and feature selection is performed again.
- Parallel Feature Selection (PARFS) – feature selection is conducted in parallel on non-grammatical and grammatical characteristics, obtained reduced feature sets are merged afterwards.
- Grammatical Feature Selection (GFS) – feature selection is conducted only on characteristics from the grammatical ranking.
- Non-Grammatical Feature Selection (NGFS) – feature selection is executed only on characteristics not listed in grammatical ranking.

Table 1 illustrates a number of features obtained for different variants of feature selection. It is evidently observed that forward search principally results in more compact reduced set of characteristics, which is naturally caused by local search nature of this algorithm. Individual characteristics most frequently chosen in experiments involving different feature selection strategies, for both, features listed (top ten) and not listed (top ten) in the grammatical ranking are shown on Fig. 1. It can be observed that most useful features are found within grammatical ranking, especially between features g_{20} and g_{100}. Furthermore features describing incidence of pronouns (g_{R+3}), adverbs (g_{R+6}) and prepositions (g_{R+7}) appeared frequently in the reduced feature sets. Among non-grammatical characteristics ratio of punctuation marks and occurrence of exclamation mark, dash and hyphen were found particularly important; however in general selection ratio for those features seems to be significantly lower.

Table 1

Obtained features set sizes

Feature selection scheme	Sequential feature selection	
	Sequential Forward Search (SFS)	Sequential Backward Search (SBS)
BFS	20	222
TPFS	26	194
PARFS	37	240
GFS	24	233
NGFS	13	7

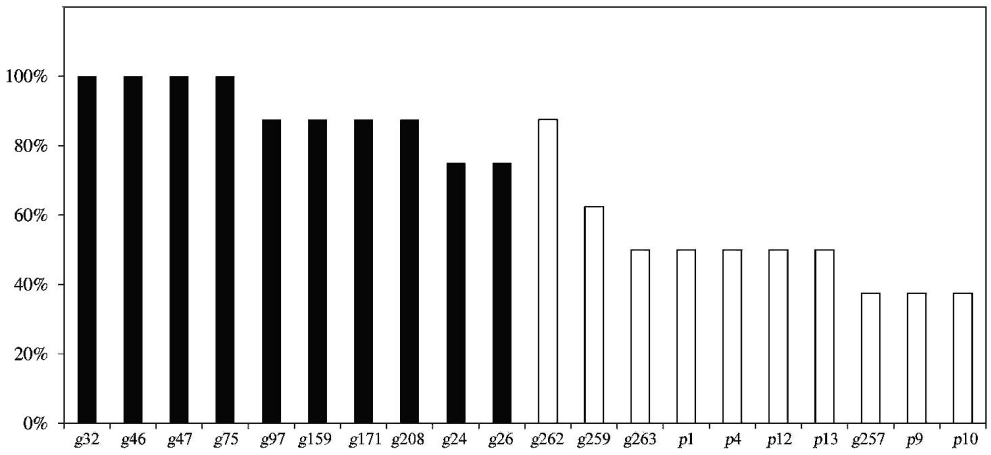


Fig. 1. Ten most frequently chosen characteristics for all variants of feature selection and features listed and not listed in the grammatical ranking

Rys. 1. Dziesięć najczęściej wybieranych charakterystyk dla wszystkich wariantów algorytmu wyboru cech i cech z i spoza rankingu gramatycznego

3.2. Performance analysis

Next series of experiments was devoted to studying performance of selected classifiers with different feature sets obtained through feature selection and varying size of grammatical ranking (where applicable). The trials were executed for training dataset first, to determine the best-performing classifier and the most appropriate feature set, with five runs involving cross-validation and average classification accuracy being reported here.

Tables 2–5 sum up obtained results for *k*-nearest neighbor, Naïve Bayes, feed-forward Neural Network and Random Forests classifiers.

Table 2

Classification accuracy [%] for training dataset and k-Nearest Neighbor classifier

Ranking Size & Sequential Selection Variant / Feature selection variant	*R* = 32		*R* = 64		*R* = 128		*R* = 256	
	SFS	SBS	**SFS**	**SBS**	**SFS**	**SBS**	**SFS**	**SBS**
BFS	42.25	22.50	37.50	32.50	55.25	22.25	56.00	61.25
TPFS	33.00	19.50	46.75	39.75	43.50	26.75	53.75	39.50
PARFS	49.75	53.00	56.25	60.75	61.00	69.25	63.00	70.50
GFS	51.00	50.75	51.25	62.00	61.50	69.75	62.75	75.00
NGFS	38.25	20.25	32.75	22.50	25.00	20.25	38.50	24.25

Table 3

Classification accuracy [%] for training dataset and Naïve Bayes classifier

Ranking Size & Sequential Selection Variant / Feature selection variant	*R* = 32		*R* = 64		*R* = 128		*R* = 256	
	SFS	**SBS**	**SFS**	**SBS**	**SFS**	**SBS**	**SFS**	**SBS**
BFS	37.50	24.00	27.25	22.50	40.50	18.75	39.75	58.75
TPFS	27.00	20.00	40.00	29.75	30.00	23.00	41.00	56.25
PARFS	42.25	42.25	45.75	59.00	48.75	62.25	50.50	55.75
GFS	38.00	44.00	41.75	57.75	43.75	62.25	45.50	56.25
NGFS	38.00	17.00	26.75	19.50	20.75	14.50	36.25	21.50

Table 4

Classification accuracy for training dataset and Neural Network classifier

Ranking Size & Sequential Selection Variant / Feature selection variant	*R* = 32		*R* = 64		*R* = 128		*R* = 256	
	SFS	**SBS**	**SFS**	**SBS**	**SFS**	**SBS**	**SFS**	**SBS**
BFS	41.75	29.25	47.25	46.25	56.75	42.50	45.00	60.50
TPFS	27.75	26.25	48.25	41.25	50.00	33.25	48.75	60.50
PARFS	53.75	42.25	45.00	46.75	50.00	62.75	63.50	60.25
GFS	35.00	43.25	41.00	51.00	52.25	56.50	56.60	61.25
NGFS	38.50	17.00	26.25	11.75	24.25	17.25	35.00	22.25

Table 5

Classification accuracy for training dataset and Random Forests classifier

Ranking Size & Sequential Selection Variant / Feature selection variant	*R* = 32		*R* = 64		*R* = 128		*R* = 256	
	SFS	**SBS**	**SFS**	**SBS**	**SFS**	**SBS**	**SFS**	**SBS**
BFS	66.00	33.50	57.25	51.50	69.25	58.75	68.00	82.00
TPFS	37.75	28.25	61.50	44.25	57.50	47.25	62.25	83.50
PARFS	62.75	64.75	60.25	69.00	69.25	80.50	76.00	79.25
GFS	61.75	63.75	53.75	70.75	71.00	80.50	68.00	80.75
NGFS	44.25	27.75	36.75	24.25	34.50	23.25	46.25	32.00

It can be seen that employing a broad range of grammatical features is crucial for obtaining high classification accuracy. Random Forests classifier was found to be the best performing one. Among various schemes of feature selection the most successful were Two Phase Feature Selection, Basic Feature Selection and Grammatical Feature Selection.

Finally selected best-performing classifiers based on Random Forests and *k*-Nearest Neighbor were more thoroughly evaluated for both datasets. We used set of characteristics with $R = 256$ and Sequential Backward Search along with Two Phase Feature Selection or Grammatical Feature Selection (for *k*-Nearest Neighbor). Results obtained for 10 runs and both training and testing datasets, with 5-fold cross-validation are shown on Fig. 2.

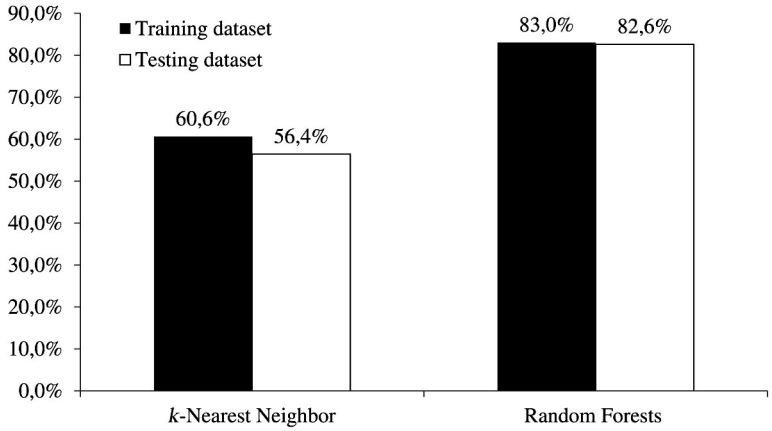


Fig. 2. Classification accuracy for training and testing datasets

Rys. 2. Trafność klasyfikacji dla zbiorów uczącego i testującego

4. Conclusion

This contribution examined the possibility of employing various characteristics of documents in computer-written form for authorship attribution. The suitability of several features was considered for parsed instances of news reports, taking into account a possibility of using few discrimination algorithms for conclusive classification. It was established here, that for ensuring proper classifier performance the most important ones are those based on occurrence of grammatical structures. We also identified tree-based classifiers as most promising in terms of accuracy – for both training and testing datasets. In general, authors' classification for selected features and both instances proved to be reasonably accurate. It can be noted that presented approach can be used for different languages – provided that suitable parsing procedure could be conducted beforehand.

Further work in the research area of this contribution will involve employing genetic feature set selection. Supplementary experimental studies on effectiveness of various classifiers are planned as well. The possibility of using authorship attribution techniques for identifying source code creator constitutes likewise a promising area of upcoming research.

The study is co-funded by the European Union from resources of the European Social Fund. Project PO KL "Information technologies: Research and their interdisciplinary applications", Agreement UDA-POKL.04.01.01-00-051/10-00.

References

[1] Argamon S., *Levitan S., Measuring the Usefulness of Function Words for Authorship Attribution*, Proc. Joint Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing, paper no 162, 2005.

[2] Broccias C., *Cognitive linguistic theories of grammar and grammar teaching*, [in:] De Knop S., De Rycker T. (eds.), *Cognitive Approaches to Pedagogical Grammar*, Walter de Gruyter, Berlin 2008, 67-90.

[3] Cheng N. Chandramouli R., Subbalakshmi K.P., *Author gender identification from text*, Digital Investigation, vol. 8, 2011, 78-88.

[4] Eder M., Rybicki J., *Do birds of a feather really flock together, or how to choose training samples for authorship attribution*, Literary and Linguistic Computing /to appear.

[5] Grieve J., *Quantitative authorship attribution: An evaluation of techniques*, Literary and Linguistic Computing, vol. 22, 2007, 251-270.

[6] Hastie T., Tibshirani R., Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York 2009.

[7] Hoover D.L., *Another Perspective on Vocabulary Richness*, Computers and the Humanities, vol. 37, 2003, 151-178.

[8] Jagadev A.K, Devi S., Mall R., *Soft Computing for Feature Selection*, [in:] Dehuri S., Cho, S.-B. (eds.), *Knowledge Mining using Intelligent Agents*, Imperial College Press, London 2011, 217-258.

[9] Jockers M.L., Witten D.M., *A comparative study of machine learning methods for authorship attribution*, Literary and Linguistic Computing, vol. 25, 2010, 215-223.

[10] Jurafsky D., Martin J.H., *Speech And Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall, New Jersey 2009.
[11] Karlgren J., Eriksson G., *Authors, genre, and linguistic convention*, Proc. SIGIR Workshop on Plagiarism Analysis, Author ship Attribution, and Near-Duplicate Detection, 2007, 23-28.
[12] Klein D., Manning C.D., *Accurate Unlexicalized Parsing*, Proceedings of the 41st Meeting of the Association for Computational Linguistics, 2003, 423-430.
[13] Koppel M. Schler J., Argamon S., *Authorship attribution in the wild*, Language Resources and Evaluation, vol. 45, 2011, 83-94.
[14] Kowalski P.A., *Procedure of feature extraction from face image for biometrical system* (in Polish), Technical Transactions, vol. 1-AC/2012, Cracow University of Technology Press, 55-79.
[15] Layton R., *Authorship Attribution for Twitter in 140 Characters or Less*, Cracow University of Technology Press, Proc. 2nd Cybercrime and Trustworthy Computing Workshop, 2010, 1-8.
[16] Łukasik S., Haręza M., Kaczor M., *Grammatical structures ranking (supplementary material)*, http://www.pk.edu.pl/~szymonl/nauka/Author_suppl.pdf (date of access: 9.10.2013).
[17] Łukasik S., Kulczycki P., *Using topology preservation measures for high-dimensional data analysis in a reduced feature space* (in Polish), Technical Transactions, vol. 1-AC/2012, Cracow University of Technology Press, 5-15.
[18] Panicheva P., Cardiff J., Rosso P., *Personal Sense and Idiolect: Combining Authorship Attribution and Opinion Analysis*, [in:] Proc. International Conference on Language Resources and Evaluation, Valletta, paper no. 10.491, 2010.
[19] Penn Treebank Project, http://www.cis.upenn.edu/~treebank/
[20] Punctuation, [in:] Merriam-Webster's Collegiate Dictionary: Eleventh Edition, 1009, Merriam-Webster, Springfield 2004.
[21] Radford A., *Minimalist Syntax: Exploring the structure of English*, Cambridge University Press, Cambridge 2004.
[22] Stamatatos E., *A survey of modern authorship attribution methods*, Journal of the American Society for Information Science and Technology, vol. 60, 2009, 538-556.
[23] Wang L.P, Fu X.J., *Data Mining with Computational Intelligence*, Springer, Berlin 2005.
[24] Zheng R., Li J., Chen H., Huang Z., *A framework for authorship identification of online messages: Writing style features and classification techniques*, Journal of the American Society of Information Science and Technology, vol. 57, 2006, 378-393.

TECHNICAL TRANSACTIONS | **CZASOPISMO TECHNICZNE**
AUTOMATIC CONTROL | AUTOMATYKA
1-AC/2013

MAŁGORZATA CHARYTANOWICZ[*], HENRYK CZACHOR[**], JERZY NIEWCZAS[***]

NONPARAMETRIC REGRESSION APPROACH: APPLICATIONS IN AGRICULTURAL SCIENCE

ZASTOSOWANIE REGRESJI NIEPARAMETRYCZNEJ W NAUKACH ROLNICZYCH

A b s t r a c t

In this paper, a method for determining the soil pore size distribution, constituting the subject of the presented investigations, is proposed. A research study was conducted using image analysis algorithms, and in turn, nonparametric statistical techniques. The results and further work will be discussed in section four. The purpose of this investigation is to discover the relationship between the pore size and volume of the corresponding pores. The algorithm presented here is based on the theory of statistical kernel estimators. This frees it of assumptions in regard to the form of regression function. The approach is universal, and can be successfully applied for many tasks in data mining, where arbitrary assumptions concerning the form of regression function are not recommended.

Keywords: nonparametric regression, kernel estimators, morphological image processing, closing procedure, pore size distribution, pore space, total porosity, soil structure, aggregation, soil compaction

S t r e s z c z e n i e

Celem niniejszego artykułu jest zaprezentowanie procedury wyznaczania rozkładu wielkości porów w agregatach glebowych. Do scharakteryzowania zależności pomiędzy badanymi zmiennymi wykorzystana zostanie funkcja regresji. W przeprowadzonych badaniach zastosowano algorytmy analizy obrazów cyfrowych oraz metodykę statystycznych estymatorów jądrowych. Przedstawiona metoda umożliwia uzyskanie właściwej charakterystyki rozkładu wielkości porów i może stanowić efektywne narzędzie stosowane w wielu zagadnieniach eksploracji danych. Jako model nieparametryczny, nie wymaga założeń dotyczących kształtu zależności funkcyjnej między rozpatrywanymi zmiennymi.

Słowa kluczowe: regresja nieparametryczna, estymatory jądrowe, przekształcenia morfologiczne, operacja zamknięcia, rozkład wielkości porów, porowatość ogólna gleby, struktura gleby, agregacja, gęstość gleby

[*] Małgorzata Charytanowicz, Ph.D., Institute of Mathematics and Computer Science, The John Paul II Catholic University of Lublin; System Research Institute, Polish Academy of Sciences.

[**] Henryk Czachor, D.Sc., Ph.D., Bohdan Dobrzański, Institute of Agrophysics, Polish Academy of Sciences.

[***] Jerzy Niewczas, D.Sc., Ph.D., Institute of Mathematics and Computer Science, The John Paul II Catholic University of Lublin.

1. Introduction

Pore size distribution is one of many physical measurements characterizing soil structure as far as plant growth is concerned. A number of scientists have reported studies of pore space as a general method for defining soil properties. In this respect, a complete analysis of the soil pore size distribution is used for predicting the gas diffusivity, water infiltration rates, water availability to plants, water-storage capacity and movement of water.

The most common measure characterizing the fraction of the pore space within a solid is the total porosity, defined by the ratio:

$$\varphi = \frac{V_P}{V_T} \tag{1}$$

where:

V_p – the volume of void-space,

V_T – the total volume of soil material, including the solid and void components.

Porosity is a dimensionless quantity and can be reported either as a decimal fraction or as a percentage. Being simply a fraction of total volume, it can range between 0 and 1, typically falling between 0.3 and 0.7 for most soils. This provides a more useful physical description of an particular soil, such as providing an estimate of compaction and the maximum space available for water. Moreover, a number of scientists have reported that studies of pore size distribution are useful as a general method for defining the soil structure. Pore sizes usually have traditionally been divided into macropores and micropores, with the division between the two being arbitrary. Because of the relationship between the pore properties and the reaction of chemicals in the soil, a more detailed pore-fraction analysis seems warranted [7, 15].

The purpose of this investigation is to elaborate a method of measuring the pore size distribution in the soil. Internal difference in porosity within an aggregate can be visualized by microtomography scanning of the air-dried samples. Once scanned, the computed tomography information allows the non-destructive visualization of slices, arbitrary sectional views and pseudo-color representations [14].

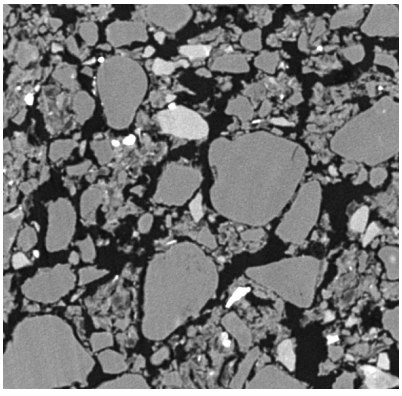


Fig. 1. Cross section of a typical soil with pore space in black

Rys. 1. Przekrój agregatu glebowego, pory zostały wyróżnione kolorem czarnym

For image processing, the authors used a program for computer image analysis package. After preprocessing methods, morphological operation closing, consisting of a dilation followed by erosion, was used to fill in holes and small gaps without changing the size and original boundary shape [5, 6, 12, 17]. Subsequently, digital images were segmented into pore space and solid. The data derived automatically from images was then statistically examined, as nonparametric statistical regression allowed for the determination of the soil pore distribution.

2. Material and Methods

These studies were conducted using soil aggregates from the cultivated soil layer explored at the Institute of Agrophysics, of the Polish Academy of Sciences in Lublin. The direct and nondestructive analyses of internal soil aggregate structures were detected using computed tomography equipment Nanotom 800, with the voxel-resolution of 2.5 microns per volume pixel [4, 8]. Three types of aggregates, different in terms of fertilization, denoted as aggregates 0 – without fertilization, NPK – mineral fertilization, and OB – pig manure, were studied. Tomography sections were processed using macros writing in the Aphelion 4.0.1 package.

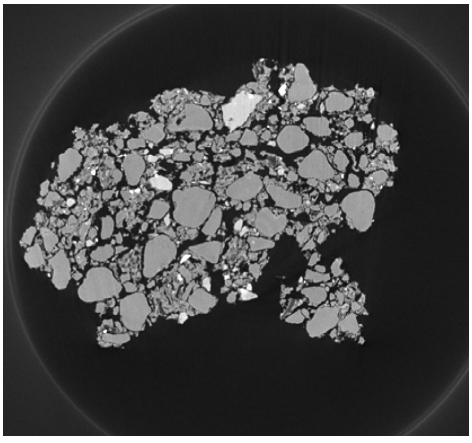


Fig. 2. The soil aggregate microtomographic image

Rys. 2. Obraz tomograficzny agregatu glebowego

The entire procedure was composed of the following steps. Firstly, grayscale images were preprocessed by removing ring artifacts using the ROI method. After selecting the region of interest, the automatic Otsu binarization method was then employed.

Subsequently, binary morphological closing with increasing size of square structuring element (starting with size 2), was processed. In each step, the source image was subtracted from the target image, and the result volumes were listed, giving soil aggregate pore distribution. The operation was repeated until all pores were filled. Subtraction of the transformed image from the original image gives the total pore volume in the sample soil pores. Finally, the pore fractions analysis was performed using the regression approach [3].

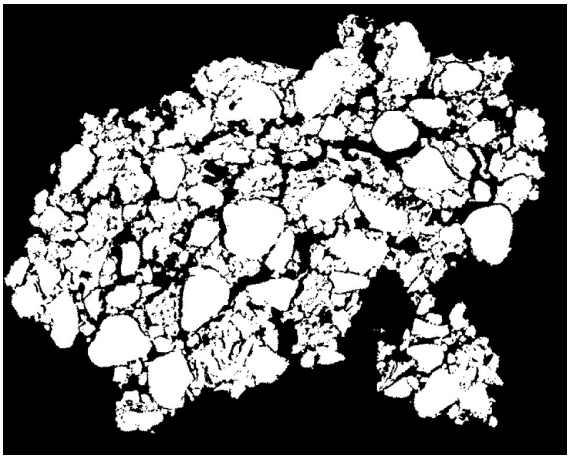


Fig. 3. The soil aggregate image after selecting ROI and binarization through using Otsu method

Rys. 3. Obraz agregatu glebowego po zastosowaniu selekcji ROI i binaryzacji metodą Otsu

The procedure for determining the soil pore size distribution
1. X-ray microtomographic image analysis of aggregates.
2. Ring artifact removal using – the ROI method.
3. Image binarization – the OTSU method.
4. Determining the pore size distribution – binary morphological closing.
5. Pore fractions analysis – regression approach.

In next section, the regression approach is shortly described. A simple linear regression model is not appropriate for the data. Therefore nonlinear estimation and nonparametric methods were concerned.

3. The regression analysis

The validation of the efficiency and accuracy of regression technique was explored by comparing results on a variety of real datasets. Both a classical parametric regression model and several nonparametric methods were examined as far as the pore size distribution was concerned.

The best fit to the data in the family of nonlinear estimation was revealed for two models: polynomial with the power of three; and logarithmic, for which the determination coefficients were in range 80–90%. The proposed regression functions did not well discover the properties of the pore size distributions, especially on the left side, because of their positively skewed and unimodal character. Therefore, the nonparametric kernel method was recommended [1, 9, 10].

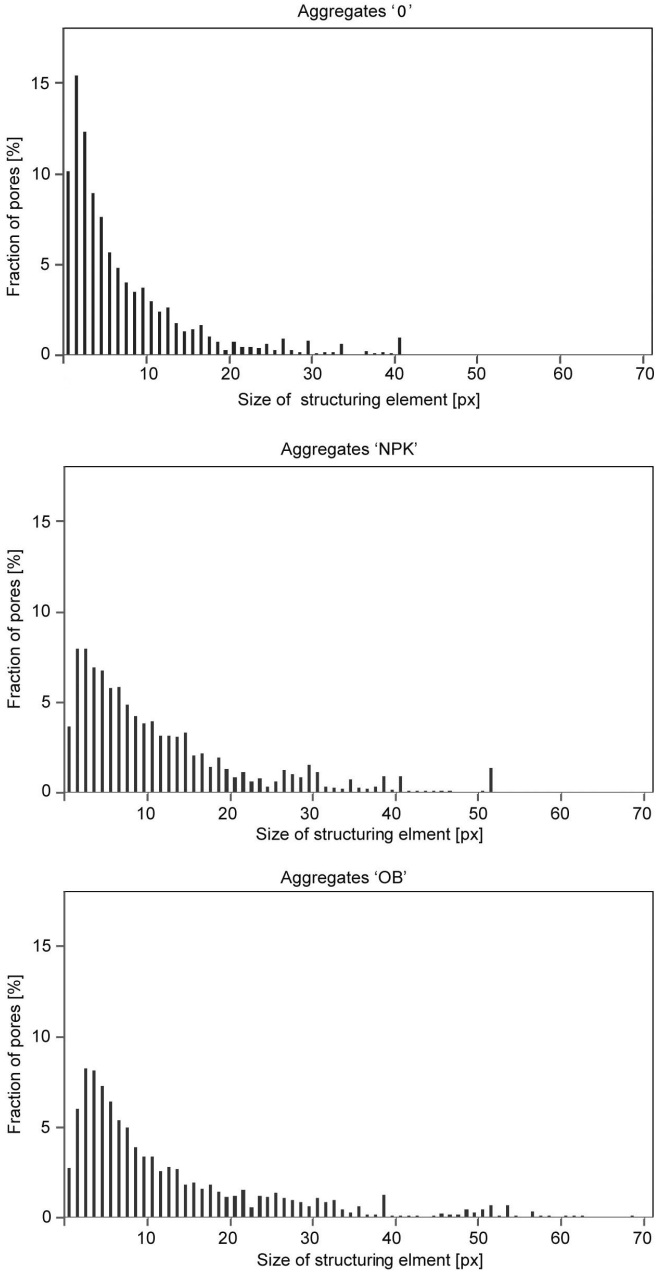


Fig. 4. The pore size distribution: the data presents three types of aggregates different in terms of fertilization, denoted as aggregates: 0 – without fertilization, NPK – mineral fertilization, and OB – pig manure

Rys. 4. Rozkłady porów, wykresy przedstawiają trzy typy agregatów: 0 – bez nawożenia, NPK – nawóz mineralny, OB – nawóz naturalny

Classical parametric methods of determining an appropriate functional relationship between the two variables impose arbitrary assumptions concerning the functional form of the regression function. Moreover, the choice of parametric model depends very much on the situation. If a chosen parametric family is not of appropriate form, then there is a danger of reaching incorrect conclusions in the regression analysis. This also makes it difficult to take into account the whole of the accessible information. However, the rigidity of this regression can be overcome by removing the restriction that the model is parametric. This approach leads to nonparametric regression. This lets the data decide which function fits them best. In this study, a class of kernel–type regression estimators called 'local polynomial kernel estimators' is presented.

Let therefore, m elements $(x_i, y_i) \in R \times R$, $i = 1, 2,..., m$ be given, where values x_i may designate some non-random numbers or realizations of the one-dimensional random variable X, whereas y_i designate realizations of the one-dimensional random variable Y. Assuming the existence of the function $f: R \rightarrow R$ having a continuous first derivative that:

$$y_i = f(x_i) + \varepsilon_i \tag{2}$$

where ε_i are independent random variables with zero mean and unit finite variance. Let then $p \in N$ be the degree of the polynomial being fit. The kernel regression estimator $\hat{f}: R \rightarrow R$, obtained by using weighted least squares with kernel weights, is given by the formula:

$$\hat{f}(x) = e(X^T W X)^{-1} X^T W y \tag{3}$$

where:

$$y = [y_1, y_2, \ldots, y_m]^T \tag{4}$$

is the vector of responses,

$$X = \begin{bmatrix} 1 & x_1 - x & (x_1 - x)^2 & \cdots & (x_1 - x)^p \\ 1 & x_2 - x & (x_2 - x)^2 & \cdots & (x_2 - x)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m - x & (x_m - x)^2 & \cdots & (x_m - x)^p \end{bmatrix} \tag{5}$$

is an $m \times (p + 1)$ design matrix, and:

$$W = \operatorname{diag}\left(\frac{1}{h} K\left(\frac{x_1 - x}{h} \right), \frac{1}{h} K\left(\frac{x_2 - x}{h} \right), \ldots, \frac{1}{h} K\left(\frac{x_m - x}{h} \right) \right) \tag{6}$$

is an $m \times m$ diagonal matrix of kernel weights, while:

$$e = [1, 0, \ldots, 0]^T \tag{7}$$

is the 1 × (p + 1) vector having 1 in the first entry and zero elsewhere. The coefficient $h > 0$ is called 'a bandwidth', while the measurable function $K : R \to [0, \infty)$ of unit integral, symmetrical with respect to zero, and having a weak global maximum in this place, takes the name of the kernel.

An important problem is the choice of the parameter p. For sufficiently smooth regression functions, the asymptotic performance of $\hat{f}$ improves for higher values of p. However, for higher p, the variance of the estimator becomes larger, and in practice, a very large sample may be required. On the other hand, the even degree polynomial kernel estimator has a more complicated bias expression which does not lend itself to simple interpretation. These facts suggests the use of either $p = 1$ or $p = 3$. Moreover, for $p = 1$, the convenient explicit formulae exists:

$$\hat{f}(x) = \frac{1}{mh} \sum_{i=1}^{m} \frac{\left(\hat{s}_2(x) - \hat{s}_1(x)(x_i - x)\right) y_i K\left(\frac{x - x_i}{h}\right)}{\hat{s}_2(x)\hat{s}_0(x) - \left(\hat{s}_1(x)\right)^2} \tag{8}$$

where:

$$\hat{s}_r(x) = \frac{1}{mh} \sum_{i=1}^{m} (x_i - x)^r K\left(\frac{x_i - x}{h}\right) \text{ for } r = 0, 1, 2 \tag{9}$$

Therefore, except in more advanced statistical applications, $p = 1$ is preferred.

The choice of the kernel form has no practical meaning and thanks to this, it is possible to take into account the primarily properties of the estimator obtained. Most often, the standard normal kernel is expressed by the convenient analytical formula:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{10}$$

is used.

The practical implementation of the kernel regression estimators requires a good choice of bandwidth. If h is too small, a spiky rough kernel estimate is obtained, and if h is too large, it results in a flat kernel estimate. A frequently used bandwidth selection technique is the 'cross-validation method' introduced by Clark [2], which chooses h to minimize:

$$CV(h) = \frac{1}{m} \sum_{i=1}^{m} \left(y_i - \hat{f}_{-i}(x_i)\right)^2 \tag{11}$$

where:

$\hat{f}_{-i}(x_i)$ – the leave-one-out kernel regression estimator based on data $(x_1, y_1), \ldots, (x_{i-1}, y_{i-1}), (x_{i+1}, y_{i+1}), \ldots, (x_m, y_m)$.

A typical regression curve obtained by the kernel regression method for m = 200 data points generated from function $f(x) = 0{,}1 \cdot \sin(5x - 2) + e^{-(x-1)^2/4}$ as defined on interval [0, 3] is demonstrated on Fig. 5. A standard normal kernel K given by rule (10) and bandwidth h calculated by the cross-validation method (11) were used.

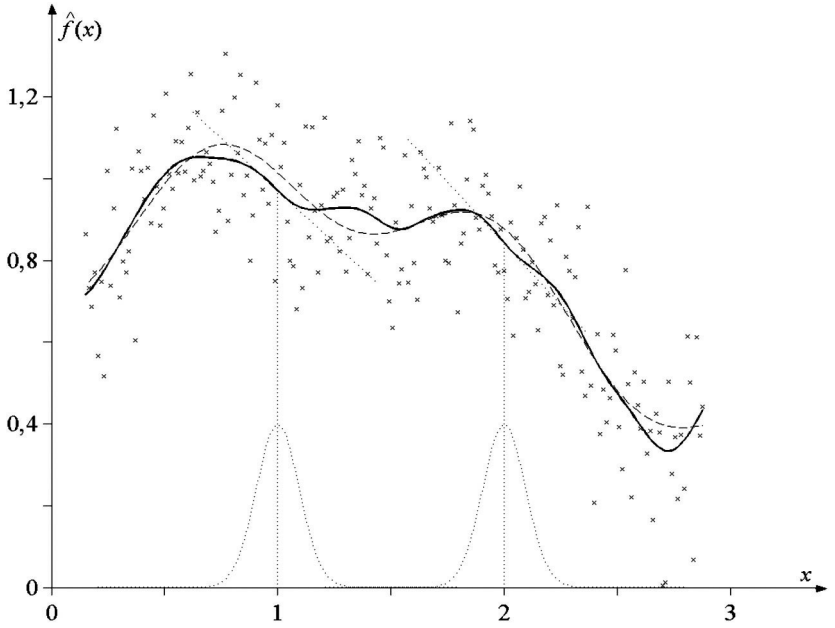


Fig. 5. The estimated regression: the data points are represented by a cross, the true function curve by a dashed line, the regression curve by a solid line, and kernels for arguments $x = 1$ and $x = 2$ by a dotted line

Rys. 5. Jądrowy estymator funkcji regresji: wartości próby losowej oznaczono krzyżykami, estymowaną funkcję linią przerywaną, jądrowy estymator funkcji regresji linią ciągłą, jądra dla argumentów $x = 1$ i $x = 2$ linią kropkowaną

The tasks concerning the choice of the kernel form, the bandwidth, as well as additional procedures improving the quality of the estimator obtained, are found in [13, 16]. The utility of local linear kernel estimators has been investigated in the context of some typical data derived from the soil aggregate images.

4. Results and discussion

The main aim of this research is to discover the relationship between the pore size and volume of the corresponding pores in soil aggregates. The data obtained by the proposed method, based on the image processing algorithms, allows the use of regression approach.

The kernel regression built upon a weighted local linear regression was used to analyze the soil pore size distribution. For ease of computation, the standard normal kernel (10) was used. The bandwidth was determined using the cross–validation method (11).

Thus, particle-size distribution was translated into an equivalent regression model, which in turn, is related to characterize water retention, crucial in any modeling study on water flow and solute transport in soil. This can be used for comparing and predicting some specific points of interest of the water retention characteristics.

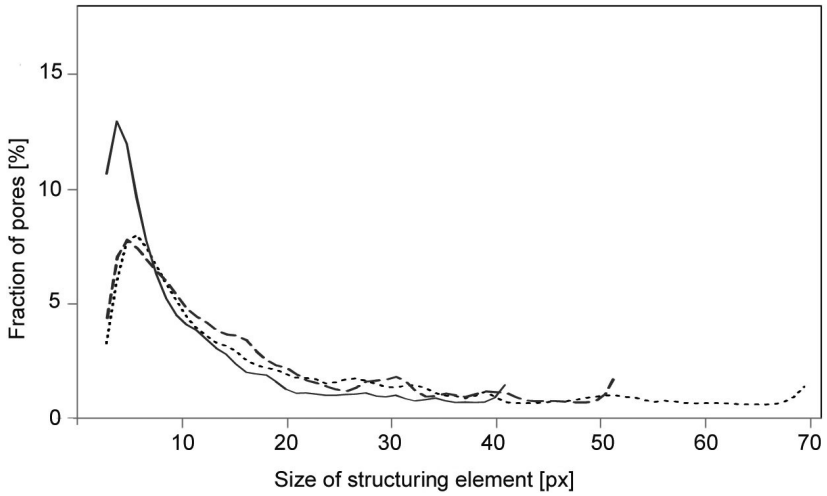


Fig. 6. The estimated regression: the data presents three types of aggregates different in terms of fertilization, denoted as aggregates 0 – without fertilization (a solid line), NPK – mineral fertilization (a dashed line), and OB – pig manure (a dotted line)

Rys. 6. Jądrowy estymator funkcji regresji dla trzech typów agregatów: 0 – bez nawożenia (linia ciągła), NPK – nawóz mineralny (linia przerywana), OB – nawóz naturalny (linia kropkowana)

Fraction of total aggregate volume occupied by pores was significantly greater in OB aggregates. There was also greater percentage of large pores in OB aggregates than in 0 or NPK aggregates. In the smallest range, however, porosity of 0 aggregates exceeded that of NPK and OB aggregates. Total porosities were higher for both NPK and OB aggregates (24% and 32%, correspondingly) than the 14% for 0 aggregates.

A combination of higher microporosity and higher percent of large pores in OB aggregates may generate more favorable conditions for microbial activity through a combination of high water-holding capacities, increased aeration and gas transport. Our current aim is comparing and relating specifics of internal pore structures in the aggregates from their water stability, composition and chemical properties. Water stable aggregates are much more porous in relation to the non stable aggregates.

Soil aggregation and its maintenance is very important for sustainable agriculture because it impacts majority of biological and physical soil properties and processes. This directly promotes better movement of air and of water in the soil, prevents runoff and soil water erosion, and indirectly determines plant growth. However soil aggregates are very sensitive against water and mechanical stresses. Discovering the main factors determining their water stability is a great challenge for worldwide agriculture which would help to prevent soil degradation and to increase a carbon sequestration in soils and their fertility.

The aim of future research is to discover the differences between the inter aggregate pore structure of water stable and non stable aggregates [8]. Additional physico-chemical parameters obtained by chromatography analysis are going to be used.

5. Conclusions

Recent advances in computed tomography and digital image processing algorithms provide technologically advanced measurement tools for studying the internal structures of soil aggregates. This seems very useful in characterizing the pore size distribution and in quantifying the differences in pore structures of the aggregates from the different types of soil.

A more detailed analysis can be obtained by deriving various methods to quantify the pore structure and developing a pore size-distribution curve to predict retention properties. The proposed algorithm, based on image analysis and kernel estimator methodology, is expected to be an effective technique for various agricultural studies.

Nowadays, kernel regression is a common tool for empirical studies in many research areas. This is also a consequence of the fact that kernel regression techniques are provided by many software packages.

The kernel regression approach is also common in image processing and reconstruction methods. Quite recently, kernel regression has experienced a kind of revival in Earth sciences on estimating some characteristics of the distribution in nonparametric models [11].

References

[1] Charytanowicz M., Kulczycki P., *Nonparametric Regression for Analyzing Correlation between Medical Parameters*, [in:] Pietka E., Kawa J. (eds), *Advances in Soft Computing – Information Technologies in Biomedicine*, Springer-Verlag, Berlin, Heidelberg 2008.

[2] Clark R.M., *Non-Parametric Estimation of a Smooth Regression Function*, Journal of the Royal Statistical Society, Series B 39, 1977, 107-113.

[3] Draper N.R., Smith H., *Applied regression analysis*, John Wiley and Sons, New York 1981.

[4] Gonet S., Czachor H., *Organic carbon and humic substances fractions in soil aggregates* 16th Meeting of the International Humic Substances Society, Functions of the Natural Organic Matter in Changing Environment, China, Hangzhou, 9-14.09.2012, 215-217, 2012.

[5] Gonzalez R.C., Woods R.E., *Digital Image Processing*, Prentice-Hall Inc., New Jersey 2002.

[6] Kowalski P., *Procedura ekstrakcji cech z obrazu twarzy dla potrzeb systemu biometrycznego*, Technical Transactions, vol. 1-AC/2012, Cracow University of Technology Press, 2012, 55-79.

[7] Kravchenko, A., Chun H.C., Mazer M., Wang W., Rose J. B., Smucker, A., Rivers M., *Relationships between intra-aggregate pore structures and distributions of Escherichia coli within soil macro-aggregates*, Applied Soil Ecology, vol. 63, 2013, 134-142.

[8] Król A., Niewczas J., Charytanowicz M., Gonet S., Lichner L., Czachor H., Lamorski K., *Water stable and non stable soil aggregates and their pore size distributions*, 20th International Poster Day and Institute of Hydrology Open Day "Transport of water, chemicals and energy in the soil – plant – atmosphere system", 2012, 870-871.

[9] Kulczycki P., *Estymatory jądrowe w analizie systemowej*, WNT, Warszawa 2005.

[10] Kulczycki P., *Kernel estimators in industrial applications*, [in:] Prasad B. (ed) Soft Computing Applications in Industry, Springer-Verlag, Berlin 2008.
[11] Perzanowski K.A., Wołoszyn-Gałęza A., Januszczak M., *Indicative factors for European bison refuges in the Bieszczady Mountains*, Ann. Zool. Fennici 45, 2008, 347-352.
[12] Pratt W.K., *Digital Image Processing*, John Wiley and Sons, New York 2001.
[13] Silverman B.W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London 1986.
[14] Smith J.R., Chang S.F., *VisualSeek: A fully automated content-based image query system*, Proc. ACM Multimedia Conf., 1996, 87-98.
[15] Van der Weerden T., Klein C., Kelliher F., *Influence of pore size distribution and soil water content on N_2O response curves*, 19th World Congress of Soil Science, Soil Solutions for a Changing World, 1–6 August 2010, Brisbane, Australia 2010.
[16] Wand M.P., Jones M.C., *Kernel Smoothing*, Chapman and Hall, London 1994.
[17] Wayne L.W.C., *Mathematical Morphology and Its Applications on Image Segmentation*, MS thesis, Dept. of Computer Science and Information Engineering, National Taiwan University, Taiwan 2000.

TECHNICAL TRANSACTIONS | CZASOPISMO TECHNICZNE
AUTOMATIC CONTROL | AUTOMATYKA
1-AC/2013

DOMINIKA GOŁUŃSKA*, MAŁGORZATA HOŁDA**

THE NEED OF FAIRNESS IN GROUP CONSENSUS REACHING PROCESS IN A FUZZY ENVIRONMENT

KONCEPCJA „SPRAWIEDLIWOŚCI" W PROCESIE OSIĄGANIA KONSENSUSU W GRUPIE W WARUNKACH ROZMYTOŚCI

Abstract

In this paper we explain the need of "fairness" in the human-consistent computational system which supports a group consensus reaching process. We propose the model which combines mathematical approach based on fuzzy environment, i.e. "soft" consensus developed by Kacprzyk and Fedrizzi and socio-psychological approach concerning fairness component. We view fairness from two possible perspectives: a fair distribution and a fair final decision. Finally, we confirm our assumptions by observations in the analyzed groups of students.

Keywords: group consensus reaching process, decision support systems, fairness, fair resource allocation, soft consensus

Streszczenie

W niniejszym artykule wyjaśniono potrzebę „sprawiedliwości" w systemie obliczeniowym wspomagającym proces osiągania konsensusu w grupie decydentów. Autorki zaproponowały model łączący podejście matematyczne oparte na środowisku rozmytym oraz czynniki społeczno-psychologiczne wyjaśniające opisywaną koncepcję. Pojęcie „sprawiedliwości" rozpatrywane jest tu w dwóch kategoriach: rozkładu zasobów oraz decyzji ostatecznej. Założenia poparte są wnioskami na podstawie obserwacji grup studentów.

Słowa kluczowe: proces osiągania konsensusu w grupie, systemy wspomagania decyzji, sprawiedliwość, sprawiedliwy rozkład zasobów, miękki konsensus

* Dominika Gołuńska, M.Sc., Department of Automatic Control and Information Technology, Faculty of Electrical and Computer Engineering, Cracow University of Technology; PhD Studies, Systems Research Institute, Polish Academy of Sciences.

** Małgorzata Hołda, Ph.D., Department of Automatic Control and Information Technology, Faculty of Electrical and Computer Engineering, Cracow University of Technology.

1. Introduction

Decision theory is an interdisciplinary domain which combines research from many disciplines, i.e. psychology, sociology, economics, philosophy, political science, etc. The formal direction can not be the only course of decision making problems since all the classical methods have a very limited capacity for explaining empirical choices.

Regardless of its origin, the essence of decision making process is always the same: there are some options to choose between and the choice is made in a goal-directed way [6]. In fact, many different models of decision making process occurred and enriched an analysis of human behavior, social interactions and other socio-economic descriptions depending on the respective purpose. All of these novel agent-based computational models appeared in order to make the process more human-consistent and believable. That is the reason why we decided to apply psychological and sociological theories to investigate and design systems in this research topic.

By the same token, the concept of fairness appears to be a multifarious issue which draws upon ideas from a whole panoply of scientific disciplines. The prevalent aspect, however, is the impossibility to achieve an ideally fair distribution of power in the decision making process. Deviations of behavior from the presumed results in decision making process suggest and also confirm that fairness influences decision making process.

Groups of individuals are known to be effective organs in decision making process. In spite of several dysfunctions of group work, there still are more crucial benefits (process gains). Namely, groups are better than individuals at understanding problems, at catching errors, so they provide learning. Moreover, a group has more information than any member and can combine this knowledge to derive better solutions and stimulate the creativity of participants and the process itself. Hence, the *group decision making process* will be the groundwork of our further consideration. Considering different types of groups [7] in decision making problem, we took into account *interpersonal orientation group*. It means that one in which the final solution of the problem is only a minor goal. Here, the priority is to ensure a good relation among group members during decision making process and to achieve consensus in the sense of some satisfactory agreement as to the chosen option.

We want to guarantee an equal participation of all decision making members in the consensus reaching process. In most cases, there is also a small group of outsiders who are isolated in their opinions as to the rest of the group and are omitted. Significantly, outsiders do not feel the satisfaction of discussion, what affects the effectiveness of an entire group. Of course, it does not exclude the final decision achievement, but decreases the opportunity of many further activities, i.e. practical implementation of the final decision, survival of the group in the long time period, etc.

All of the socio-psychological aspects forced us to seek a novel approach of consensus degree which will consider the satisfaction of every individual throughout the consensus reaching process. In Section two we will briefly present the overall structure of consensus reaching process. Section three shows an explanation of group decision support systems with a wider description of a moderator who plays here the main role. Furthermore, we attempt to reduce the complexity of the proposed system with a detailed description of relevant aspects only. In the previous work [3], we mainly focused on performance of the core of multistage model of consensus reaching process under fuzziness. In the light of the fact that most human behaviors have not been formalized mathematically yet, our purpose in this paper

is to get a better understanding of how social mechanism in group decision support systems works. Therefore, in Section four we will name some psychological studies and sociological explanations of actual fair behavior and confirm them with our observations in the analyzed groups of students. For comparison, we investigated the problematics of fairness in its versatility on the example of a study of two groups of students. An analysis of the problem of fairness in the groups of students undergoing an examination shows the manifold factors involved in the assessment of what kind of behavior is fair and what kind is not, and how to achieve fairness in decision making process. Our proposal concerning division of fairness approach in two possible directions: a fair distribution and a fair final decision is described in Section five and six respectively.

2. Group consensus reaching process

Basically, we assume the following settings: there is a finite set of individuals (experts, agents) and the finite set of alternatives (options). Experts openly express their preferences by means of pairwise comparison as to the every pairs of available options [13]. What matters here is that the main goal of group decision making process is to find a solution that all decision makers are willing to support [2].

Decision making problem is an iterative and interactive process which includes several different levels, i.e. aggregating all individual preferences into one common decision, elaborating on the agreement in the spirit of consensus reaching process etc. Reaching consensus requires: time, active participating of all members, creative thinking and being open-minded, active listening, considering ideas, feelings and situations of every participant. The model of consensus reaching process is manageable only if individuals are able to negotiate and change their preferences, thus they are willing to support.

Consensus is reached when each expert in the group agrees to support the selected final decision, though it may not have been his or her first choice. It forces the group to consider all aspects of a problem and voice objections to possible alternatives [2]. Hence, the main part of this process is discussion, which gives the opportunity to exchange knowledge, clarify point of view, defend own preferences or to become convinced to different opinions. Any member can block consensus. That is why these kinds of decisions are more difficult and complex than others. Thus, to achieve the main goal, we assume that individuals are “committed to reaching consensus” – they are expected to iteratively update their testimonies, and as a result to finally attain a satisfactory agreement. We assume the topological approach of where agreement is measured on the basis of distance between individuals during every stage of the process. Initially, experts disagree in their preferences, so they are far away from consensus. The aim is to minimize this distance, and consequently lead the group closer to the acceptable agreement [14]. What matters here, is that these initial differences of opinions are a strength of the group and a key to gather additional information, clarify issues and force the group to search for better solutions with bigger benefits for everyone [2].

3. Group decision support systems for consensus reaching

Since the development of modern technology, computerized support in making decision has enormously progressed. Today's tools are flexible, efficient, easy in use and allow to create an interactive user-friendly interface to view data, configure models, etc. This class of computer-based information systems including knowledge based systems that support decision making activities is defined by one term – *decision support systems*. They combine the intellectual reserves of individuals with the proficiency of computer to enhance the quality of final decision [16, p. 13]. Similarly, *group decision support systems* mean interactive, intelligent, computer-based systems that facilitate solution to unstructured problems by a set of decision-makers working together as a group. Unstructured problems are "fuzzy, complex processes to which there are no cut-and-dried solution methods and where human intuition is often a basis for decision making [16, p. 11]." Software products provide collaborative support to groups, i.e. supply a mechanism for teams to share opinions, data, information, knowledge, and other resources. What matters here is that group decision support system is an adjunct to decision makers to facilitate their decision making process but not to replace their judgments. Moreover, it is a dynamic system which is adaptive over time, therefore, decision makers should be reactive and able to change their opinions quickly. Group decision support systems attempt to improve the effectiveness of decision making (accuracy, quality) rather than its efficiency (the cost of making decisions).

The key to success is to create more *human consistent* and *human centered* tools and techniques to grasp and deal with difficult (decision making type) problems. These systems should provide computational tools, cognitive aspects and social dimension. In the GDSS consideration it means that a computer asks a group to solve a problem, then collects, interprets and integrates the solutions obtained by humans.

The main role of this computer-based system plays *moderator* or *facilitator* who takes care of running the whole discussion process. Moderator constantly measures distances between individuals and checks whether consensus is reached or not. Moreover, his most important task is to support the discussion, i.e. he stimulates the exchange of information, suggests arguments, convinces decision makers to change their preferences, focuses the discussion on the issues which may resolve the conflict of opinions in the group. This is repeated until the group gets sufficiently close to consensus, i.e. until the individual fuzzy preference relations become similar enough, or until we reach some time limit [11]. Doubtless, the moderator affects the general sense of satisfaction within the group and has a direct influence on the quality of final decision. What matters here is that he only tries to persuade proper experts to change their opinions and suggests some rational arguments – he does not force, argue or push individuals to change their testimonies.

Our task is to develop and enhance the discussion part and provide the moderator with a specific knowledge about group members. Briefly speaking, we want to facilitate the work of moderator, provide him with some useful guidelines and additional indicators and, as a result, make the consensus reaching process easier, faster, more effective.

4. Notion of fairness in the consensus reaching support systems – a socio-psychological explanations with observations in the analyzed groups of students

One of the definition of *fairness* says that "fairness means the satisfaction of justified expectations of agents that participate in the system, according to rules that apply in a specific context based on reason and precedent [19]". Fairness is an intricate idea that depends on many factors, e.g.. cultural values or the context of the problem. It combines many different research areas such as mathematics, philosophy, economics and other social sciences, especially social psychology. The last research area is crucial because it gives an answer to a question: "how people understand fair behavior [18, p. 15]?" Generally, fairness is understood as equality and becomes an essential element of the new agent-based computational models which aim at explaining actual behavior. Thus, another question arises: whether one can talk about a simple unified principle which could possibly solve the quandary of decision making process? The range of factors influencing the equilibrium in a group usually turns out to be versatile and depending on group dynamics, psychological and situational, as well as personal characteristics of group members. The overall idea is that subjects may be simply prejudiced in their understanding of fairness. Opponents in the decision making process may be attuned to a different perception of fairness, which impedes decision making process, or causes that consensus becomes impossible. Group decision making is usually dependable upon the actual state of emotions pervading the circle of people who set themselves a task of reaching consensus.

Our analysis of group behavior in two groups of students consisting of 14 and 12 members respectively (group A and B – students of the third year at the Department of Automatic Control and Information Technology, Faculty of Electrical and Computer Engineering, Cracow University of Technology) allowed to draw the following conclusions. It was noticeable that if at least some people in a group cared about equity, consensus was reached more quickly and the decision making process became smoother. The second of the analyzed groups (B) showed a slightly more consensus prone character. The distinctions between the standpoints of the individual agents and subgroups were not so sharp as in group A. The major factor in reaching consensus in a smoother and quicker way in group B than in group A was a lower state of psychological agitation characteristic of the group, and a calmer, less aggressive level of communication. In its major part, the group consisted of agents with much less individualized personalities, and a better developed sense of cooperation resulting in a subsequent faster implementation of the final decision.

The other crucial and decisive factor in reaching consensus in this group was a very efficient cooperation in small subgroups. The individuals in group B usually exerted a much more open attitude to the preferences of other individuals, and were less likely to fall prey to a self-serving bias. The overall mechanism of the functioning of this group and reaching consensus process was based on a very conspicuous supportive system of the presentation of ideas and the skillful putting forth of the carefully selected arguments. Generally speaking, most of the members of the group showed a relatively high level of emotional intelligence allowing to analyze almost emotionlessly the arguments stated by their opponents. What was also well observable was the fact that some members of group B exerted a strong positive influence on the entire group by exhibiting good diplomatic and negotiating skills. Two of the agents in this group were extremely flexible in swift changes of the communicative strategies they deployed to invigorate the other members of the group to share the most exposed and most successfully supported opinion.

The explanations as for the behavioral patterns in this group can be supported by the definition of the *cooperative game theory* which virtually is a game where players can enforce fair behavior. Cooperative game theory is connected with the distribution of benefits that a group of agents achieves from cooperation. The model assumes that the group of individuals wishes to solve a common problem and by cooperating they can solve the problem more efficiently [18, p. 56]. In fact, research in psychology has shown that in group situations, decisions of individuals are influenced by motives such as "group performance, sense of responsibility for others, or social concerns [18, p. 61]". It is also worth noticing that a greater level of fairness and more successful achievement of goals in decision making process were attributed to some characteristics of the observed groups such as: an eagerness to learn the actual differences between their own opinions and those of the opponents, and an ability to eliminate such opinions which were impeding agreement the most. The openness and the markedly interactive character of the group relations proved to be the core of success in meeting the point of balance.

By comparison, the first group of students (A) which blocked consensus reaching process was characterized very often by an overall unwillingness to come to a satisfactory solution for all of the members, and an insistence on particular interests of mini groups, or individuals within the group. The mini groups were very assertive in expressing their opinions and often ignorant of the opinions of opponents. Sticking to the proclaimed opinions and not being open to the proposals of others, especially those which were very different ended up in an impasse of discussion and an impossibility of reaching an agreement. Quite frequent was the situation in which some members of the group showed a total disregard for the opinions of their opponents, closing themselves in a very limited mind framework. The creativity in reaching satisfactory solutions was decreased, which was not an optimistic forecast. We may assume that the diminished level of the coherence and creative unity of the group will not lead in the long run to an expected consensus, as well as a very much needed implementation of the possible solutions.

It cannot go unnoticed that the entire context of coming to an agreement is fundamentally a situational, contextual, and psychological phenomenon. Such factors as cultural unification or cultural clashes, and much in the same manner, economic differences or similar economic status may be of significance in group consensus reaching process. These elements may have influence on whether the more egoistically oriented individuals or the more fair prone agents overbalance the process of reaching an agreement. The economic element of group consensus reaching process was thoroughly examined, for instance, by Ernest Fehr and Klaus M. Schmidt [5]. In our analysis we ignored these factors focusing on other prevalent ingredients of the situational context such as, for instance, persuasive and manipulative capabilities of the group members.

In our study we observed that agents pushed equity principles which were advantageous for them more than for other parties, in particular, those which were disadvantageous to parties with great persuasion power. The first of the two analyzed groups of students showed a complexity of reactions within the small groups or subgroups in decision making process, highly dependable on psychological factors. One of the first things which were noticed proved that group behavior to some extent was in line with individual behavior of particular agents. Basically, group behavior depended on the decision rules that agents selected and used to arrive at group decision in a biased way. Another thing which was very well seen was that subgroups expressing similar standpoints within the observed groups ignored the decision

rule of their so called "opponents" – which meant that they followed the self-serving, or self-interest preference principle. What is more, a given subgroup sharing a similar viewpoint typically disregarded other subgroups and thought of themselves as if they were single agents. Above all, one must stress that expectations as for the level of fairness were often not consistent with the outcome; we expected the level of fairness to be higher.

These observations bear out the importance of the self-interest preference in group behavioral patterns. The choices of the respective members of the group underpin our theoretical assumptions as for the self-centeredness of the individuals influencing the final group decision making process.

However, psychological studies have revealed that in real life, decision makers are not as selfish as what is shown in the solutions received using mechanisms of rational choice approaches, in the sense of maximization of some utility function [6]. Experiments showed that individuals tend to cooperate and give priority to fairness over greedy behavior. *Trust game* will transparently perform this activity. In the *trust game*, A has an initial amount of money he or she could either keep or transfer to B. If A transfers it, the sum is tripled. B could keep this amount, or transfer it (partially or totally) to A. Traditional game theory suggests that A should keep everything, or if A transfers any amount to B, then B should keep all. Experimental studies have revealed that agents tend to transfer about 50% of their money and this fairness and cooperation is related to all cultures, sexes, etc [1].

With reference to our assumption that fairness means the satisfaction of expectations of agents, group decision support system should provide the sense of satisfaction among the group members during the discussion and after the process completion. According to psychological research, satisfaction of decision makers has a direct influence on higher quality of final decision and several further activities, i.e. practical implementation of the final decision or survival of the group in the long time period [15].

5. Fair share of distributed resources

In our research we mainly reflected on one of fairness judgments identified by social psychology, namely *distributive fairness* [17]. It is usually related to the distribution of resources, goods or costs, thus to *fair resource allocation problems*. Resource allocation problems are concerned with the distribution of constrained resources within competing activities so as to achieve the best general implementation of the system with respect to fair management of all the participants. Briefly speaking, the aim is to take a *fair share* of the distributed goods, thus to find such distribution that is perceived as fair by all individuals.

The overall resource allocation problem might be stated in the following way. There is a set of $I = \{1, 2, ..., m\}$ of m activities. There is given a set P of location patterns (location decisions). For each activity i, $i \in I$, a nonlinear function $f_i(x)$ of the location pattern x is defined. This function measures the outcome $y_i = f_i(x)$ of the location pattern for each activity i.

To get the individuals closer to each other (to obtain an agreement between them), there have to occur some changes in their initial preferences. The overall amount of changes in the individuals' preferences constitutes the resource. Thus, for a given set of decision makers moderator wants to allocate fairly the resource with the objective of minimizing the outcome (i.e. the distances between individuals).

The main goal of our system is to take into account preferences of every individual and get the entire group closer to the consensus with fair treatment of all the participants. We neglect the situation when the moderator gets decision makers closer to the consensus by argumentation and persuasion only aiming at the most promising directions of further discussion (those who reach consensus quickly), while individuals who are isolated in their opinion are omitted. We found confirmation of this assumption in our observations. As we observed group behavior in two selected groups there were many situations when consensus reaching process was blocked and the domineering members of the group, as well as the dominant subgroups exerting a powerful influence on the entire group caused that some individuals found themselves in a situation of a total isolation. One of the analyzed groups (group A) was likely to undergo all the distortions resulting from the self-interest prejudice. The interpersonal relations in the observed group were not good enough to overcome all the difficulties caused by the differences in opinions. Some of the members of the group were left on the margin and could not feel the satisfaction ensuing from discussion. This, in turn, brought about a situation of little effectiveness. All in all, the image of the group was negative, there were many lost opportunities of achieving consensus. Briefly speaking, moderator can not ignore the individuals who are isolated in their opinions as to the rest of the group members, quite on the contrary, he has to convince them to change their previous preferences. This attitude undoubtedly confirms one of our assumptions, namely, an importance of active participation of every individual during the entire consensus reaching process.

As we assumed, our research should be done with respect to fair distribution. The theory of distributive fairness can be applied whenever it is possible to precisely define a fair distribution problem and to find a solution that is accepted by participants (or proposed by the moderator). If we consider the distances of the individuals' opinion to the final opinion, naturally, the final opinion should be fair in the sense that the distances of the individuals' preferences to the final decision should be fairly distributed.

6. Fair final decision

While considering the concept of fairness with reference to consensus reaching process, we decided to view the basic idea of this notion from two possible perspectives. The first one, presented in the previous section, concerns a fair distribution of resources, whereas the second is directly connected with the outcome of decision making process, namely a *fair final decision*.

We define a *fair final decision* as a possibility to reach a final consensus during a series of discussions. However, the majority here refers directly to the outcome and can be defined as the *soft consensus,* a conceptual human-consistent framework proposed by Kacprzyk and Fedrizzi [9, 10], and Zadrożny [4]. The developed idea is meant basically as an agreement of a considerable majority of individuals with regard to a considerable majority of alternatives. This operational definition of consensus can be, for instance, expressed by a linguistically quantified preposition: *most of the individuals agree in their preferences to almost all of the options*, and the consensus degree (in the range [0, 1]) is computed. It means that, except none or total agreement between agents as to the chosen solution, this approach allows some partial, acceptable consistency.

Notice that to define a fuzzy majority for measuring a degree of consensus the application of *fuzzy linguistic quantifiers* (most, almost all etc.) has been performed. The computations of this relative type of linguistic quantity can be handled via Zadeh's classic calculus [20]. Regardless of the way of implementation, the main condition of this novel approach is that it definitely overcomes the conventional concept in which consensus was understood as a "full and unanimous agreement", which means that the preferences of all the decision makers should be exactly the same. Obviously, this scenario is utopian and unrealistic in practice because individuals usually expose relevant differences in their standpoints, flexibility, tendency to change opinions, etc. All of these factors generally block the group from gaining a full and unanimous agreement.

7. Conclusions

In this article we proposed a new concept of supporting group consensus reaching process enhanced by the concept of "fairness". We showed that this notion should be definitely taken into account while creating human-consistent support systems because it is strongly connected with psychology, economics, game theory, etc. and, as a result, takes cognizance of socio-psychological aspects of group behavior. In fact, it helps us understand the typical human behavior within a group of individuals and to develop more intelligent, human-centric and human-consistent systems for supporting consensus reaching in the future development.

In our research we came to a conclusion that a degree of consensus obtained by including aspect of fairness would be higher than the previous approach based solely on soft consensus with the use of fuzzy logic proposed and successfully implemented by Kacprzyk an Zadrożny [12]. We enriched our concept by the novel fairness component. Hence, we take liberty of putting forth a hypothesis that the concept of novel approach affects directly the effectiveness of decision making process, satisfaction among group members and, as the result, the quality of the final decision, which becomes highly justified. The ultimate goal of our further research is the mathematical formalization of the fair group consensus reaching process (building a model with regard to real events and psychological factors) in order to verify – confirm or reject – our assumptions.

Thus, our comprehensive approach proposed for supporting consensus reaching process under fuzziness refers particularly to a degree of agreement in a group, individual preferences of group members and the moderation of the discussion in order to gain satisfactory solutions in the more effective and efficient way. What matters here, is that we respect the fair distribution in the sense of contribution of all decision members to choose the final solution. Hence, the situation when minority must obey majority and change their opinions accordingly is in the proposed system ignored.

Dominika Gołuńska's contribution is partially supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing. Project financed from The European Union within the Innovative Economy Operational Programme 2007–2013 and European Regional Development Fund.

References

[1] Camerer C.F., *Psychology and economics. Strategizing in the brain*, Science, 300, 2003, 1673-1675.

[2] Center for Excellence in Government, Facilitator's Toolbox, 1996, http://www.employeesu.com/docs/EmployeeResources/Creating%20Consensus.pdf, [date of access: 23.03.13].

[3] Falkiewicz D., Kacprzyk J., *Different Aspects of Supporting Group Consensus Reaching Process Under Fuzzines*, Technical Transactions, vol. 1-AC/2012, Cracow University of Technology Press, 17-27.

[4] Fedrizzi M., Kacprzyk J., Zadrożny S., *An interactive multi-user decision support system for consensus reaching process using fuzzy logic with linguistic quantifiers*, Decision Support Systems, vol. 4, no. 3, 1988, 313-327.

[5] Fehr E., Schmidt K., *A Theory of Fairness, Competition and Cooperation*, The Quarterly Journal of Economics, MIT Press, no. 3, vol. 114, 1999, 817-868.

[6] Hanson S.O., *Decision Theory: A Brief Introduction*, Philosophy, vol. 23, Royal Institute of Technology, Stockholm, 1994, 1-94.

[7] Jennings H.H., *Sociometric differentiation of the psychegroup and sociogroup*, Sociometry, vol. 10, 1947, 71-79.

[8] Kacprzyk J., *Neuroeconomics: yet another field where rough sets can be useful?*, Chac C.C et al. (Eds.): RSCTC 2008, LNCS, vol. 5306, 2008, 1-12.

[9] Kacprzyk J., Fedrizzi M., *A 'human-consistent' degree of consensus based on fuzzy logic with linguistic quantifiers*, Mathematical Social Sciences, vol. 18, 1989, 275-290.

[10] Kacprzyk J., Fedrizzi M., *A 'soft' measure of consensus in the setting of partial (fuzzy) preferences*, European Journal of Operational Research, vol. 34, 1988, 315-325.

[11] Kacprzyk J., Zadrożny S., *On a concept of a consensus reaching process support system based on the use of soft computing and Web techniques*, [in:] D. Ruan, J. Montero, J. Lu, L. Martínez, P. D'hondt, E.E. Kerre (Eds.): Computational Intelligence in Decision and Control. World Scientific, 2008, 859-864.

[12] Kacprzyk J., Zadrożny S., *Soft computing and Web intelligence for supporting consensus reaching*, Soft Computing, vol. 14, no. 8, 2010, 833-846.

[13] Kacprzyk J., Zadrożny S., *Supporting Consensus Reaching Processes under Fuzzy Preferences and a Fuzzy Majority via Linguistic Summaries*, Preferences and Decisions Studies in Fuzziness and Soft Computing, Springer, vol. 257, 2010, 261-279.

[14] Kacprzyk J., Zadrożny S., Raś Z., *How to Support Consensus Reaching Using Action Rules: a Novel Approach*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, World Scientific, vol. 18, no. 4, 2010, 451-470.

[15] Priem R.L., Harrison D.A., Muir N.K., *Structured Conflict and Consensus Outcomes in Group Decision Making*, Journal of Management, 1995, vol. 21, no. 4, 691-710.

[16] Turban E., Aronson J.E., Liang T.P., *Decision Support Systems and Intelligent* Systems, 6th Edition, Prentice Hall, 11-19, 2005, 94-101.

[17] Tyler T.R., Smith H.J., *The handbook of social psychology*, vol. 2, *Social justice and social movements*, McGraw-Hill, Boston, 1998, 595-629.

[18] Wierzbicki A., *Trust and Fairness in Open, Distributed Systems*, Springer, 2010, 11-19, 56-70.

[19] Young H.P., *Equity: In Theory and Practice*, Princeton University Press, 1994.

[20] Zadeh, L, *A computational approach to fuzzy quantifiers in natural languages*, Computers and Mathematics with Applications, no. 9, 1983, 149-184.

TECHNICAL TRANSACTIONS | CZASOPISMO TECHNICZNE
AUTOMATIC CONTROL | AUTOMATYKA
1-AC/2013

KRZYSZTOF SCHIFF*

ANT COLONY OPTIMIZATION ALGORITHM FOR THE SET COVERING PROBLEM

ALGORYTM MRÓWKOWY DLA ZAGADNIENIA POKRYCIA ZBIORU

Abstract

This article describes a new hybrid ant colony optimization algorithms for the set covering problem. The problem is modeled by means of a bipartite graph. New heuristic patterns, which are used in order to choose a vertex to a created covering set have been incorporated into modified hybrid algorithms. Results of tests on investigated algorithms are discussed.

Keywords: set covering problem, ant colony optimization, heuristic rule

Streszczenie

W artykule przedstawiono nowy hybrydowy algorytm mrówkowy dla problemu zagadnienia pokrycia zbioru o minimalnym koszcie. Problem jest zamodelowany za pomocą grafu dwudzielnego. W modyfikowanym algorytmie wprowadzono nową heurystykę wyboru wierzchołków do podzbioru wierzchołków pokrywających. Opracowany algorytm przetestowano i porównano, a wyniki tych badań omówiono.

Słowa kluczowe: zagadnienie pokrycia zbioru, algorytm mrówkowy, heurystyka

* Krzysztof Schiff, Ph.D., Department of Automatic Control and Information Technology, Faculty of Electrical and Computer Engineering, Cracow University of Technology.

1. Introduction

Many practical optimization problems such as for example facility location problem, airline crew scheduling, nurse scheduling, vehicle routing and resource allocation problem can be described and modeled as the Set Covering Problem (SCP) [1, 2, 10–12] and many other combinatorial problems can be modeled in such a way. There are many kinds of computer algorithms that have been designed so far to solve SCP such as exact algorithms [13, 14], heuristic algorithms [15–18], meta-heuristics algorithms [19–21], also algorithms, which are based on ant colony optimization strategy [3, 6, 22, 23] and hybrid of Ant Colony Optimization Algorithm (ACO) with so called Constraint Programming (CP) [26, 27]. Ant algorithms were designed to solve many combinatorial problems [5, 8, 24, 25]. This paper presents new algorithms, which are based on an ant colony optimization strategy, for the set covering problem with a minimum covering cost and a new heuristic information patterns, which have been used in these algorithms. Transition probability rule and pheromone update rule and also a mechanism checking constraints consistency are used all together in order to solve the SCP and to minimize the total covering cost. The remainder of this paper is structured as follows: in section 2 the SCP is introduced, in section 3 the structure of ACO algorithm is described, in section 4 pseudo-code of ant algorithms with new heuristic patterns and transition probability rules, which is used in new elaborated ant algorithms, are discussed and in section 5 results of the conducted computational experiments on a special kind of a graph with an almost equal density and in section 6 conclusions are presented.

2. Set covering problem

The set covering problem can be modeled as a bipartite graph network $G(V_1 + V_2, E, w)$ with weights w_{ij} assigned to edges e_{ij}, such that $e_{ij} = (v_i \in V_1, v_j \in V_2)$, $e_{ij} \in E$ as it is shown in Fig. 1 and in the same way as it is presented in [9]. The degree of vertex i is the sum of edges adjacent to this vertex i. All vertices v_{1i} can be grouped into subsets in such a way that all vertices from the set V_2 are covered by vertices from the some subset V_s of vertices V_1. A vertex v_{2i} is covered by a vertex v_{1j} if an edge e_{ij} exists between a vertex i and a vertex j in a bipartite graph, for example subset $V_s \subset V_1$ which consists of vertices v_{11}, v_{13}, v_{14} and v_{16} covers all vertices from the set V_2 and the another example of set V_s is subset which consists of vertices v_{14} and v_{17}. In general if the cardinality number of set V_s is lower than the total set covering cost is higher. In the set covering problem with minimum the cardinality number of set V_s and the cost the total cost of set covering, this means the sum of weights assigned to all graph edges, which are participating in the set covering, has to be minimized.

The objective function F is to find a cover set with a minimum cost and is described in (1) and (2):

$$\min F = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(x_{ij} w_{ij} \right), \; i \in V_1, \; j \in V_2 \tag{1}$$

and subject to constraint such that:

$$\sum_{i \in Vs}^{n} x_{ij} \geq 1, \; i \in V_1, \; j \in V_2 \tag{2}$$

where:

$x_{ij} = 1$, when a vertex j is covered by a vertex i,
$x_{ij} = 0$, when a vertex j is not covered by a vertex i,
w_{ij} – this is a weight associated to edge e_{ij} (a cost of covering a vertex j by a vertex i),
V_s – is a subset of vertices V_1 which cover all vertices V_2.

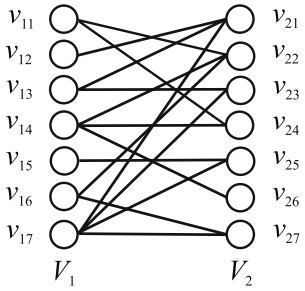


Fig. 1. The set covering problem modeled by a bipartite graph

Rys. 1. Problem pokrycia zbioru modelowany grafem dwudzielnym

3. Structure of ACO algorithm

In ant algorithms a colony of artificial ants is looking for a good quality solution of the investigated problem. The pseudo-code of ACO procedure is presented as algorithm 1. Each artificial ant constructs an entire solution of the problem in some number of steps, called intermediate solutions. Any of intermediate solutions are referred to as solution states. In each step m of the algorithm each ant k goes from a one state i to an another state j and thus constructs a new intermediate solution called later a partial solution of the problem since the entire solution is received in some number of steps and at each of these steps there is an intermediate solution called a partial solution or a solution under construction. At each step each ant k computes a set of feasible expansions to its current state and moves to one of these in probability. This set of feasible expansions is called a neighborhood of current state. In presented algorithms concerning processing of bipartite graph, this means working on a graph model of the set covering problem at each state each ant chooses a vertex from the set V_1 and adds it to a partial solution in order to construct finally at the end of algorithm action the entire solution to the SCP problem. At the end of algorithm action the set of vertices V_s constitute a solution to the SCP problem. Each ant k starts with an empty set V_s and successively adds to this set V_s a vertex chosen one after the other from the set V_1 with probability p^k_{ij} moving from a one to another state. At each state i there are some vertices in the set V_s and these vertices from the set V_s constitute a partial solution of a problem at step m, this means at state m. Each ant in order to construct a solution uses common information which is encoded in pheromone trails, this means the trail level of the move, indicating how proficient it has been in the past to make that particular move. Each ant also deposits a pheromone on a trail when a solution has been found and a quantity of the pheromone deposited depend on a quality of this solution. The move of each ant also depends on so called the attractiveness of the move, as computed by some heuristic indicating the a priori desirability of that move. In order to avoid a very fast

convergence to a locally optimal solution an evaporation mechanism is used, this means that over the time the pheromone trail evaporates, thus reducing its attractive strength.

Algorithm 1
ACO procedure

```
begin
    while (exist cycle) do
        while (exist any ant, which has not worked) do
            while (a solution has not been completed) do
                choose a next vertex to a constructed solution with a probability p^k_ij;
                update neighborhood of current state;
            end
            update a best solution if a better solution has been found;
        end
        update a global best solution if a better solution has been found;
        use an evaporation mechanism;
        update a pheromone trails τ(i) = τ(i) + Δτ;
    end
end.
```

Each ant k moves from one state i to another state j with a transition probability rule $p^k_{ij}(t)$, which is described by the formula:

$$p_{ij} = \left\{ \begin{array}{c} \dfrac{\left(\tau_{ij}^{\alpha}\mu_{ij}^{\beta}\right)}{\sum_{j\in N_i}\left(\tau_{ij}^{\alpha}\mu_{ij}^{\beta}\right)}, \quad \text{for} \quad j \in N_i \\ 0, \quad \text{for } j \text{ which not} \in N_1 \end{array} \right\} \tag{3}$$

using the pheromone trail τ_{ij} and the attractiveness μ_{ij} of the move. The pheromone trail τ_{ij} is the useful information, which is deposited by others ants, for each ant during its work on construction of solution, about the usage of vertex j in the past by others ants. The attractiveness μ_{ij} is a desire of choosing a vertex j from the neighborhood N_i of current state when there is a partial solution yet constructed in state i and the attractiveness μ_{ij} can by expressed by a some heuristic formula. The attractiveness μ_{ij} allows to better choose a some vertex from all vertices, from the neighborhood N_i of current state, to be added to a solution under construction taking an objective function into a consideration. The neighborhood N_i of state i is constituted by vertices which can be added to a constructed partial solution. At the start all vertices can be added to a partial solution of the problem, this means to a solution of a problem under construction and the number of these vertices is reduced not only because of their inclusion into the solution, which is under the construction, but also because some of these vertices cannot be yet added to a solution, which is under construction, since these vertices does not satisfied solution constraints and only these vertices can be added to constructed partial solution which still satisfied solution constraints. The partial solution of the problem is a part of solution and the partial solution is a subset of vertices, which constitute a solution of the problem. Parameters α and β which is used in the transition

probability rule $p^k_{ij}(t)$ expressed by (3), indicate about this, how important the pheromone trail τ_{ij} and the attractiveness μ_{ij} are during transition from one to another state. Values of these parameters α and β should be set by experiment and tuned to the set covering problem with minimum covering cost.

After a solution has been found each ant deposits a pheromone with a quantity $\Delta\tau$ on all vertices, which constitute the solution V_s, in accordance with the pattern:

$$\tau_{ij}(t) = \tau_{ij}(t) + \Delta\tau \tag{4}$$

Thus these vertices which were included into a solution have received an additional quantity of a pheromone and can be chosen to a solution that would be constructed next with a higher probability than others vertices from the set V_1.

An evaporation mechanism is incorporated into an ant algorithm in order to avoid a too fast convergence to a sub-optimal solution. An intensity of evaporation is controlled by a parameter ρ and a quantity of a pheromone on each vertex from the set V_1 is update at the end of each cycle in accordance with the pattern:

$$\tau_{ij}(t) = (1-\rho)\,\tau_{ij}(t), \quad \rho \in (0, 1] \tag{5}$$

Thus a diversity of a solution is granted. Values of a parameter ρ should be set by experiment.

A quantity of deposited pheromone $\Delta\tau$ depends on a quality of solution Q and if the better is a solution than the more pheromone is deposited and in general can be stated as formula:

$$\Delta\tau = f(Q) \tag{6}$$

and in particular can be expressed by some specific formula, which take into account the covering cost.

4. Hybrid ACO algorithm

Both ACO-SCP algorithms, which are discussed in this paper, are modified versions of the hybrid algorithm described in [6] and in this paper the general pseudo-code of these algorithms is presented as algorithm 2. In the algorithm presented in this paper a new dynamic heuristic rule is proposed. The dynamic heuristic information:

$$\mu_1(i) = \frac{\nu c}{\sum_{j=1} w_{ij}}, \ \textit{if } x_{ij} = 1 \text{ and } j \in S_2 \text{ and } j \text{ is not covered yet} \tag{7}$$

is defined in the same way as in the paper [6] and the dynamic heuristic information $\mu_2(i)$ and $\mu_3(i)$ can be defined adequately:

$$\mu_2(i) = \sum_{j=1}^{n} \left(w_{\max} - w_{ij} \right) if \left[\left(w_{ij} - w_{kj} \right), k \in V_s \right] \tag{8a}$$

$$\mu_3(i) = \sum_{j=1}^{n} \left(w_{kj} - w_{ij} \right) if \left[\left(w_{ij} < w_{kj} \right), k \in V_s \right] \tag{8b}$$

where:

vc – this is a number of additionally covered vertices from S_2 if vertex i would be included into a solution V_s,
$w_{\max}$ – this is the maximal weight from weights associated to an edges $e_{ij,}$
w_{ij} – this is a weight associated to an edge e_{ij},
V_s – this is a constructed yet subset of V_1 vertices,
k – this is a vertex already included into set V_s,
$\mu_1(i)$ – desirability of vertex i when not covered vertices j from the set V_2 are taken into consideration,
$\mu_2(i)$ and $\mu_3(i)$ – desirability of vertex i when edge covered vertices j from the set V_2 are taken into consideration,

$x_{ij} = 1$ when an edge e_{ij} exists between vertex i and vertex j and $x_{ij} = 0$ otherwise.

In both algorithms a following vertex that should be added to a partial solution is chosen with a probability that depends on a pheromone trail, heuristic information and transition rule. Main differences between elaborated algorithms and algorithm, which is presented in the paper [6] concern a transition probability rule (9), (10) and heuristic information (8a), (8b). Since the quality of a solution depends on a total weight of covering, this means depends on a sum of weights assigned to all edges between all vertices of the set V_2 and the set V_s, the attractiveness μ_{ij} of choosing vertex j expressed as a function of a weight is very important. At any state only these vertices from the set V_1 which can improve quality of solution should be considered when any ant choose following vertex that should be added to a partial solution and to a set of vertices V_s. Such vertices from the set V_1 which can improve quality of a solution will be called available vertices and will be constitute a set V_A and these vertices will be also constitute the neighborhood N_i of a current state. These vertices from the set V_1, which cannot improve a quality of solution are excluded as a result of consistency checking from available vertices V_A and such vertices constitute a set V_{ex}. It is obvious that if any vertex is included into a partial solution V_s it cannot belongs to a set of available vertices V_A, so taking the above into consideration the number of available vertices V_A can be computed in accordance with expression $V_A = V_1 - V_s - V_{ex}$. The attractiveness $\mu_1(i)$ and $\mu_2(i)$ and $\mu_3(i)$ of choosing vertex i from available vertices V_A depend on weights of its edges and not only it concerns not covered yet vertices from the set V_2 expressed by the attractiveness $\mu_1(i)$, but also these vertices from the set V_2, which have been covered up till now, this means up to this moment of choosing from the neighborhood N_i of a current state the next following vertex i expressed by the attractiveness $\mu_2(i)$ and $\mu_3(i)$:

a) for HACO1-SCP:

$$p(i) = \frac{\tau(i)\mu_1(i)\mu_2(i)}{\sum_{i \in V_A} \left(\tau(i)\mu_1(i)\mu_2(i) \right)}, \; V_A \in V_1 \tag{9}$$

b) for HACO2-SCP:

$$p(i) = \frac{\tau(i)\mu_1(i)\mu_3(i)}{\sum_{i \in V_A} \left(\tau(i)\mu_1(i)\mu_3(i) \right)}, \; V_A \in V_1 \tag{10}$$

where:

V_A – this is a set of available vertices, $V_A = V_1 - V_s - V_{ex}$,
V_{ex} – these are vertices, which are excluded as a result of consistency checking,
$\tau(i)$ – this is a pheromone trail on a vertex i,
$(\mu_1(i)\ \mu_2(i))$ – this is a heuristic information associated with a vertex i in part a),
$(\mu_1(i)\ \mu_3(i))$ – this is a heuristic information associated with a vertex i in part b).

A quantity of pheromone $\Delta\tau$ is deposited by ants during one cycle of algorithm action on all vertices of the set V_s, which were included into the best constructed solution, in accordance with the formula:

$$\Delta\tau = \frac{1}{1 - \frac{c_{\text{best}} - c}{c_{\text{best}}}} \tag{11}$$

where:

c_{best} – this is the best cost of covering,
c – this is an actual cost of covering.

Algorithm 2
Hybrid ACO procedure for SCP

```
begin
    while (exist cycle not done) do
        for (k:=1 to n Ants) do
            while (a solution (a subset V_s) is not completed ) do
                Update Available Vertices;
                Choose next vertex i with probability p(i) and consistency checking;
                Add to a Partial Solution;
                Update Partial Solution;
            end
            Save a Better Solution;
        end
        Update Optimum;
        Use an evaporation mechanism;
        Update Pheromone;
    end
Return Best Solution Founded;
end.
```

Constraint Programming based on Edge Consistency with pre and post-processing.

An edge adjacent to a vertex i from the set V_1 and exactly from the set V_A can be added to a partial solution only if a cost (weight) of this edge is lower in comparison to a cost of any edges which are already included to a partial solution V_S and when both these edges, which weights are compared, cover the same vertex j from the set V_2. Adjacent edges of this vertex i, whose costs are higher, cannot be added to any partial solution, thus to a solution of the problem.

A vertex i from set V_1, which has been already included into a partial solution V_s, will be excluded from this partial solution V_S only when costs (weights) all of its edges are higher in comparison to costs of other edges adjacent to vertices which are too already included into a partial solution, this means that a vertex i will be excluded if this vertex i has no edge, this means that its edges were with higher weights and were excluded before from constructed partial solution up till now and thus a vertex i has now no edge at all and a vertex i can be now excluded from a partial solution V_s.

A vertex i from set V_1, which can be chosen to a partial solution V_s, is excluded from these available vertices to be chosen to a partial solution V_A if costs of all its all edges are higher in comparison to costs of edges from a constructed yet partial solution, this means from the set V_s covering these some vertices j from the set V_2 and any of its edges cover any vertex j from the set V_2, which is not yet included into a partial solution V_s.

This new modified heuristic pattern lets to receive a better solution than a solution which is received by the hybrid ACO algorithm, which is presented in [6] for a bipartite graph with almost equal degree of all vertices. The constraint programming technique used in this paper is based on the edge consistency with pre and post processing [4, 6, 7]. Thus a number of available vertices which can be potentially included into a partial solution is minimized and any no longer needed vertices are eliminated from a partial solution and also only these edges with lower weights are added to a solution under construction and these with higher weights are eliminated from a partial yet constructed solution.

5. Experiments

There are three algorithms which were studied during experiments. The first is the HACO algorithm, which was described in [6], the second is the HACO1-SCP algorithm with desirability $\mu_2(i)$ and the third is the HACO2-SCP with desirability $\mu_3(i)$, which are described in this paper. Two parameters were under observation during conducted experiments: an average minimum cost of set covering and an average cardinality number of the set V_s, which were received as a result of 10 measures. All algorithms were studied for a bipartite graph with 100 x 100 vertices and for a different graph densities q, which were generated in random. Later all algorithms were studied for a bipartite graph with random generated edges for each its vertex, this means with random generated vertex degree and for measure cases with different number of vertices, this means for 50 x 50, 100 x 100, 150 x 150, 200 x 200 and 250 x 250 vertices in a bipartite graph. These bipartite graphs belong to the particular kind of a bipartite graph since each edge in each of these graphs exists with a probability q and thus each vertex of graph has almost equal degree, this means has almost equal number of adjacent edges. An average minimum set covering costs for 10 measures were presented in Table 1 and Fig. 2 and an average minimum cardinality numbers of the set V_s for 10 measures were presented in Table 2 and Fig. 3. These all three algorithms ran with the following common parameters setting for each measure cases: the evaporation rate was set to 0.995, the number of ants was set to 200 and the number of cycles was set to 300.

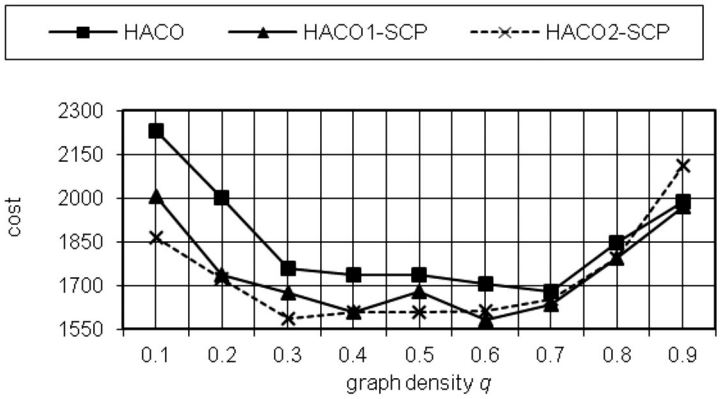


Fig. 2. An average cost in dependency on a graph density q

Rys. 2. Średni koszt pokrycia w zależności od gęstości grafu

Table 1

An average cost in dependency on a graph density q

q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
HACO	2233.0	2000.4	1757.8	1738.5	1736.6	1706.8	1678.9	1845.6	1988.3
HACO1-SCP	2006.9	1736.5	1676.0	1608.3	1677.4	1583.5	1633.1	1794.4	1969.8
HACO2-SCP	1865.5	1721.7	1587.7	1606.3	1606.4	1610.8	1651.3	1792.3	2111.1

Table 2

An average cardinality number of V_s in dependency on a graph density q

q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
HACO	25.9	15.6	11.6	8.7	6.9	5.7	5,0	4,0	3.2
HACO1-SCP	25.3	15,0	10.3	7.9	6.1	5.0	4,0	3.1	2.2
HACO2-SCP	26.4	15,0	10.8	7.9	6.3	4.9	4,0	3,0	2,0

There is an improvement in quality of the solution when the HACO1-SCP or HACO2--SCP algorithm is used instead of the HACO algorithm since there are lower cardinality numbers of the set V_s and there are lower covering costs for all investigated graph densities q = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. The HACO2-SCP algorithm is better than the HACO1-SCP algorithm when average costs are taken into consideration for rare graphs $q \leq 0.5$ and there is not a difference between both algorithms for dense graphs $0.5 < q$. As concern average cardinality numbers HACO1-SCP and HACO2-SCP algorithm do not differ from one another.

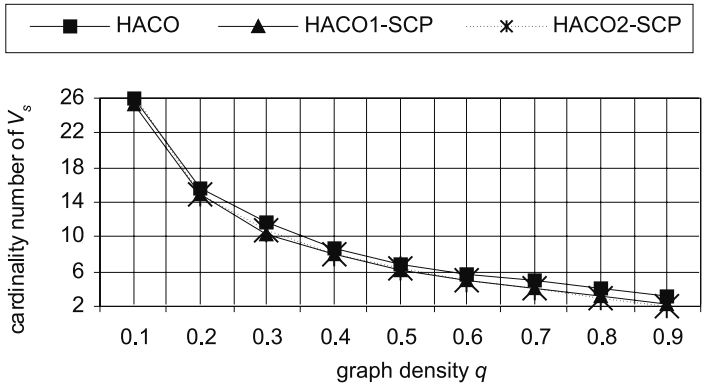


Fig. 3. An average cardinality number of V_s in dependency on a graph density q

Rys. 3. Średnia liczba kardynalna zbioru V_s w zależności od gęstości grafu q

The HACO1-SCP and the HACO2-SCP algorithm are also better than the HACO-SCP algorithm when average costs of set covering and average cardinality numbers of the set V_s are taken into consideration in function of a number of bipartite graph vertices n with different vertices degree. These two above parameters has been observed during conducted tests when a graph density was differentiated for different number of graph vertices and received values of two above parameters have been shown in the Table 3 and in the Table 4 or in the Fig. 4 and in the Fig. 5. In order to get a bipartite graph with a different density for each graph vertex each edge was generated with any probability so degree of each graph vertex has different values, this means each graph vertex has a different number of adjacent edges and thus a bipartite graph has vertices with different degree.

All experiments have shown that both elaborated algorithms give a better quality of solution than the HACO algorithm which has been presented in the paper [9].

Table 3

An average cost in dependency on a number of vertices

n	50	100	150	200	250
HACO	856.1	1696.1	2538.6	3314.4	4026.3
HACO1-SCP	839.8	1636.5	2433.8	3215.3	3933.9
HACO2-SCP	829.9	1571.7	2374.4	3106.9	3927.9

All algorithms, the HACO algorithm and these both elaborated algorithms HACO1--SCP and HACO2-SCP, which are presented in this paper, are implemented in Microsoft Visual C++ under Microsoft Windows XP on Intel Celeron CPU 1.7GHz, 256 Mb RAM and the running time of these algorithms are proportional to the time complexity expressed by a multiplication of a quadratic number of vertices n^2 existing in a bipartite graph, a number of cycles, a number of ants and a cardinal number of a set V_s.

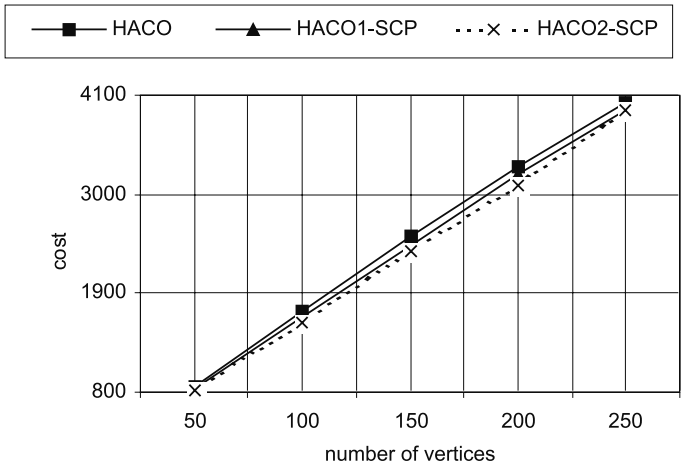


Fig. 4. An average cost in dependency on a number of vertices

Rys. 4. Średni koszt w zależności od liczby wierzchołków

Table 4

An average cardinality number of V_s in dependency on a number of vertices

n	50	100	150	200	250
HACO	6.1	7.0	7.8	8.4	9.0
HACO1-SCP	5.0	6.1	7.0	7.6	8.0
HACO2-SCP	5.0	6.1	7.0	7.8	8.0

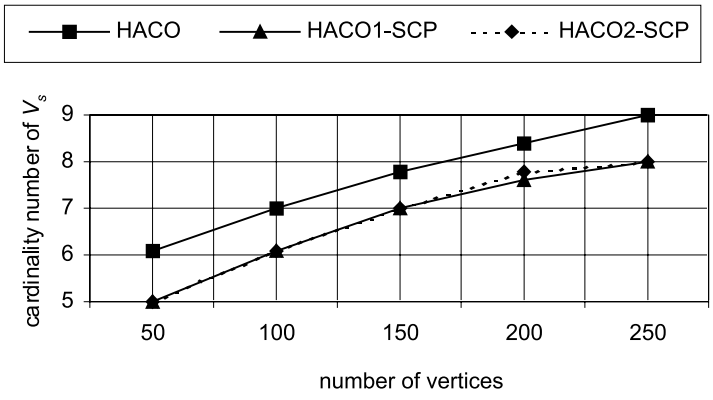


Fig. 5. An average number of cardinality number V_s in dependency on a number of vertices

Rys. 5. Średnia liczba kardynalna zbioru V_s w zależności od liczby wierzchołków

6. Conclusions

In this article the minimum cost set covering problem was solved by using the ACO algorithm with Constraint Programming and with new heuristic patterns. These new proposed heuristics, which have been used in the HACO1-SCP and the HACO2-SCP algorithm, lets to match in a better way an available vertex i from the neighborhood N_i of state to an already constructed partial solution V_s in case of graphs with an almost equal degree of vertices and with edges, which have been generated in random with a determined probability q and in case of graphs with different number of vertices and different degree of vertices and with edges, which are generated in random with any probability q for each graph vertex. Both the HACO1-SCP and the HACO2-SCP algorithm look for a new vertex form the set V_A to be added to a partial solution V_s with the highest number of additional edges and with the lowest corresponding overall cost of partial solution. The HACO algorithm is taking into account only these available vertices from the set V_A which are outside of a partial solution V_S and which can be added to a partial solution with a minimum additional average cost, this means with an average minimum sum of edge weights and thus omits these available vertices from the set V_A which can be added to a partial solution V_s with a higher number of additional edges and a little higher cost than an average minimum sum of edge weights and which can minimize the overall cost of a constructed partial solution because of lower weights of its edges covering already covered yet vertices from the set V_2.

References

[1] Valenzuela C., Crawford B., Soto R., Monfroy E., Paredes F., *A 2-level Metaheuristic for the Set Covering Problem*, International Journal of Computers, Communications & Control, vol. 7/2, 2012, 377-387.

[2] Ren Z.G., Feng Z.R., Ke L.J., Zhang Z.J., *New ideas for applying ant colony optimization to the set covering problem*, Journal Computers and Industrial Engineering, vol. 58/4, 2010, 774-784.

[3] Lessing L., Dumitrescu I., Stutze T., *A comparision between ACO algorithms for the set covering problem*, LNCS, vol. 3172, 2004, 1-12.

[4] Ren Z.G., Ke L.J., Chang H., *A fast and efficient Ant colony Optimization Approach for the Set covering problem*, Proc. of IEEE World Congress on Computational Intelligence, Hong Kong 2008, 1839-1844.

[5] Hadji R., Rahoual M., Talbi E., Bachelet V., *Ant colonies for the set covering problem*, [in:] Abstract Proc. of From Ant Colonies to Artificial Ants: Second International Workshop on Ant Colony Optimization, Brussels 2000, 63-66.

[6] Crawford B., Soto R., Monfroy E., Paredes F., Palma W., *A hybrid Ant algorithm for the set covering problem*, International Journal of the Physical Sciences, vol. 6/19, 2011, 4667-4673.

[7] Dechter R, Frost D., *Backjump-based backtracking for constraint satisfaction problems*, Artificial Intelligence, vol. 136/2, 2002, 147-188.

[8] Dorigo M., Di Caro G., *The Ant Colony Optimization meta-heuristics. New Ideas in Optimization*, McGraw Hill, New York 1999.

[9] Mihelic J., Robic B., *Facility Location and Covering Problems*, Proc. of the 7th International Multiconference Information Society, Volume D – Theoretical Computer Science, Ljubljana, Slovenia 2004.
[10] Housos E., Elmoth T., *Automatic optimization of subproblems in scheduling airlines crews*, Interfaces, vol. 27/5, 1997, 68-77.
[11] Vasko F.J., Wilson G.R., *Using a facility location algorithm to solve large set covering problems*, Operations Research Letters, vol. 3/2, 1984, 85-90.
[12] Vasko F.J., Wolf F.E., *Optimal selection of ingot sizes via set covering*, Operations Research, vol. 35/3, 1987, 346-353.
[13] Balas E., Carrera M.C., *A dynamic subgradient-based branch-and-bound procedure for set covering*, Operations Research, vol. 44/6, 1996, 875-890.
[14] Fisher M.L., Kedia P., *Optimal solution of set covering/partitioning problems using dual heuristics*, Management Science, vol. 36/6, 1990, 674-688.
[15] Chvatal V., *A greedy heuristic for the set-covering problem*, Mathematics of Operations Research, vol. 4/3, 1979, 233-235.
[16] Lan G., DePuy G.W., *On the effectiveness of incorporating randomness and memory into a multi-start meta-heuristic with application to the set covering problem*, Computer & Industrial Engineering, vol. 51/3, 2006, 362-374.
[17] Ceria S., Nobili P., Sassano A., *A Lagrangian-based heuristic for large-scale set covering problems*, Mathematical Programming, vol. 81, 1998, 215-228.
[18] Caprara A., Fischetti M., Toth P., *A heuristic method for the set covering problem*, Operations Research, vol. 47/5, 1999, 730-743.
[19] Beasley J.E., Chu P.C., *A genetic algorithm for the set covering problem*, European Journal of Operational Research, vol. 94/2, 1996, 392-404.
[20] Brusco M.J., Jacobs L.W., Thompson G.M., *A morphing procedure to supplement a simulated annealing heuristic for cost- and coverage-correlated set-covering problems*, Annals of Operations Research, vol. 86, 1999, 611-627.
[21] Caserta M., *Tabu search-based meta-heuristic algorithm for large-scale set covering problems*, [in:] Doerner K.F., Gendreau M., Greistorfer P., Gutjahr W., Hartl R., Reimann M. (Eds.), *Metaheuristics: Progress in complex systems optimization*, Springer, New York 2007, 43-63.
[22] Lessing L., Dumitrescu I., Stützle T., *A comparison between ACO algorithms for the set covering problem*, [in:] Dorigo M., Mauro Birattari M., Blum C., Gambardella L.M., Mondada F., Stützle T. (Eds.), *Ant Colony Optimization and Swarm Intelligence*, LNCS, vol. 3172, Springer-Verlag, Berlin, Heidelberg, 2004, 1-12.
[23] Crawford B., Castro C., *Integrating look ahead and post processing procedures with ACO for solving set partitioning and covering problems*, [in:] Rutkowski L., Tadeusiewicz R., Zadeh L.A., Żurada J.M. (Eds.), *Proc. of the 8th international conference on Artificial Intelligence and Soft Computing*, Springer-Verlag, Berlin, Heidelberg 2006, 1082-1090.
[24] Finger M., Stützle T., Ramalhinho H., *Exploiting fitness distance correlation of set covering problems*, [in:] Cagnoni S., Gottlieb J., Hart E., Middendorf M.R., Raidl G.R. (Eds.), *Proc. of the Applications of Evolutionary Computing on EvoWorkshops*, LNCS, vol. 2279, Springer, Berlin, Heidelberg 2002, 61-71.
[25] Dorigo M., Gambardella L.M., *Ant colony system: A cooperative learning approach to the traveling salesman problem*, IEEE Transactions on Evolutionary Computation, vol. 1/1, 1997, 53-66.

[26] Gagne C., Gravel M., Price W., *A look-ahead addition to the ant colony optimization metaheuristic and its application to an industrial scheduling problem*, Proc. of the fourth Metaheuristics International Conference, vol. 1, Proto 2001, 79-84.

[27] Michel R., Middendorf M., *An island model based ant system with lookahead for the shortest super sequence problem*, [in:] Eiben A.E., Bäck T., Schoenauer M., Schwefel H.P. (Eds.), *Parallel Problem Solving from Nature*, LNCS, Springer Verlag, vol. 1498, 1998, 692-701.

TECHNICAL TRANSACTIONS | **CZASOPISMO TECHNICZNE**
AUTOMATIC CONTROL | AUTOMATYKA
1-AC/2013

GRZEGORZ NOWAKOWSKI*

OPEN SOURCE RELATIONAL DATABASES AND THEIR CAPABILITIES IN CONSTRUCTING A WEB-BASED SYSTEM DESIGNED TO SUPPORT THE FUNCTIONING OF A HEALTH CLINIC

MOŻLIWOŚCI WYKORZYSTANIA RELACYJNYCH BAZ DANYCH *OPEN SOURCE* DO BUDOWY INTERNETOWEGO SYSTEMU WSPIERAJĄCEGO PRACĘ PRZYCHODNI ZDROWIA

Abstract

In this paper the capabilities of using open source relational databases to construct a web-based system designed to support the functioning of a health clinic have been presented as an alternative to commercial solutions. The author introduced a prototype of the system, which is based on selected database. Obtained results confirm the assumption that the medical system does not have to strain health care budget, while providing an acceptable standard of services.

Keywords: open source, relational database, internet system, health clinic, prototype

Streszczenie

W niniejszym artykule przedstawiono możliwości wykorzystania relacyjnych baz danych *open source* do budowy internetowego systemu wspierającego pracę przychodni zdrowia jako alternatywy dla komercyjnych rozwiązań. Autor zaprezentował stworzony na podstawie wybranej bazy prototyp systemu. Opisane rezultaty potwierdzają założenie, że system medyczny nie musi nadwyrężać budżetu służby zdrowia, zapewniając jednocześnie akceptowalny poziom świadczonych usług.

Słowa kluczowe: open source, relacyjna baza danych, system internetowy, przychodnia zdrowia, prototyp

* Grzegorz Nowakowski, M.Sc., Department of Automatic Control and Information Technology, Faculty of Electrical and Computer Engineering, Cracow University of Technology.

1. Introduction

In recent years the internet technologies have gained immense importance in the management of health care services, not only in case of an individual practitioner, but also at national and regional levels. They offer the possibility to reduce administrative costs and the provision of distance health care in order to avoid an unnecessary duplication of the same medical examinations.

The role of the internet has been constantly growing as it has been widely used by citizens to obtain medical information at present. This situation has drawn attention to the urgent need to elaborate on further information and services in the field of health care and make them available to citizens as well as to ensure the proper quality and safety of internet websites.

Implementation of the system supporting the functioning of a health clinic is usually a very expensive enterprise. Hence, designing a system meant meeting the above-made assumptions, not only the functionality aspects but also the economic option should be considered.

The purpose of this paper is to show the possibility of using open source relational databases to build a web-based system designed to support a health clinic's functioning as an alternative to commercial solutions. For the sake of implementing the assumption, some fully-functional databases systems (including mechanisms such as procedures, functions, triggers and cursors) were tested and compared. Then, the author presented a prototype of the system, which, based on the selected database system, will support health clinics. Obtained results confirm the assumption that the medical system does not have to strain the health care budget, while providing an acceptable level of a given service.

2. Open source relational databases review

Open source is a way of creating and distributing software based on a free-sharing system with the source code [1]. The specificity of an open source product distribution allows for an analysis, an independent compilation (creation of the executable option) and modifications to the code by independent developers.

The idea of open source [2] software is gaining more and more importance and attention, and all indications are that we will see its further development in the future. So far, this type of software has been developing in the environment using the internet as a medium. Open source been introduced to the commercial market and changed the prevailing rules of thereof.

Open source software offer covers a wide range of products – from device drivers, usage packages, to server applications and development tools [3].

A relational database is a computer database in which all data is stored in relations which (to the user) are tables with rows and columns [4]. Each table consists of records (called tuples) and each record is identified by a field (attribute) containing a unique value. Every table shares at least one field with another table in 'one to one', 'one to many', or 'many to many' relationships. These relationships allow an database user to access the data in almost an unlimited number of ways, and to combine the tables as building blocks to create complex

and very large databases. It has an internal programming languages, typically using SQL to manipulate the data, by which it creates advanced data services [5].

Relational databases are the basis for most modern systems [6]. Although various database management systems differ from one another in many aspects, they are based on a common theoretical basis.

A brief description of five fully functional (including mechanisms such as procedures, functions, triggers, and cursors) relational open source databases has been presented below.

MySQL is one of the most popular open source relational database management system (RDBMS) [8]. Right now it is being developed by the Oracle company. MySQL is used by some large-scale projects such as Wikipedia, Twitter and Facebook. Oracle Company has optimized MySQL's functioning to work on Windows, so the database system has become fully cross-platform.

MariaDB is a binary equivalent of the MySQL relational database management system [8]. Each month MariaDB code is combined with MySQL code in order to provide support for all the functionalities and add modifications introduced by Oracle. MariaDB is available under GNU license.

PostgreSQL has its roots in the Ingres project, conducted at the University of Berkeley in the 80's of last century. The first version was presented in 1995. It is an open source project based on PostgreSQL license. PostgreSQL is an object-relational database management system (ORDBMS) [10].

Apache Derby is a relational database management system (RDBMS) available as an Apache project [11]. It is written in Java and designed to be used in software written in this language. It has very low system requirements – less than 3 MB of memory, including a built-in JDBC driver. It is available under the Apache license, version 2.0. The database system can operate in two modes: as a full-featured embedded database system or as a traditional database system working in client-server mode using Derby Network Server.

Firebird is an open source SQL relational database management system [12]. The current version of Firebird is based on InterBase, created and developed by Borland in 2000. Developed modules are available on the license of Initial Developer's Public License. Firebird runs under the control of Linux, Unix and Windows [7].

3. Comparison of open source relational databases

Discussed databases systems were tested on large data tables. Testing data was a log file (from one of the web server) consisting of a million entries. The same data set was used for all the tested products. Standardized tests such as: mass import of data and basic data operations like SELECT, UPDATE, and DELETE were started. 64-bit Dell servers, running on Windows Server 2008 R2, were used in a test environment. All the tested databases are cross-platform [7].

Comparison of relational open source databases [7–13]

Database	MySQL	MariaDB	PostgreSQL	Apache Derby	Firebird
Version	5.5.8	10.0	9.2	10.9.1	2.5.1
Management tool	phpMyAdmin, HeidiSQL	HeidiSQL	pgAdmin	interactive scripting tool called 'ij'	FlameRobin
Dataset import time (one million records)	time similar to MariaDB	one minute	thirty seconds	few seconds	slower than MySQL, MariaDB and PostgreSQL
Returning time of the first hundred thousand records	time similar to MariaDB	two seconds	thirty seconds	five seconds	slower than MySQL, MariaDB and PostgreSQL
Performing time of the basic operations (SELECT, UPDATE, DELETE, orders) on smaller datasets	slower than MariaDB	fractions of a second	half a second	slower than MySQL and MariaDB	slower than MySQL, MariaDB and PostgreSQL
Security	access control lists for all objects and operations SSL communication between the client and server cryptographic functions, etc. (everything contained in the manual)	access control lists for all objects and operations	authentication through a number of mechanisms from the Kerberosto LDAP SSL communication between the client and server database can be encrypted with using the pgCrypto extension	authentication and authorizing of users and restrictions on the level of objects possibility to encrypt the database	access control at the level of objects and the ability to remove the source code triggers, procedures and views

Database	MySQL	MariaDB	PostgreSQL	Apache Derby	Firebird
Backup and recovery	offers a variety of backup and recovery strategies from which we can choose the methods that best suit the requirements for our installation (everything contained in the manual)	backup/ restore tools (mysqldump, xtrabackup, filesystem snapshots, etc)	backup/ restore tools (SQL dump, cluster dump, offline copy backup, continuous archiving, pg_basebackup, point-in-time recovery, pg_upgrade)	provides a way to back up a database while it is either offline or online restore a full backup from a specified location (everything contained in the manual)	backup/ restore tools (gbak commands, nbackup)
Technical support	available from Oracle (both free of charge and paid) and many other online sources (for diverse levels)	website design as well as internet forums and blogs	extensive documentation and FAQ, active community of users	apache website contains complete documentation active developer community	website contains complete documentation
Additional advantages	9 out of 10 websites use MySQL databases (according to Oracle)	mechanism of sub-query (sub-queries)	available in most Linux distributions	low system requirements	possibility of encoding the procedures in Java and C + + (version 3.0)

As shown in the above comparison MariaDB server is a bit more efficient than the others. Also, it is an open product designed and developed by the same team that created MySOL prior to its acquisition by Oracle. MariaDB can be used by any server that has been used by MySQL up until now.

Table 2

Summary of tested database servers [7]

Database server	Arguments for	Arguments against
MySQL	cross-platform, high performance, a large number of installations, the ability to migrate to commercial versions	some functionalities reserved for commercial versions, the lack of built-in management tools
MariaDB	performance, cross-platform, commercial independence, completely compatible replacement for MySQL	the lack of built-in management tools
PostgreSQL	cross-platform, high performance, management tools in the basic installation	sometimes a little intuitive syntax, poor performance of logging
Apache Derby	cross-platform, low system requirements, can be used as an embedded solution or in a client-server environment	incomplete management tools, trouble with the syntax
Firebird	cross-platform, easy installation	performance, scalability

4. A prototype of an internet system to support the functioning of a health clinic

Working with an internet system which supports the functioning of a healthcare center logistically corresponds with the order of the stages which the patient goes through in order to benefit from a healthcare center [14]. The system is expected to enhance the doctor's work.

A prototype of the system was developed on the basis of a few monthly interviews conducted with the staff that is employed in Scanmed Multimedis health center in Krakow. This fact enabled the author to design particular modules of the system, and a special emphasis was put on their compatibility with a real activity.

4.1. The system concept

Among the users of the system three main groups should be distinguished: patients, doctors and administrators. Access to particular modules of the system is granted for each user on different authorization levels, according to their needs and competencies.

All data in the system are stored on a server. There is information regarding patients, personnel, visits, tests, et cetera there. The common means of communication is the internet.

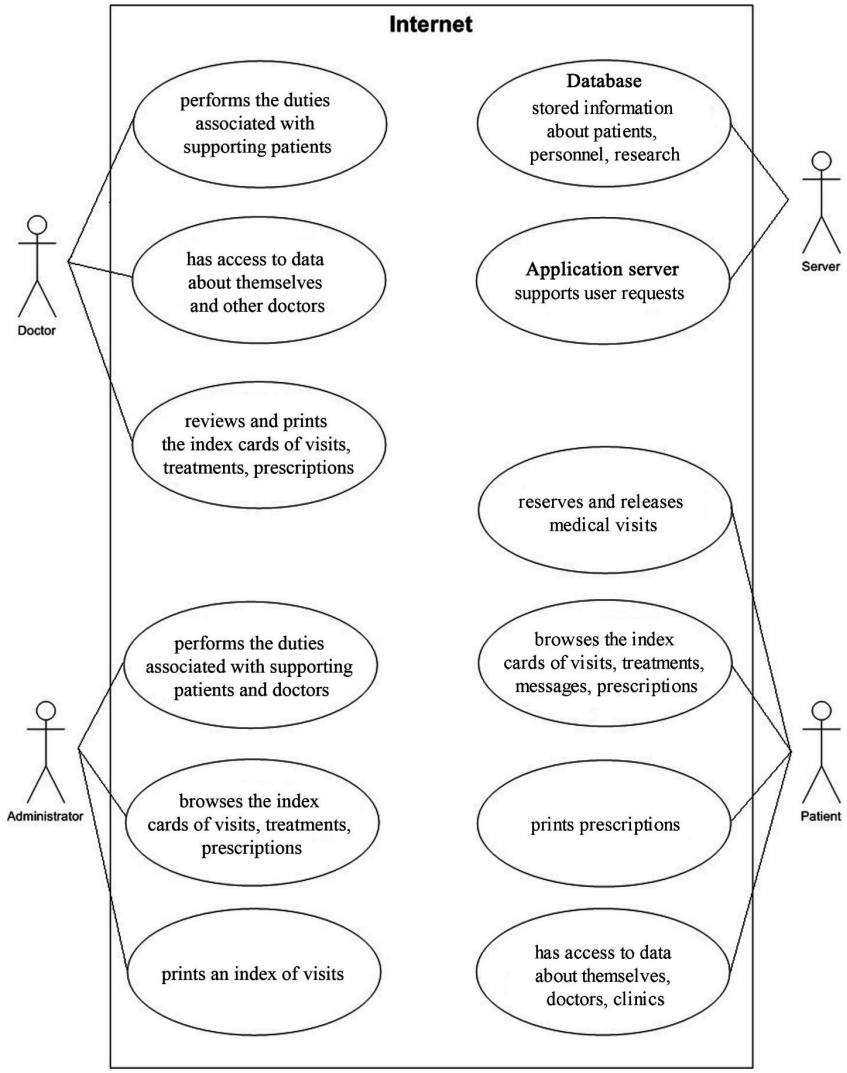


Fig. 1. The system concept

Rys. 1. Koncepcja systemu

4.2. System Model Information Flow

The below chart illustrates data processing in each iteration. The starting point depends on the authorization granted based on the assignment to one of the three user groups.

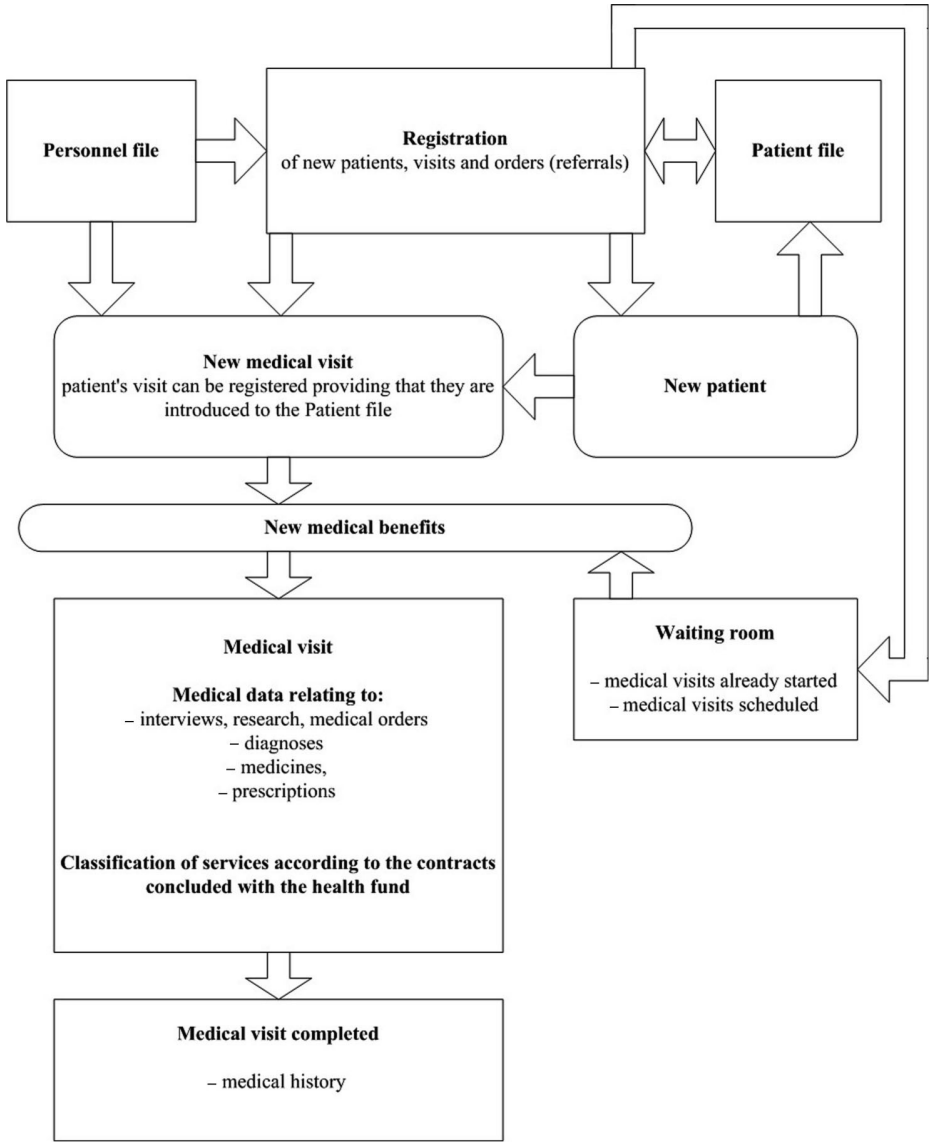


Fig. 2. System Model Information Flow

Rys. 2. Model przepływu informacji w systemie

According to the chart, the system contains the following modules:

- **Personnel and patients file** – for managing information regarding the doctors and the patients of the clinic;
- **Registration** (reservation) – for planning and booking new visits, patients, making accessible the schedule which includes the doctors' allocations and their office hours, daily patients limits;

- **Waiting room** – for browsing through the planned and started visits, ordered by the date and hour of the appointment and by the names of the doctors assigned to carry out the planned visit;
- **New medical visit** – collecting all information regarding the patient's appointments, enabling printing prescriptions;
- **New medical benefits** – collecting all information regarding medical records and list of services according to the contracts concluded with the health fund;
- **Medical visit completed** – for developing medical history records, issuing prescriptions, reporting the registered and elaborated data during the visit.

4.3. System Requirements

A proper operation of the system requires the provision of appropriate hardware and network infrastructure. The patient and the doctor find it convenient. It is enough for the system to operate on mid-range computer with access to the internet and mainstream pre--installed software (web browser). A key role of the system is to guarantee adequate storage space of the database and the continuous high-speed internet.

4.4. A system architecture

The structure of the system is based on a three-tier architecture [15]. Each of the modules created for the system contains components from one of the three application server layers enumerated below:

- **the presentation tier** consists of the user interface elements of the site, and includes all the logic that manages the interaction between the visitor and the client's business. This tier makes the whole site eye-catching, and the way it is designed is of a crucial importance for the site's success;
- **the business tier** (also called the middle tier) adopts requests from the presentation tier and returns a result to the presentation tier depending on the business logic it contains. Almost any event that happens in the presentation tier usually results in the business tier being called (utilized), except events that can be handled locally by the presentation tier, such as simple input data validation, and so on;
- **the data tier** (sometimes referred to as the database tier) responsible for managing the application's data and sending it to the business tier when requested.

These tiers are purely logical – there is no constraint on the physical location of each tier. In theory, it is possible to place all of the application, and implicitly all of its tiers, on a single server machine, or place each tier on a separate machine if the application permits this [15].

As the below chart shows, most of the data shown to the system users are retrieved from the relational open source database system. As mentioned before, the database is free of charge, fast and reliable. It is also significant, that many companies which provide internet services offer an access to this sort of databases systems, which facilitates launching new websites. This is the reason why this particular platform has been chosen for creating the system prototype.

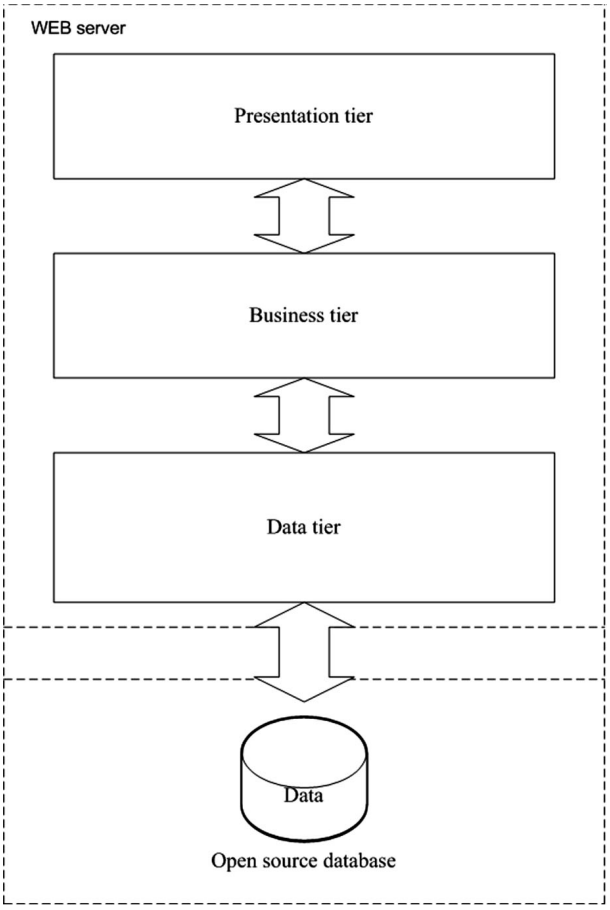


Fig. 3. Open source database server in three-tier architecture

Rys. 3. Serwer bazy danych open source w trójwarstwowej architekturze

It should be mentioned that the important constraint in the three-layered architecture model is that information must flow in sequential order between tiers. The presentation tier may communicate only with the business tier. It is impossible to approach the data tier directly. The business tier constitutes the 'brain' of the application, which communicates with the remaining tiers and processes and coordinates all the information flow.

If the presentation tier directly accessed the data tier, the rules of three-tier architecture programming would be broken [15, 16]. When you implement a three-tier architecture, you must be consistent and obey its rules to reap the benefits.

In order to make the data transfer and processing within separate system tiers understandable, an uncomplicated example will be analyzed: 'The display of a patient's personal details' when the 'My data' button is pressed by the patient. Also it will be indicated which technologies/protocols can be used for a given tier.

When the visitor clicks on the 'My data' button associated with a particular patient (Step 1), the presentation tier (which contains buttons and can be made by means of smarty

componentized templates) forwards the request to the business tier. Information is the order type of 'display personal data of the patient' (Step 2).

The business tier (which can be made in PHP – open source technology for building dynamic, interactive web content) receives the request and interprets it as a command to show the personal data of a particular patient. The service request is forwarded to the command data tier to search for a patient's data (Step 3). The data tier needs to be called because it stores and manages the entire web site's data.

The data tier (in which can be used MySQL) searches for the specified personal information of the patient (Step 4) and eventually returns a success code to the business tier. The business tier (the software instructions to access data (SQL) can be placed in several classes PHP) handles the return code (Step 5) and any errors that might have occurred in the data tier and then returns the output to the presentation tier.

Finally, the presentation tier generates a view of personal information of the patient (Step 6). The results of the execution are wrapped up by generating a HTML web page that is returned to the visitor (Step 7). The personal data of the patient is then displayed in the browser of the patient.

Note that in this simple example, the business tier doesn't do a lot of processing and its business logic isn't very complex. However, if new business rules appear for your application, you would change the business tier. In any case, the presentation tier is informed about the status and provides human-readable feedback to the visitor.

4.5. Prospects for the development of the system

A web-based system designed to support a health clinic's functioning should develop very dynamically. The construction of the system's individual parts [17] should change along with the changing legal situation in health care. Continuous cooperation with users (patient, doctor, administrator) should allow for the insertion of facilities for even more efficient operation.

To improve the performance of the existing system the following modules [14] could be added:

- **Statements** – a module would enable the storage of information concerning declarations that has been made (or withdrawn) by patients for primary care physicians within the General Health Insurance. That could control the number of declared patients in the office or clinic and generate reports to the National Health Fund of the declarations made to a particular provider;
- **Medical benefits** – the module would be responsible for the registration of all orders of diagnostic tests, laboratory tests and treatments. Service orders would be controlled from the adoption referrals to the service until the end of the job, which is to be settled with the payer;
- **Refunds** – a module would be responsible for the allocation of costs of the services of the various Departments of the National Health Fund – in accordance with the principles of healthcare financing. The system automatically would generate required reports by the National Health Fund. Thus, a medical unit could be reimbursed for expenses connected with treating patients;
- **Lists** – with increased options for generating different configurations (number and types of services provided, financial statements, cost, etc.) the module allows you to obtain information on the results of the medical unit quickly;

– **Base** – a module would be responsible for managing all databases (dictionaries), e.g. diagnosis, medical services, medicines, dictionary print forms allowing them to define new models. etc. The following dictionary database would be delivered with the system (supplemented partly or as a whole): an address book Branches of the National Health Fund, an address book of the country, the medicine database KS-BLOZ, the diseases database according to X revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) – three or four characters and database of medical procedures (International Classification of medical procedures: ICD-9-CM).

4.6. A system infrastructure – conclusions

The solution of this infrastructure supports the following arguments:

– availability (each user of the system with an account can use it from the internet);
– possibility to store data (patients, doctors) as well as all documentation (visits, cards, prescriptions) in one place – a relational database open source;
– intuitive user system and thus diminishing the education needs of the system;
– reduction of documentation expenses;
– better organization of work and the use of personnel;
– using the system only when necessary, using it to such an extent that the user is interested in (modularity);
– the PCs and the internet are currently available in each clinic (the result of savings, they do not need to buy new hardware).

5. Conclusions

The purpose of this paper was to show the possibility of using an open source relational database system to build a web-based system designed to support health clinic's functioning as an alternative to commercial solutions.

It should be noted that the choice of database system determines the fact how the system will operate in real-world conditions to a large extent. The time of creation of such a system also depends on the database.

The results of standard tests carried out on large tables of data confirmed the assumptions and demonstrated that the database systems of this type are fast, reliable and above all, free of charge. It is also significant that many companies which provide internet services offer access to these databases, which further supports the choice of such a solution.

A prototype system has been designed (by developing a flexible architecture) in such a way that it can be extended with new modules (which can be implemented as separate components) or existing ones may be verified to facilitate the work of a doctor or patient service.

The prototype indicated a relational database open source as a basis system, thus proving the assumption that the medical system does not have to strain the health care budget, while providing an acceptable level of services.

References

[1] Open Source Initiative, http://opensource.org/ (date of access: 22.06.13).

[2] Gonzales-Barahona J. M., Heras Quiros P., Bollinger T., *A brief history of free software and open source*, IEEE Software, January 1999, 32-33.

[3] Gacek C., Lawrie T., Arief B., *The many meanings of open source*, IEEE Software, vol. 21, 2004, 34-40.

[4] James Hernandez M., *Database Design for Mere Mortals: A Hands-On Guide to Relational Database Design (2nd Edition)*, Sybex, 2003.

[5] Whitehorn M., Marklyn B., *Inside Relational Databases 2nd Edition*, Springer, 2001.

[6] Beynon-Davies P., *Database Systems*, Palgrave Macmillan, 2003.

[7] Networld, 2013, http://www.networld.pl/artykuly/387854/ (date of access: 22.06.13).

[8] MySQL http://www.mysql.com/products/, (date of access: 22.06.13).

[9] HeidiSQL http://www.heidisql.com/, (date of access: 22.06.13).

[10] Apache Derby, http://db.apache.org/derby/, (date of access: 22.06.13).

[11] pgADMIN, http://www.pgadmin.org/, (date of access: 22.06.13).

[12] FlameRobin, http://www.flamerobin.org/, (date of access: 22.06.13).

[13] Berkus J., *Ten ways to wreck your database*, O'Reilly Webcast, July 2009.

[14] Pietka E., *An integrated information system for hospital work*, PWN, Warsaw 2004.

[15] Darie C., Balanescu E., *Beginning PHP and MySQL E-Commerce: From Novice to Professional*, *Second Edition*, Apress, 2008.

[16] Software Developer's Journal, 2011, http://sdjournal.pl/magazine/1758-projektowanie-systemow-informatycznych-w-praktyce/, (date of access: 22.06.13).

[17] Coelho F., *Database quality survey projects and results*, January 2012, detailed list of projects surveyed in *On the Quality of Relational Database Schemas in Open Source Software.*

TECHNICAL TRANSACTIONS | CZASOPISMO TECHNICZNE
AUTOMATIC CONTROL | AUTOMATYKA
1-AC/2013

ŁUKASZ ŚCISŁO*

CONTROLLER–SENSING ELEMENT COMMUNICATION USING UDP PROTOCOL

KOMUNIKACJA STEROWNIK–URZĄDZENIE POMIAROWE Z WYKORZYSTANIEM PROTOKOŁU UDP

Abstract

The main goal of this paper is an attempt to create a new approach to measuring equipment-PLC controller communication. Beside the individual communicational protocols supported and developed by PLC manufacturers, PLC controllers support also popular protocols like TCP or UDP. This allows an easy communication with computers and many devices supporting these protocols. This paper presents a control system, which allows the distance measurement where the data is sent to the PLC controller (SIEMENS S7-1200) using Ethernet bus. The data can be also accessed using Human Machine Interface on an operators screen. In the paper it is shown how to create such a communication system which can be implemented for measurement purposes and is ready for farther development in the future.

Keywords: Ethernet, UDP,UART, PLC, control systems, HMI, Profinet, industrial automation

Streszczenie

Głównym celem niniejszego artykułu jest opracowanie nowego podejścia do komunikacji urządzenie pomiarowe–sterownik PLC. Sterowniki przemysłowe, oprócz własnych protokołów komunikacyjnych, coraz częściej obsługują popularne protokoły komunikacyjne, takie jak TCP czy UDP. Umożliwia to komunikację z komputerami PC oraz wieloma urządzeniami obsługującymi wyżej wymienione standardy. W artykule pokazano sposób zaprojektowania i realizacji układu wykonującego pomiar odległości, który wyniki pomiaru przesyła do sterownika programowalnego SIEMENS S7-1200 za pośrednictwem magistrali Ethernet i protokołu UDP oraz do systemu kontroli na ekranie operatorskim HMI. Zaprezentowano sposób realizacji systemu komunikacji, łatwość jego implementacji dla systemów pomiarowych i możliwość rozbudowy o kolejne urządzenia.

Keywords: Ethernet, UDP, UART, sterowniki przemysłowe, HMI, Profinet, automatyka przemysłowa

* Łukasz Ścisło, M.Sc., Department of Automatic Control and Information Technology, Faculty of Electrical and Computer Engineering, Cracow University of Technology.

1. Introduction

Modern production processes require the use of very accurate and reliable controlling devices. For such a task Programmable Logic Controllers can be a very good choice. PLC`s are being widely used for controlling devices, machines and whole technological processes [7, 14, 15]. They have replaced systems based on transmitters. The advantages of using PLC controllers are their high reliability, relatively low costs, compact built and module structure which make the further development of a control system possible [9].

One of the most important advantages of controllers are many different protocols which can be used for communication purposes (TCP, UDP). Most of the modern controllers are provided with interfaces which allow connecting them into networks [6]. Moreover, most of the new PLC`s, HMI screens and programmators can communicate using Ethernet protocol (e.g. Siemens S7-1200 which communicates with computers and HMI devices using Ethernet/ Profinet protocol), it is difficult to find sensing equipment which can in an easy way send/ receive data from the PLC, using similar kind of communication [3, 4, 10, 12].

The paper has two main purposes. The first one is to show the way to use modern protocols for acquiring measurement data, which can be sent to any kind of the control platform. The second purpose is to present a method of creating a UDP connection for Siemens PLC family which is not mentioned in professional journals.

The article will be divided into four chapters. The first one is an introduction to the PLC systems; the examined problem is introduced and an explanation of the paper's aim is presented. The second chapter concentrates on problem analysis. The architecture of the proposed communication between measuring equipment and PLC controller is introduced. Moreover, the way to create such a communication system is explained in detail. The following chapter shows the test of laboratory prototype and the obtained results for the distance measurements. The last chapter consists of the summary and the description of possible future development of the proposed system.

2. Problem analysis

For many modern control systems, especially responsible for continuous monitoring of crucial elements, it is very important to obtain current measurements as fast as possible. It is especially needed for fast location of faults in mechanical systems [16]. Using programmable logic controllers it is possible to use many different communication methods. This paper especially shows the possibility of using UDP protocol instead of very popular TCP. UDP communication can increase significantly data transmission which can be very useful in industrial control systems.

The proposed system includes programmable logic controller Siemens S7-1200, HMI operators screen, two-way Ethernet/UART converter, microcontroller with analogue-to--digital converter and a distance sensor. Siemens S7-1200 includes two modules: central processing unit 1214C AC/DC/RLY and a network switch CSM 1277 SIMATIC NET. All the parts of the proposed system and their communication concept are shown in Fig. 1.

The measurement data from a range-finder is sent through Ethernet bus with the use of UDP protocol. The range-finder measures the distance from the barrier, data is being sent to the microcontroller which converts the data in a proper way and sends it to the PLC. The current measurement will be observed on HMI screen. The operator is able to decide when to start and stop the measurement and what the measurement parameters are.

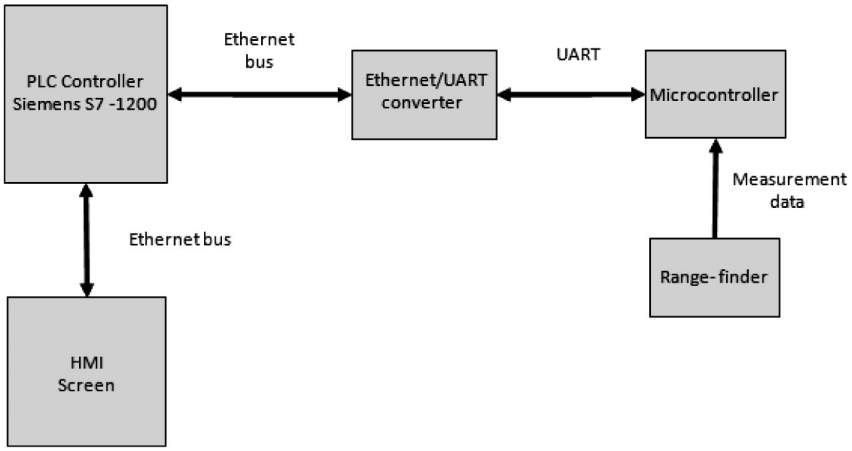


Fig. 1. The block diagram of the proposed system

Rys. 1. Schemat blokowy proponowanego systemu

2.1. UART Communication

UART (*Universal Asynchronous Receiver and Transmitter*) is a device used for serial asynchronous communication, which allows for an easy data transmission in a RS-232C standard [11]. UART has two lines for data transfer. For sending data "Tx" line is used and for receiving data "Rx" line is used (Fig. 2). Data is being transferred in frames which start from a start bit and next bits are data bits. Typically, there are 5 to 8 data bits [13]. The frame is ended with one or two stop bits. UART is serial asynchronous communication, and that is why no additional synchronizing signal is needed. Therefore, less conductors are needed for the transmission. However, in devices communicating using UART, it is important to set the same data transmission speed and the same frame format. Most of the modern microcontrollers have hardware support for serial UART communication. This communication can take place in half-duplex mode, which means that in the specific time only one device transmits data, or in full-duplex mode (both devices transmit data at the same time).

However, the serial communication is slower than parallel, so it is important from automation machines control point of view that much less wires are being used, which is a particularly great advantage when measurement data needs to be transmitted over long distances.

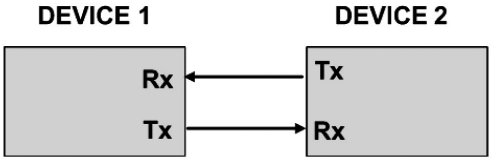


Fig. 2. UART communication

Rys. 2. Komunikacja UART

2.2. Ethernet/UART converter configuration

ATV5340 system converts messages received from programmable logic controller which are being sent by Ethernet bus. UDP protocol is used to communicate using the Ethernet bus. Finally the data is sent to microcontroller by UART interface. The configuration of AVT5340 system is made using a computer which is in the same network as ATV5340. In the web browser it is possible to insert the IP address assigned to the converter.

In the *network* tab the following attributes can be changed:
- MAC – only singular change is possible,
- IP – which is an address of the converter in the network,
- comment,
- subnetwork mask – for the converter,
- gate – parameter used when the converter communicates through the router.

In the *Hosts* tab IP addresses and port used for sending the data can be configured. In the COM tab two other parameters can be set:
- COM – which is responsible for transfer speed using UART interface. The choice between 2.4 kb/s to 1 Mb/s is possible,
- Overtime.

The converter can start data transmission using Ethernet bus when one of the two conditions is met:
- the buffer is full,
- the specific time since the first UART frame received is reached (time set in the Overtime parameter).

2.3. Microcontroller configuration

The microcontroller operates with 1 MHz frequency. To start voltage measurement and data transmission the correct command needs to be received by the microcontroller. The command format is shown in Fig. 3.

Bit 7	Bit 6	Bit 5	Bit 4	Bit 3	Bit 2	Bit 1	Bit 0
Not used					Sampling period		Measurement

Fig. 3. Command format

Rys. 3. Format rozkazu

Where bits:
- 2 and 1 – are responsible for sampling time:
  - 00 – measurement every 1 s,
  - 01 – measurement every 500 ms,
  - 10 – measurement every 250 ms,
  - 11 – measurement every 125 ms.

– 0 – start and end of the measurement:
  • 0 – measurement stop,
  • 1 – measurement start.

Commands which contain not used bits cause the measurement to stop (regardless of the bit 0 value). When the UART frame is received from the converter (connected to the microcontroller) an interruption is generated and the data received is compared to the command stored in the microcontrollers' memory. Four possible settings of voltage sampling periods are available. The change of the sampling period takes place when T0 timers' frequency divider is changed. The first voltage measurement takes place immediately after the command is received and the next three are taken every 20 ms from the first measurement. The results of the measurements are averaged and sent by serial connection to Ethernet/UART converter, which sends it farther to the PLC controller (S7-1200). The format of this data is 16-bit word, which is later translated to the distance value by the PLC.

Timer T0 is responsible for sampling period and every following measuring cycle starts with the interruption generated by the timer. Every interruption of T0 timer starts counting by timer T1. This timer generates interrupts every 20ms which start range-meters` voltage measurement. T1 executes three such cycles and goes to stop state. The measurement (sampling and A/D conversion) lasts about 0.5 ms and the whole measurement cycle lasts about 60.5 ms plus few clock cycles needed for voltage value averaging. That relatively long measurement cycle was introduced only because the range-meter used in the laboratory model has very low amplitude of oscillations of output voltage. When the measurement of long distances is introduced, the small change in output voltage causes significant changes in measured distance (that is one of the reasons to perform multiple measurements which are later averaged).

The microcontroller communicates with the AVT5340 converter with 9.6 kb/s velocity. The transmission time of 16 bit word takes about 2.1 ms. Both AVT5340 converter and microcontroller MSP430G2553 can reach up to 1 Mb/s transmission speed, but this requires the use of microcontrollers with the frequency of 16 MHz.

2.4. PLC and HMI configuration

PLC LAD application is created in the TIA (Totally integrated automation) Portal environment for the purpose of measurement process standard control. Moreover, to allow an easy access and presentation of the data, an additional HMI application is designed for KPN600 Panel. TIA is a system in the automation technology field which has been developed by Siemens for the last few years [5]. This strategy defines the interaction of extensive single components, tools and the services to achieve a full automation solution for the problem. The interaction performs integration across the four automation levels of the automation pyramid: management level (ERP, MES), operator's level (SCADA), controller and field's level (commonly called floor control level) level (Fig. 4) [8].

Although, usually the control level consists only of PLC or PLC with HMI panel it is proposed to include also other devices like: PDA's, smartphone's and tablets. All those devices are really able to connect directly to the PLC, but current fast development of network communication for programmable logic controllers including Ethernet and GSM communication may suggest that it is a matter of a short time for such application to be used

for control purposes. Developing a fully functional Android application for PLC – operator communication is currently under development. For this paper purpose only floor control level is considered (Fig. 4).

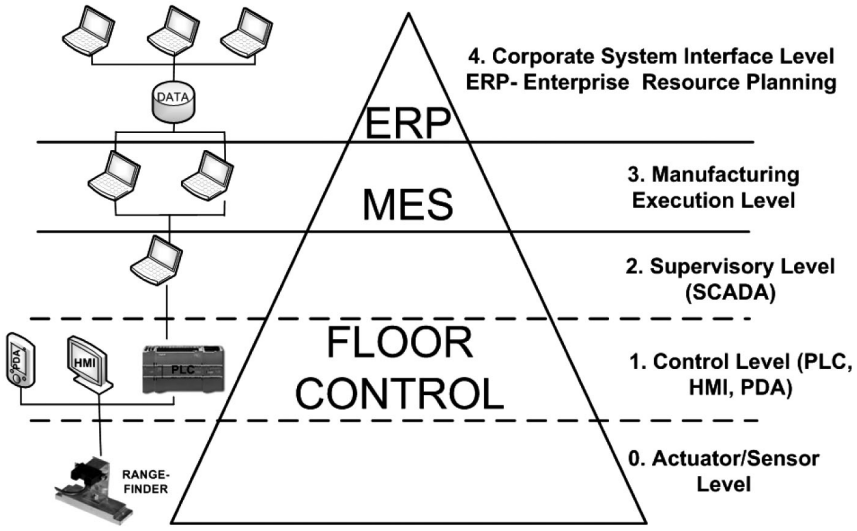


Fig. 4. Automation pyramid concept

Rys. 4. Piramida automatyzacji i zarządzania produkcją

There are two requirements for PLC control program for the proposed device:

– it has to allow the communication using UDP protocol,
– it has to be able to convert measurement data received from the converter.

To allow UDP connection three of the program blocks need to be configured (Fig. 5):

– TCON – which allows for the manual set-up of the connection with TCP, ISO-on-TCP and UDP protocol. TCON sets the connection parameters and gives the connection identification number,
– TUDEND – allows data transfer using UDP protocol,
– TURCV – allows data to be received using UDP protocol.

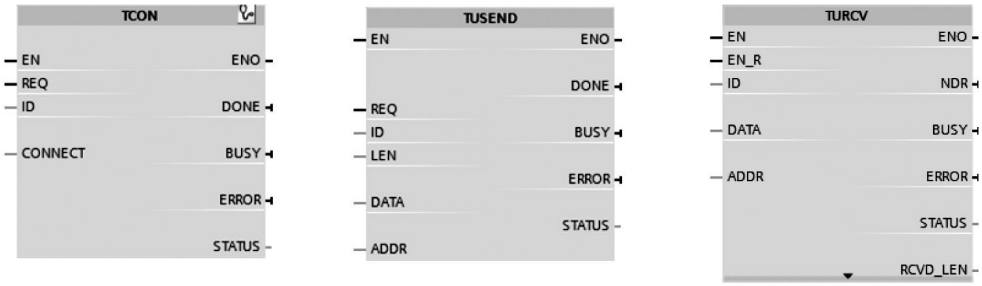


Fig. 5. PLC program blocks which need to be configured for UDP communication

Rys. 5. Bloki programowe PLC wymagające konfiguracji dla komunikacji UDP

The PLC program allows also for the conversion of the voltage measured by the range-finder and sent by the microcontroller. The value of the voltage is calculated using fallowing equation:

$$U = \frac{N_{\mathrm{ADC}} \cdot 3{,}3}{1023} \tag{1}$$

where:

U – voltage value,
N_{ADC} – value sent by the microcontroller.

3. Laboratory test of proposed system

The laboratory demonstration the device shown in Fig. 6 is used. The equipment consists of the microcontroller (MSP430G2553), Ethernet/UART converter (AVT5340) and a power supply unit. The system requires 3.3 V (microcontroller, converter) and 5 V (range-finder).

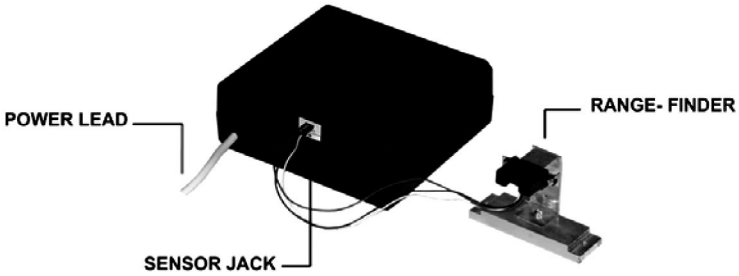


Fig. 6. The overview of the laboratory test set

Rys. 6. Części składowe układu laboratoryjnego

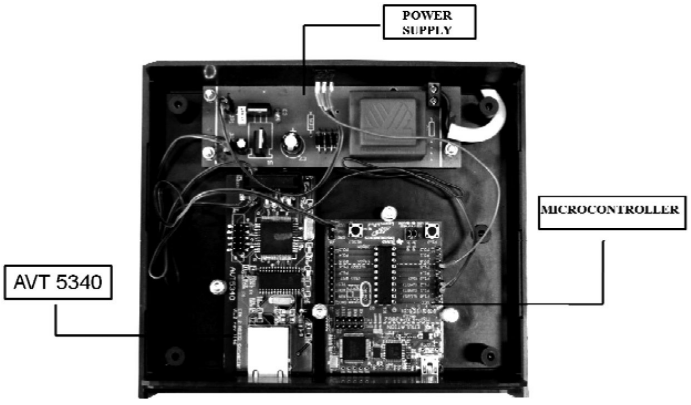


Fig. 7. The inside view if a laboratory test set

Rys. 7. Widok wnętrza układu laboratoryjnego

The configuration of the measuring cycle of the microcontroller is shown in Fig. 8:
- the measurements are held every 125–1000 ms,
- sampling time – 60.5 ms,
- data transmission to the converter – 2.1 ms.

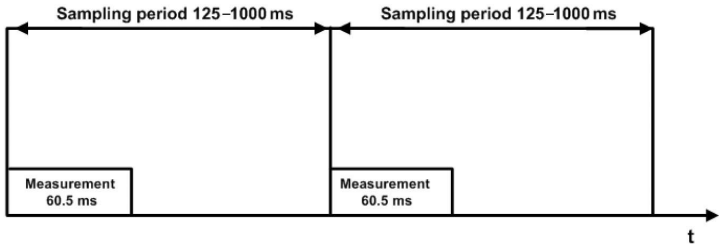


Fig. 8. Microcontrollers measuring cycle diagram

Rys. 8. Diagram cyklu pomiarowego mikrokontrolera

For a laboratory test of the proposed device two applications are used:
- PLC (Siemens S7-1200) application using TIA Portal environment,
- Human Machine Interface application for operators screen (KPN600).

TIA Portal is a new environment for industrial automation applications – especially PLC and HMI systems. It includes of STEP7 and WinCC environments. STEP7 is a system for PLC configuration and programming, which allows for using LAD and FBD languages for programming [7, 11]. WinCC flexible is a system for development of industrial operators' panels applications. The advantage of TIA Portal is an universal access to the program variables. The tags created in one of the programs mentioned above can be used by other applications and by other devices, which allows for an easier creation of a fault free control programs.

The test screen gives the possibility of choosing one of the three buttons (Fig. 9):
- measurement start – which causes a message to start the measurement to be sent to the microcontroller,
- measurement stop – allows to stop the measurement,
- options – opens a screen with additional options (Fig. 10).

Additionally, the field in the top right corner allows to choose a different sampling period. This option allows for a selection of up to four different values and a choice of a different period than the actual one which does not require a pause in a current measuring process.

Additional options allows the user to set the new variables:
- change of IP address and port of the device to which PLC wants to be connected to,
- change of the approximation function for distance measurement according to the voltage measurement (depends of the sensor used).

Figure 11 shows two of the characteristics where distance to output voltage fuction is obeserved. The rang-finder used for this device causes a small change in the output voltage when measuring long distances.

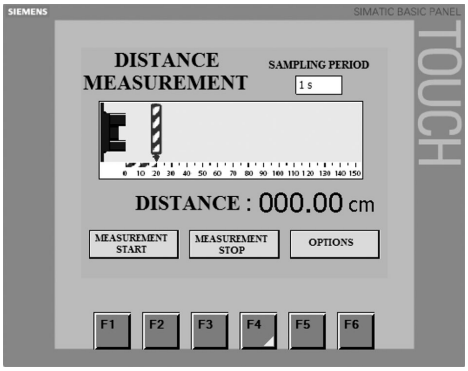


Fig. 9. HMI measurement test screen

Rys. 9. Ekran HMI pomiaru odległości

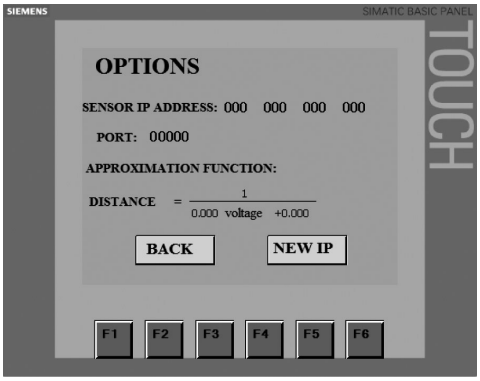


Fig. 10. HMI additional options screen

Rys. 10. Ekran HMI dodatkowych ustawień

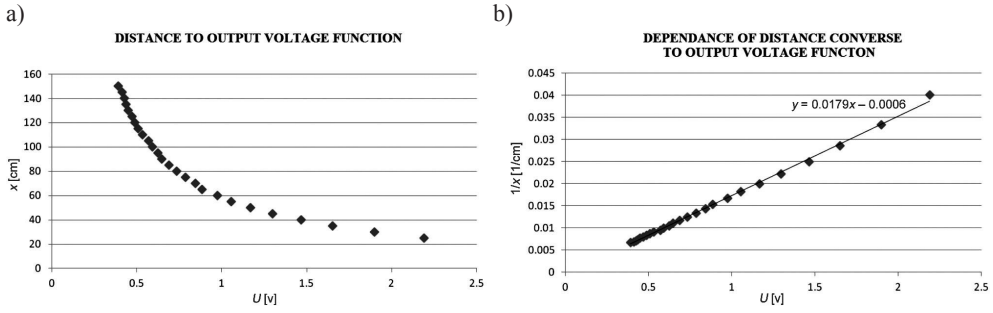


Fig. 11a) Distance to output voltage function (measuring range 25–150 cm), b) dependance of distance to output voltage function of the range-finder (measuring range 25–150 cm)

Rys. 11. Rys. 11a) Odległość w funkcji napięcia wyjściowego dalmierza (w zakresie od 25 do 150 cm), b) odwrotność odległości w funkcji napięcia wyjściowego dalmierza (w zakresie od 25 do 150 cm)

4. Conclusions

Usually, when sending measurement data to the programmable logic controller, analogue or digital modules of the controllers are used. In modern systems a wide range of network protocols is being used as well. The main advantage of using UDP communication comparing to TCP/IP is that it is much faster. It does not restrict the user to a connection based on communication model, so startup latency in distributed applications is much lower, as is operating system overhead. All the main settings like: flow control, acknowledging, transaction logging, etc. are up to program users. This means that the system user can only use the features that he really needs. The great advantage of UDP is also that broadcast and multicast transmission are available for the programmer. As it was mentioned the UDP allows a high transmission speed, with the maximum speeds enforced only by real network bandwidth. It must be remembered that actual speed is agreement of sender and receiver. The higher speed is an effect of two things:

– in case of the TCP/IP the receiver sends an acknowledgement to the sender to show that the data has been successfully received or has to be repeated. UDP protocol only acknowledges a successful sending of the data into the network and not the arrival of the data at the target station. The user program must take care of securing consistency and data preparation,
– the UDP header is rather short, compared to the TCP/IP header, therefore the UDP telegram can be created and processed much faster than the TCP.

Although it may seem that UDP is really unreliable, in case of continuous measurements where data is sent to PLC even when some packets are missing, it is no problem because usually the acquisition cycle is much slower for the user to be aware of missing the data. It must be also stated that UDP is suitable for small-to medium volumes of data (data volume: 1–2048 bytes) and TCP is suitable for transferring medium-to-large data volumes (data volume: 1–8192 bytes).

It is proposed to use a UDP protocol for measurement transmission to PLC instead of TCP. The biggest advantage of this, apart of higher speed, is that the configuration of devices and development of the software is much easier. Also, devices using this protocol can be used with any programmable logic controller regardless of the brand.

However, the presented in this article device is only used for distance measurement, it also can be used to connect other measurement equipment. Receiving data simultaneously from multiple hosts is much easier with UDP than with other protocols. The aim of the project is to find an efficient way for communication of industrial sensor equipment with the PLC controllers using Ethernet protocol. The laboratory test station allows not only to send data to such equipment, but also to any device supporting UDP protocol. However, the typical sensor is usually connected to the control device like PLC but using UDP communication can be used separately and can be, for example, connected directly to the computer.

The designed device is fully functional, however, it uses only a part of its potential because the rest fifteen input/outputs of the microcontroller are not used. Additionally five of those can be configured as analogue inputs. The device can be significantly expanded by means of connecting other sensing equipment. Moreover, Ethernet/UART AVT5340 converter, which is used for the device has two serial UART communication lines which allows to connect second microcontroller. This gives even greater possibility of farther development and operators availability to connect other sensors or actuators.

The concept of connecting measurement equipment using one of the internet protocol suite members gives an easier way to collect data for upper levels of automation pyramid (Fig. 4) or send the data directly to databases on the MES or ERP level. Such an approach can be used for distributed production scheduling. It can be very flexible, requirement-driven and reconfigurable so it acts as an open manufacturing system which can easily adapt to raid changes in market demands. Faster sensitive data transfer, using network protocols, in a more distributed way (accessed through PDA, phone, tablet and etc.) allows fast response to emergencies (machine failure, operator's absence, material shortage) and gives solution to a production problems (especially scheduling).

References

[1] Berger H., *Automating with SIMATIC S7-1200: Configuring, Programming and Testing with STEP 7 Basic V11; Visualization with WinCC Basic V11*, Publicis Publishing, 2013.

[2] Berger H., *Automating with SIMATIC S7-400 inside TIA Portal*, Publicis Publishing 2013.

[3] Berger H., *Automating with SIMATIC*, Publicis Kommunikationsag, 2003.

[4] Berger H., *Automating with STEP 7 in STL and SCL*, Publicis Corporate Pub, 2005.

[5] Berger H., *Automating with SIMATIC S7-300 inside TIA Portal*, Wiley-VCH, 2012.

[6] Bolton W., *Programmable Logic Controllers – Fifth Edition*, Newnes, 2009.

[7] Borden T., Cox R., *Technician's Guide to Programmable Controllers*, Cengage Learning, 2012.

[8] Koen P., Stromsdorfer C., *Distributed Applications in Manufacturing Environments*, The Architecture Journal, 10/2008, 40-44.

[9] Mitchell R., *PROFIBUS: A Pocket Guide*, ISA: The Instrumentation, Systems, and Automation Society, 2003.

[10] Müller J., *Controlling with SIMATIC*, Siemens, 2005.

[11] Petruzella F., *Programmable Logic Controllers*, McGraw-Hill, 2010.

[12] Pigan R., Metter M., *Automating with PROFINET*, Publicis Publishing, 2008.

[13] Raijkumar N., *Communication and Debug Platform for Smart Antenna DOA System: UART System Design in VHDL and Debug Platform in Labview*, VDM Verlag Dr. Müller, 2010.

[14] Ścisło Ł., *Multifunctional Control of a Proportioner*, Proceedings of the PhD Students and Young Scientists Conference, 2009.

[15] Ścisło Ł., *Integrated Control of the Heat Pump Heating and Cooling System*, Proceedings of the PhD Students and Young Scientists Conference, 2010.

[16] Zając M., *Wavelet Analysis for Location of Faults on Drive Transmission System*, Technical Transactions, vol. 1-AC/2012, Cracow University of Technology Press, 139-156.

TECHNICAL TRANSACTIONS | CZASOPISMO TECHNICZNE
AUTOMATIC CONTROL | AUTOMATYKA
1-AC/2013

VOLODYMYR SAMOTYY*, ULYANA DZELENDZYAK**

MATHEMATICAL MODEL OF THYRISTOR'S SYSTEM CONTROL OF DC MOTOR WITH INDEPENDENT EXCITATION

MODEL MATEMATYCZNY TYRYSTOROWEGO SYSTEMU STEROWANIA SILNIKIEM PRĄDU STAŁEGO Z NIEZALEŻNYM WZBUDZANIEM

Abstract

The result of the research in this article is a mathematical model of thyristor's system control of DC motor with independent excitation, which includes containing controlled three-phase rectifier. The nonlinear characteristics of transformer's magnetization was being taken into account. Tyristors' work in shown by means of a scheme of an ideal key.

Keywords: motor of direct-current, controlled three-phase rectifier, equation of dynamics, ideal key, logic variables

Streszczenie

W artykule przedstawiono model matematyczny tyrystorowego systemu sterowania silnikiem prądu stalego z niezależnym wzbudzaniem, który zawiera sterowany prostownik trójfazowy. Uwzględniono nieliniowe charakterystyki magnesowania transformatora. Działanie tyrystorów podano zgodnie z charaterystyką klucza idealnego.

Słowa kluczowe: silnik prądu stałego, sterowany prostownik trójfazowy, równania dynamiki, klucz idealny, zmienne logiczne

* Volodymyr Samotyy, Prof., D.Sc., Ph.D., Department of Automatic Control and Information Technology, Faculty of Electrical and Computer Engineering, Cracow University of Technology; Department of Computerized Automatic Systems, Lviv Polytechnic National University.

** Ulyana Dzelendzyak, Ph.D., Department of Computerized Automatic Systems, Lviv Polytechnic National University.

1. Introduction

To compare other motors with DC motors, the last have considerable starting torque, that's why they have proliferated in various actuators such as drives for electric, hoisting machines, and others. These motors are fed and followed by DC voltage, which is applied to the stator windings and armature motor. If there is no source of constant voltage it can be attained by transforming alternating voltage into constant. In such rectification, associated with the presence of pulsations, even when filters are used, the pulsation will appear, and as a result, we will have a deterioration of the dynamic characteristics of the system as a whole. Because of reducing such fluctuations, rectification of three-phase voltage should be applied. In this paper a mathematical model, which provides the use of three-phase half-wave rectifier with thyristors is considered. DC motor connected to the rectifier through capacitive filter, which provides filtration of output voltage ripple of rectifier.

Depending on combination of the motor windings, it is divided into engines with independent, parallel and consequent excitations. We consider the first type of combinations, when winding of excitation is fed using independent source of direct voltage. Armature's winding is fed by output voltage of rectifier in which thyristors are managed by external circuit. Changing the angle of thyristors opens this scheme and changes the output voltage of rectifier, and therefore angular velocity of motor's armature is also changed.

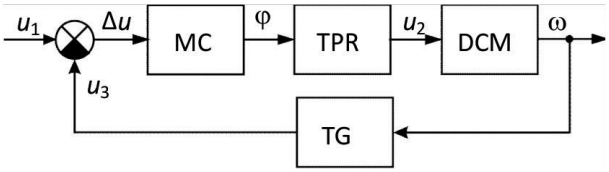


Fig. 1. Structural scheme of system's control of DC motor

Rys. 1. Schemat blokowy układu sterowania silnikiem prądu stałego

Anchor of motor is connected to tachogenerators, its output voltage is compared with the input voltage of control system. The achieved difference is transformed by microcontroller into angle of thyristors rectifier open. A block diagram of the system is shown in Fig. 1. The following notation are adopted: MC – microcontroller; TPR – three-phase rectifier; DCM – DC motor; TG – tachogenerators; u_1 – input system signal; Δu – system's error; φ – the angle of thyristors opening; u_2 – output voltage of rectifier; ω – speed of motor's armature; u_3 – output tachogenerators's voltage.

The development of a mathematical model of such a system is associated with certain difficulties, which are conditioned by nonlinear characteristics of its elements. It is the following features: magnetization cores of transformer, current-voltage characteristics of semiconductor valves and magnetic characteristics of the DC motor. The consideration of these types of nonlinearities makes the task of mathematical modeling extremely difficult. Therefore, to resolve it means to adopt certain assumptions which would take into account the most of elements' characteristics of the system and, at the same time, it must simplify algorithm for analyzing its dynamics, which is suitable for numerical implementation on a computer. The presence of such a model enables us to investigate the influence of parameters on the system, and therefore, using appropriate methods to perform its optimization with an aim of output improvement of the mechanical properties of the system "controlled three-phase half-wave rectifier – DC motor with independent excitation".

2. The review of literature

It is obvious that the closer the mathematical model of control system to real physical processes are, the more accurate the results of computer simulation we will achieve. Our problem will be not too cumbersome if we are restricted by problem with lumped parameters. It means that a mathematical model will be confined to the record system of nonlinear differential equations of state. Otherwise, the problem should be considered from the point of view of the theory of electromagnetic field. The target system can be considered as a certain nonlinear electromagnetic range, which contains a transformer, semiconductor valves, DC motors, tachogenerator and microcontroller.

An executive mechanism in robotics is one of the most widespread uses of motors [4]. To reach the maximum capacity and minimal losses, the phenomenon of superconductivity in electrical winding in low temperature can be used [5]. Linear mathematical models of DC motors are described in the works [1, 7]. Here the windings' inductive was not dependent on magnetic stream. The question of thyristor's modeling of voltage transformers is not examined in this work. The publications [2, 3, 6, 9], offer the following information: [2] – the regulator which was based on neuron network, [3] – illegible logic, [6, 9] – parametric optimization which is executed by genetic algorithm. In [13] an experimental approach of DC motor's parameter identification is observed. This question is especially important if the analysis of its work's dynamics is used. But for counting of periodical processes it will be better to use the extrapolation e-algorithm, because the method of Poincarego-Lindstedta [14] has discrepancy.

In [17] an adaptive observer estimating all parameters and load torque is proposed for DC servo motors. The observer uses no direct feedback but the adaptation schemes' use and speed measurements are actually valid. Both the observer and adaptations are simple to implement for real-time applications. Simulation results are satisfactory for the full adaptive observer. If the observer works in parallel with only load torque and armature resistance adaptations, the results are very good even if very low-quality sensors are used. In this simulation, only a single hall sensor is used as a rotational transducer, which produces a single pulse per revolution, and very high level noise and disturbance are added in order to provide a more realistic simulation.

However, we are more interested in electromagnetic devices which contain semiconductor valves. A large amount of work has been devoted to RLC-valve model [10–12]. Its main drawback lies in generating additional differential equations. Moreover, these equations are stiff, which makes the algorithm of solving more difficult.

Modern mathematical model of electromagnetic circles as an example of various electromagnetic devices can be met in papers [15, 16]. In these works the equation of dynamics is recorded concerning state's variables, which include working magnetic flux, current and voltage windings on the capacitors. In this method the equation of dynamics is recorded in the normal form of Cauchy, which is very handy when explicit numerical integration methods are used. It must be remarked that to make the record more compact the equation of dynamics must be written in matrix form. The example is shown in work [8].

In [18] the mathematical models of thyristor converters, thyristors are modeled on the first proposed scheme of perfect key. Here, there is no disadvantage, which is inherent in the models where switching functions, namely duplication of equations of state are used. The introduction of a new concept of additional logical variables, which can take the values 0, ±1,

makes it possible. These variables are governed by the terms of opening and closing valves. It gives an opportunity to describe the whole – a set of states of open and closed valves of electromagnetic transducers as a single system of equations.

3. The equation of dynamics system of thyristor DC motor's control with independent excitation

First of all, we consider the dynamic equation dynamic three-phase's half-wave rectifier with capacitor in load. Parameters of the secondary winding must be presumed as leading to the use of the number of winding to the original. In fact, the equation of three-phase's half-wave rectifier is reduced to a three-phase transformer with zero conductors, though the load of current is interfluent. For equations of state writing it is necessary to determine the total number of combinations of open and closed thyristors. To write an equation for each combination and with additional logical variables, to compile them into one system of algebraic-differential, equations are also necessary.

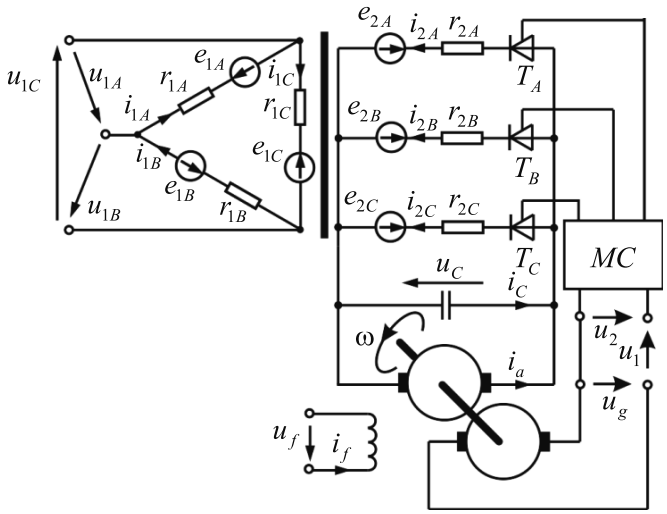


Fig. 2. Thyristor's system control of DC motor with independent excitation

Rys. 2. Tyrystorowy układ sterowania silnikiem prądu stałego z niezależnym wzbudzeniem

In this scheme a large number of such combinations is available (Fig. 2). You must examine all options when there is one or two gates will be open. So, there are three combinations when one valve remains open, and three combinations when two gates are open, namely:

1) T_A – open, T_B, T_C – closed,
2) T_B – open, T_A, T_C – closed,
3) T_C – open, T_A, T_B – closed,
4) T_B, T_C – open, T_A – closed,
5) T_A, T_C – open, T_B – closed,
6) T_A, T_B – open, T_C – closed.

The combinations of two are also possible, when all valves are closed and all open, but such regimes almost never occur.

Closing of any valve will result in disabling the corresponding phase. When writing equations for three-phase transformer the switch off of phases can be modeled, equating inverse scattering inductance corresponding phase to zero, or multiplying it on additional logic variable, which takes the values of 0, 1. So there is no need for the transducer to record the equation of all combinations of open and closed valves. Just write the equation for the case when all the valves are open.

Equation circuit of primary side of the transformer can be written in a matrix form:

$$\frac{d\Psi_1}{dt} = U_1 - R_1 I_1 \tag{1}$$

where:

Ψ_1, U_1, I_1 – matrix columns of complete linkages, voltages and currents $h(h = \Psi_1, U_1, I_1) = (h_A, h_B, h_C)^T$,

$R_1 = \text{diag}(r_{1A}, r_{1B}, r_{1C})$ – diagonal matrix of resistance.

Subscripts A, B, C indicates the involvement in the relevant phases, 1 indicates the involvement in the primary side of the transformer.

Equation of secondary side can be written in a matrix form:

$$\frac{d\Psi_2}{dt} = -H_C u_C - R_2 I_2 \tag{2}$$

where:

Ψ_2, I_2 – column matrix of complete linkages and currents $h(h = \Psi_2, I_2) = (h_A, h_B, h_C)^T$,

$R_2 = \text{diag}(r_{2A}, r_{2B}, r_{2C})$ – diagonal matrix of resistance,

$H_C = (1, 1, 1)^T$ – structural matrix.

Subscript 2 indicates the involvement in the secondary side of the transformer.

Equations of currents are written in differential form:

$$\frac{dI_1}{dt} = \alpha_1 \left(\frac{d\Psi_1}{dt} - \frac{d\psi}{dt} \right) \tag{3}$$

$$\frac{dI_2}{dt} = K\alpha_2 \left(\frac{d\Psi_2}{dt} - \frac{d\psi}{dt} \right) \tag{4}$$

where:

$\psi = (\psi_A, \psi_B, \psi_C)^T$ – column matrix of basic linkages' phases,

$\alpha_1 = \text{diag}(\alpha_{1A}, \alpha_{1B}, \alpha_{1C})$, $\alpha_2 = \text{diag}(\alpha_{2A}, \alpha_{2B}, \alpha_{2C})$ – diagonal matrix of inverse leakage of inductances' winding of the primary and secondary sides of the transformer,

$K = \text{diag}(k_A, k_B, k_C)$ – diagonal matrix of additional logical variables.

It takes the values of 0, 1.

Magnetic circuits are described by the equations:

$$I_1 + I_2 = \Phi(\psi) + \alpha_0 E_1 \psi \tag{5}$$

where:

$\Phi(\psi) = (\varphi_A(\psi_A), (\varphi_B(\psi_B), (\varphi_C(\psi_C))^T$ – column matrix of magnetic voltages in transformer's core. It is defined by its magnetization curves,

φ_0 – inverse leakage's inductance of zero sequence,

E_1 – matrix of dimension 3 × 3, in which all elements are equated to 1.

The equations (5) are differentiated by time (5):

$$\frac{dI_1}{dt} + \frac{dI_2}{dt} = \mathrm{P}\frac{d\psi}{dt} + \alpha_0 E_1 \frac{d\psi}{dt} \tag{6}$$

where:

$\mathrm{P} = \mathrm{diag}(\alpha''_A,\ \alpha''_B,\ \alpha''_C)$ – a diagonal matrix of inverse incremental inductances.

Equations of currents are substituted (3), (4) in the equation of state's magnetic circuits (6):

$$\alpha_1\left(\frac{d\Psi_1}{dt} - \frac{d\psi}{dt}\right) + K\alpha_2\left(\frac{d\Psi_2}{dt} - \frac{d\psi}{dt}\right) = \mathrm{P}\frac{d\psi}{dt} + \alpha_0 E_1 \frac{d\psi}{dt} \tag{7}$$

or:

$$\left(\mathrm{P} + \alpha_0 E_1 + \alpha_1 + K\alpha_2\right)\frac{d\psi}{dt} = \alpha_1 \frac{d\Psi_1}{dt} + K\alpha_2 \frac{d\Psi_2}{dt} \tag{8}$$

whence it appears:

$$\frac{d\psi}{dt} = D_1 \frac{d\Psi_1}{dt} + D_2 \frac{d\Psi_2}{dt} \tag{9}$$

$D_1 = G\alpha_1,\ D_2 = GK\alpha_2,\ G = \left(\mathrm{P} + \alpha_0 E_1 + \alpha_1 + K\alpha_2\right)^{-1}$.

Equations (1), (2) are written as a single expression:

$$\frac{d\Psi}{dt} = U - RI \tag{10}$$

where:

$\Psi = (\Psi_1, \Psi_2)^T$ – column matrix of complete linkages,

$I = (I_1, I_2)^T$ – column matrix of currents,

$U = (U_1, -H_C u_C)^T$ – column matrix of voltages,

$R = \mathrm{diag}(R_1, R_2)$ – diagonal matrix of active resistances.

Taking into consideration (10) the equation (9) can be written as:

$$\frac{d\psi}{dt} = D(U - RI), \quad D = (D_1,\ D_2) \tag{11}$$

The differential equations (9) are substituted in equations of currents (3), (4):

$$\frac{dI_1}{dt} = \alpha_1 \left(\frac{d\Psi_1}{dt} - D_1 \frac{d\Psi_1}{dt} - D_2 \frac{d\Psi_2}{dt} \right) \tag{12}$$

$$\frac{dI_2}{dt} = K\alpha_2 \left(\frac{d\Psi_2}{dt} - D_1 \frac{d\Psi_1}{dt} - D_2 \frac{d\Psi_2}{dt} \right) \tag{13}$$

Taking into consideration the notation (10) the system of equations (12), (13) can be written as:

$$\frac{dI_1}{dt} = A_1(U - RI), \quad A_1 = (\alpha_1(E - D_1), \ -\alpha_1 D_2) \tag{14}$$

$$\frac{dI_2}{dt} = A_2(U - RI), \quad A_2 = (-K\alpha_2 D_1, \ K\alpha_2(E - D_2)) \tag{15}$$

Equation (15) will be added by capacitor equation:

$$\frac{du_C}{dt} = C^{-1} \left(H_C^T I_2 - i_a \right) \tag{16}$$

Equation (11), (15) and (16) can be written as single expression:

$$\frac{dX}{dt} = B(\psi)Z(t), \quad Z(t) = (U - RI, \ H_C^T I_2 - i_a)^T \tag{17}$$

where:

$X = (\psi, I_2, u_C)^T$ – vector of state variables,

$B = \text{diag}(M, C^{-1})$ – matrix of coefficients; $M = (D, A_2)^T$.

Primary currents are determined in accordance with (5):

$$I_1 = \Phi(\psi) + \alpha_0 E_1 \psi - I_2 \tag{18}$$

These equations must be supplemented by the conditions of closing and opening thyristors. If the valve of the j phase ($k_j = 0$) is closed the dependence $\Psi_j = \psi_j$, $j = A, B, C$ will be correct. That's why its opening condition will be determined by attaching positive voltage onto it:

$$-\frac{d\psi_j}{dt} - u_C > 0 \tag{19}$$

If condition (19) is executed, the counting of angle of opening thyristors will begin. For the opening angle of thyristors must be $\pi/6$. The period of the supply voltage is $\tau = 0.02$ c, which corresponds to a frequency of $f = 50$ Hz. This period corresponds to the meaning of 2π rad. Therefore angle $\pi/6$ are formed by 1/12 period, it means $\Delta\tau = 0.01/6$ s. From the moment of executing condition (9) the time $\Delta\tau = 0.01/6$ was counted. The signal of permitting for thyristor's opening T_j is submitted and additional logical variable value $k_j = 1$ are assigned. Thyristors are closed by their currents:

$$i_{2j} = 0, \quad \frac{di_{2j}}{dt} < 0 \tag{20}$$

If condition (20) is executed the valve T_j will be closed and additional logical variable value will be assigned $k_j = 0$.

Now the equation of DC motor with independent excitation is being considered. Differential equations of motor's winding currents take this form:

$$\frac{di_A}{dt} = S_A u_C + T_A u_F + E_A, \quad \frac{di_F}{dt} = T_F u_C + S_F u_F + E_F \tag{21}$$

where:

$$S_A = 1/\left(L_A + L_{AF} L_{FA} / L_{FF}\right), \quad T_A = -S_A L_{AF} / L_{FF}, \quad T_F = -S_A L_{FA} / L_{FF}$$

$$E_F = -\left(L_{FA} E_A + r_F i_F\right) / L_{FF}, \; E_A = S_A \left(L_{AF} r_F i_F / L_{FF} - c\omega\Phi - \Delta u - r_A i_A\right)$$

$$S_F = \left(1 - L_{FA} T_A\right) / L_{FF}$$

L_A – total inductance of consecutive circuit's armature,
L_{FF} – inductance of exciting winding,
L_{AF}, L_{FA} – mutual inductances of circuit's armature and circuit of excitation,
r_A, r_F – active resistances of circuit's armature and circuit of excitation,
ω – angular velocity of motor armature,
Φ – magnetic flux of motor,
c – constitutive constant of motor armature,
Δu – voltage drop in the brush contact.

In compensated motors the account of magnetic saturation can be performed approximately using the magnetization curve $\Phi = \Phi(i_F)$, $L_{FF} = L_{FF}(i_F)$. In the unsaturated motor it can be performed by $\Phi = ki_F$, $L_{FF} = \text{const}$.

The equations of motion can be written by d'Alembert equation:

$$\frac{d\omega}{dt} = \left(c\Phi i_A - M_O\right) / J \tag{22}$$

where:

J – a moment of inertia of the rotor motor,
M_O – a moment of resistance.

The motor is connected with tachogenerator, which can be considered as a linear object. It is described by the equation of the first order:

$$\frac{du_g}{dt} = \left(k_g \omega - u_g\right) / T_g \tag{23}$$

where:

u_g – input voltage of tachogenerator,
k_g – a coefficient of tachogenerator's transmission,
T_g – constant of tachogenerator's.

The equation of mechanic elements can be written in matrix form:

$$\frac{dX_M}{dt} = B_M Z_M(t) \tag{24}$$

where:

$X_M = (i_A, i_F, \omega, u_g)^T$ – a vector of mechanical state variable,
$Z_M(t) = (u_C, u_F, 1, c\Phi i_A - M_O, k_g\omega - u_g)^T$ – a vector of time functions,
B_M – a matrix of coefficient.

$$B_M = \begin{vmatrix} S_A & T_A & E_A & 0 & 0 \\ T_F & S_F & E_F & 0 & 0 \\ 0 & 0 & 0 & J^{-1} & 0 \\ 0 & 0 & 0 & 0 & T_g^{-1} \end{vmatrix} \tag{25}$$

The error of control system is formed as the difference between control voltage u_1 and the output voltage of the generator u_g:

$$u_2 = u_1 - u_g \tag{26}$$

The microcontroller performs the role of the regulator of the system. It reconstructs a linear dependence of the transistor's opening angle on system's error. It played with the linear dependence of the angle of opening on thyristors system error:

$$\varphi_O = \begin{cases} \varphi_m(1 - u_2 / u_m), & if \quad u_2 \le u_m \\ 0, & if \quad u_2 > u_m \end{cases} \tag{27}$$

where:

u_m – voltage at which the opening angle of thyristors becomes equal to zero,
φ_m – the meaning of opening angle of thyristors, when the error of system is equal to 0.

Thus, the microcontroller must:

a) fix the moment of time t_1 of condition's (19) executor,
b) according to (26) determine the opening angle of thyristors,
c) transfer this angle in time $t_O = \varphi_O\tau/(2\pi)$,
d) when the passage time will exceed the moment t_1 on value t_O the signal of opening the thyristor will be given.

The last point is an identity of the condition $t \ge t_1 + t_O$. Mathematically, it boils down to assigning additional logical variable value the meaning of $k_j = 1$. In the program code, this delay can be realized by usual transfer time t_O in the number of discrete steps, namely:

$$n = t_O/\Delta t$$

where:

Δt – an integration step.

4. Results of computer simulations

The Program was developed and computer simulation of the dynamic processes of controlling the speed of rotation of DC motor with separate excitation was carried out. Here the managing element is a three-phase half-wave rectifier which is built on three thyristors. Voltage feed is set by expressions:

$$u_{1A} = U_m \sin\left(\omega_0 t + \gamma\right),\ u_{1B} = U_m \sin\left(\omega_0 t + \gamma - \gamma_0\right),\ u_{1C} = U_m \sin\left(\omega_0 t + \gamma + \gamma_0\right)$$

$$U_m = 311 \text{ V}, \omega_0 = 314.1593 \text{ rad/s}, \gamma = 1.8 \text{ rad}, \gamma_0 = 2\pi/3 \text{ rad}.$$

Using these parameters of rectifier calculations was executed: $r_{1A} = r_{1B} = r_{1C} = 2.0$ Ohm; $r_{2A} = r_{2B} = r_{2C} = 1.0$ Ohm; $\alpha_{1A} = \alpha_{1B} = \alpha_{1C} = 172$ H^{-1}; $\alpha_{2A} = \alpha_{2B} = \alpha_{2C} = 200$ H^{-1}; $\alpha_0 = 1.2$ H^{-1}; $C = 9.0$ mF. Magnetized curve was approximated by the equation of formula's choosing:

$$\varphi(\psi) = \begin{cases} a_1\psi, & |\psi| > \psi_1 \\ S_3(\psi), & \psi_1 \le |\psi| \le \psi_2 \\ a_2\psi - a_0, & |\psi| > \psi_2 \end{cases} \tag{28}$$

where:

$a_1 = 0.25$ H^{-1};
$a_2 = 3.0$ H^{-1};
$a_0 = 1.8$ A;
$\psi_1 = 0.2$ Wb;
$\psi_2 = 0.9$ Wb;
$\varphi(\psi_1) = 0.05$ A;
$\varphi(\psi_2) = 0.9$ A;
$S_3(\psi)$ – a cube spline;
$\alpha''(\psi_1) = a_1$, $\alpha''(\psi_2) = a_2$ must be taken into account.

The parameters of motor and generator are $L_A = 9.67$ H; $L_{FF} = 110.8$ H; $L_{AF} = L_{FA}$ 0 H; $r_A = 33.2$ Ohm; $r_F = 173$ Ohm; $J = 0.09$ N·m·s^2/rad; $M_O = 4.0$ N·m; $c = 70.8$ N·m/(Wb·A); $u_f = 300.0$ V; $T_g = 0.04$ s; $k_g = 0.1$ V·s/rad; $u_m = 10.0$ V; $\varphi_m = 5.934$ rad.

In Fig. 3 the time dependence of the rate's rotation of the DC motor for different values of the input voltage of the system u_1 is shown. Thus, the curve 1 corresponds to the meaning 5 V, the curve 2 corresponds to the meaning 10 V, curve 3 corresponds to the meaning 15 V, curve 4 corresponds to the meaning 20 V. We see that the time of transition process of control system increases with input voltage. When the voltage is maximum (curve 4) the transition process will have exponential character without control.

The dependence of constant meaning of DC motors winding speed on input voltage u_1 is shown in Fig. 4. This dependence has practically a lineal character. When $u_1 \le 3.5$ V the motor does not wind because his electromagnetic moment is less than the moment of resistance.

Constant meaning of transformer's secondary phase current for two meanings of input voltage u_1. is shown in Fig. 5. For meaning $u_1 = 10$ V (Fig. 5a) the opening angle of thyristors was approximately equal to 3 rad. Instead of it, when u_1 was equal to 20 V this angle would become equal to 0.7 rad. Looking at Fig. 4, we can simply say that after 20 V the speed of motor winding remains constant. Comparing Fig. 5b with Fig. 5a, the form of current phase is narrower and has greater amplitude in Fig. 5a.

Constant meaning of capacitor for two meanings of input voltage is shown on Fig. 6. For meaning u_1 = 10 V (Fig. 6a) amplitude of capacitor voltage is equal to 64 V. This voltage has constant and variable component. The constant component dominates and it is approximately equal to 163,5 V. Amplitude of variable component is not overstated 0,5 V. For meaning u_1 = 20 V (Fig. 6b) the constant component of capacitor's voltage approximately grew up to 101 V, but variable component practically did not change.

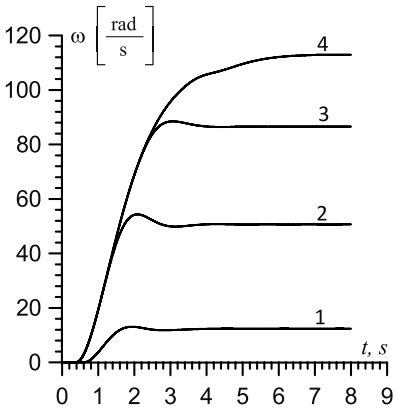


Fig. 3. Time dependence of the speed's rotation of the motor for different values of the input voltage u_1: 1–5 V, 2–10 V, 3–15 V, 4–20 V

Rys. 3. Zależności czasowe prędkości obrotowej silnika dla różnych wartości napięcia wejściowego u_1: 1–5 V, 2–10 V, 3–15 V, 4–20 V

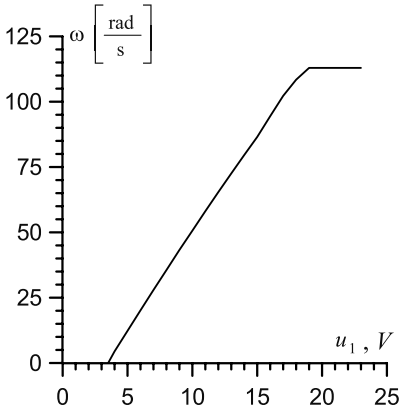


Fig. 4. The dependence of speed's rotation of the motor input voltage

Rys. 4. 4. Zależność prędkości obrotowej silnika od napięcia wejściowego

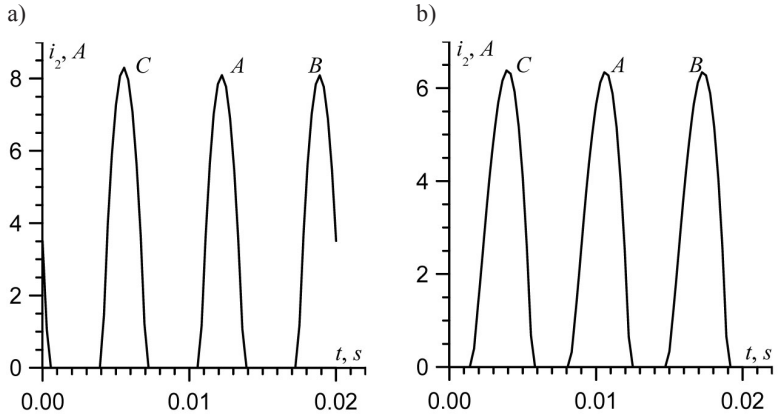


Fig. 5. Well-established values of currents phase transformer secondary for two values of the input voltage u_1: a) 10 V, b) 20 V

Rys. 5. Przebiegi ustalone prądów faz na wyjściu transformatora dla dwu wartości napięcia wejściowego a) 10 V, b) 20 V

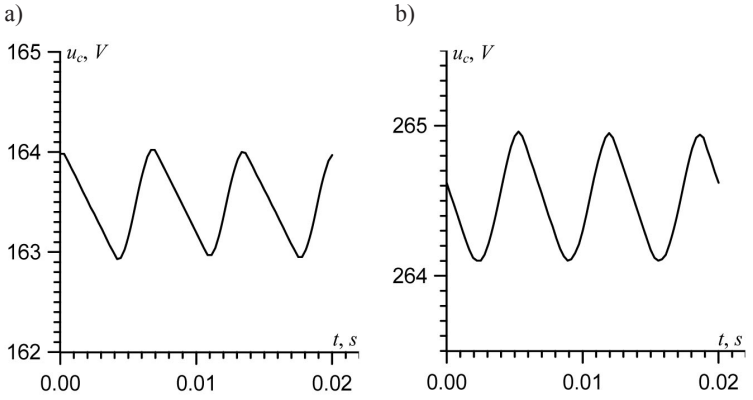


Fig. 6. Well-established values of voltage capacitor for two values of the input voltage u_1: a) 10 V, b) 20 V

Rys. 6. Przebiegi ustalone napięcia kondensatora dla dwu wartości napięcia wejściowego u_1: a) 10 V, b) 20 V

5. Conclusion

This article proposed a mathematical model of thyristor control system of winding speed of DC motor's control with independent excitation. Tyristors' work is shown by scheme of an ideal key, which gives an opportunity to minimize number of system's differential equation. The influence of tyristor's opening angle on winding motor's speed was investigated. Microcontroller controls tyristors. Using linear law, it counts the opening angle which depends on system's error meaning. It was clarified that the speed of armature motor depends on input voltage in the strip from 4 V to 18 V. This dependence has a linear character. When system's input voltage grows up, the time of transitional process will grow up also. Thus, the developed mathematical model of the system's work regimes is distinguished by universality and essence of algorithmization.

References

[1] Yildiz A.B., *Electrical equivalent circuit based modeling and analysis of direct current motors*, Electrical Power and Energy Systems, vol. 43, 2012, 1043-1047.

[2] Betechuoh B.L., Marwala T., Tettey T., *Using inverse neural networks for HIV adaptive control*, International Journal of Computational Intelligence Research, vol. 3, 2007, 11-15.

[3] Chow M.Y., Tipsuwan Y., *Gain adaptation of networked DC motor controllers based on QOS variations*, IEEE Transactions on Industrial Electronics, vol. 50, 2003, 936-943.

[4] Craig J.J., Hsu P., Sastry S.S., *Adaptive control of mechanical manipulators*, The International Journal of Robotics Research, vol. 6, 1987, 16-28.

[5] Crapo A.D., Lloyd J.D. *Homopolar DC motor and trapped flux brushless DC motor using high temperature superconductor materials*, Magnetics, IEEE Transactions on Magnetics, vol. 27, 1991, 2244-2247.

[6] Delibasi A., Turker T., Canserver G., *Real time DC position control by fuzzy logic and PID controllers using labview*, International IEEE Conference on Mechatronics & Robotics 04, Aachen, 13–15 September 2004, 9-15.
[7] Gurbuz F., Apkinar E., *Stability analysis of a closed loop control for a pulse width modulated DC motor drive*, Turkish Journal of Electrical Engineering & Computer Sciences, vol. 10, 2002, 427-438.
[8] Kaczorek T., *Teoria sterowania i systemów*, PWN, Warsaw 1999.
[9] Lankarany M., Rezazade A., *Parameter Estimation Optimization Based on Genetic Algorithm Applied to DC Motor*, IEEE International Conference on Electrical Engineering, Lahore, Pakista, 11–12 April 2007, 1-6.
[10] Porębski J., Korohoda P., *SPICE; program analizy nieliniowej układów elektronicznych*, WNT, Warsaw 1993.
[11] Prigozy S., *PSPICE computer modeling of hysteresis effects*, IEEE Transactions on Education, vol. 36, 1993, 2-5.
[12] Ramotowski M. *Zastosowanie programu NAP do obliczania układów elektrycznych*, WNT, Warsaw 1992.
[13] Saab S.S., Kaed-Bey R.A., *Parameter Identification of a DC Motor: An Experimental Approach*, IEEE Transactions on Electronic, Circuit and Systems, vol. 2, 2001, 981-984.
[14] Samborska M., *Aproksymacja rozwiązań okresowych nieliniowych rownań obwodow elektrycznych metodą Poincarego-Lindstedta*, Czasopismo Techniczne, vol. 1-AC/2012, Wydawnictwo Politechniki Krakowskiej, 67-77.
[15] Samotyy V., *Analiza procesów okresowych nieliniowych urządzeń elektromagnetycznych systemów automatyki*, Czasopismo Techniczne, vol. 1-E/2010, Wydawnictwo Politechniki Krakowskiej, 67-77.
[16] Samotyj V., Shakya S., *Mathematical Model of Frequency Multiplier «Symmetric Scheme of Rectification with Voltage Doubling – Bridge Parallel Thyristor Chopper»*, Technology, vol. 9, 2000, 58-62.
[17] Sevinç A., *A full adaptive observer for DC servo motors*, Turkish Journal of Electrical Engineering & Computer Sciences, vol. 11, 2003, 427-438.
[18] Samotyj V., *Matematyczne modeluwannia stacionarnych procesiw elektromagnetnych pryctrojiw system keruwannia*, Feniks, Lviv 1997.

TECHNICAL TRANSACTIONS | CZASOPISMO TECHNICZNE
AUTOMATIC CONTROL | AUTOMATYKA
1-AC/2013

DAMIAN KRUSZEWSKI*

NONPARAMETRIC MODELING OF MEDICAL SCHEME DATA

MODELOWANIE NIEPARAMETRYCZNE DANYCH MEDYCZNYCH

Abstract

The goal of this paper is to apply Generalized Additive Models to medical scheme data. The flexibility of the nonparametric approach is demonstrated based on a real-life empirical example that seeks to model hypertension and the interplay of determinants, such as physiological measurements, medical attributes, demographic and socioeconomic characteristics in predicting blood pressure. The assessment of nonlinear patterns in the response-predictor relationship and the strength of this association are investigated. The extended Generalized Additive Models allow for modeling not only location and scale, but also other distribution parameters, such as kurtosis and skewness.

Keywords: Generalized Additive Models, smoothing, hypertension, systolic/diastolic blood pressure

Streszczenie

Celem niniejszego artykułu jest aplikacja uogólnionych modeli addytywnych do danych medycznych. Elastyczność nieparametrycznych rozwiązań przedstawiono na przykładzie modelowania zmiennych determinujących poziom nadciśnienia tętniczego krwi, takich jak atrybuty zdrowotne, fizjologiczne, demograficzne czy charakterystyki społeczno-ekonomiczne. W artykule zbadano nieliniowe zależności (oraz ich siłę) pomiędzy zmiennymi objaśniającymi a nadciśnieniem tętniczym krwi. Rozszerzona wersja modelu pozwala wyznaczyć nie tylko parametry skali i położenia, lecz również inne parametry charakterystyczne rozkładu, takie jak kurtoza i skośność.

Słowa kluczowe: Uogólnione Modele Addytywne, wygładzanie, nadciśnienie tętnicze krwi, ciśnienie skurczone/rozkurczowe krwi

* Damian Kruszewski, M.Sc., PAREXEL International; Ph.D. Studies, Systems Research Institute, Polish Academy of Sciences.

1. Introduction

There are loads of methods and techniques for nonparametric/semiparametric regression, such as Locally Weighted Regression [2], Regression Splines, Smoothing Splines, B-Splines [16], P-Splines [4], etc. All of them are aimed at one problem – to make a precise prediction. Compared with standard parametric methods such as Linear/Binary/Logistic Regression Models or Generalized Linear Models (GLM), the methodology behind nonparametric modeling relaxes the assumption of linearity in the response-predictor relationship. It enables to uncover structural behavior of the response with the independent variables that may otherwise be missed. The notion of exploring data nonparametrically has been proven to be successful in the statistical modeling. Unfortunately, this success sometimes is accompanied with a weak interpretability and greater variance for greater dimensionality. Proposed by Hastie and Tibshirani [9], Generalized Additive Models (GAM) allow for multidimensional data and provide the ability to detect the nonlinear associations without any damaging repercussion on interpretability.

The aim of this paper is to use Generalized Additive Models (GAM) to predict hypertension by multiple independent variables whose effect is modeled nonparametrically. The presented results demonstrate the importance of nonparametric solutions, especially in the context of a real-life data set. The study on the example of hypertension shows that Generalized Additive Models (GAM) provide flexible statistical methods for identification of nonlinear regression effects and complex shapes in the relationship between the response and the predictors which are missed by standard parametric solutions. The models built for Systolic/Diastolic Blood Pressure account for scale, location, kurtosis and skewness of the continuous response distribution.

There is a plethora of studies trying to find and explain the factors influencing blood pressure. Unfortunately, most of them follow parametric assumptions [18], other allow for semiparametric inferences but strictly restricted to pre-specified response distribution belonging to the exponential family and thus, disregarding kurtotic or skewed distributions [8]. As presented on the example of Systolic/Diastolic Blood Pressure, the analyzed response variables do not follow exponential family features.

2. Rationale

Modeling hypertension with Generalized Additive Models (GAM) has by its nature interdisciplinary scope. Broadening the spectrum of medical applications of Generalized Additive Models (GAM) contributes to both IT and medicine. Hypertension is a chronic health condition prevalent in most developed nations. Its prevalence in the western population exceeds 20%. Untreated high blood pressure is a major risk factor for coronary heart disease, cardiovascular disease, stroke or diabetes. Thus, it is of crucial importance to develop models identifying potential markers for its prediction. Better knowledge of the blood pressure drivers supports the decision making process concerning hypertension diagnosis and its treatment.

The rationale behind incorporating Generalized Additive Models (GAM) are:

a) relaxing the assumptions of parametric models: 1) linear form of the relationship between response and predictors, 2) diagnostic checking of the residuals (normality and independence),

b) enabling the use of multidimensional data,
c) reasonably easy interpretation.

Generalized Additive Models (GAM) have greater flexibility than their parametric counterparts. Traditional methods, although attractively simple, often fail in many applied settings. In real-life, effects are generally not linear.

The rationale of relaxing the parametric assumptions of linearity and normality (a) enables to explore the data visually and uncovers structural behavior that may be otherwise missed. Of note is the fact, that these properties of Generalized Additive Models (GAM) are shared by other nonparametric solutions. What make them distinctive from other nonparametric models are their properties of multidimensionality and interpretability (b and c). In light of the necessity of including large number of explanatory variables in the real-life applications, the practical capabilities of the most commonly used nonparametric regression methods, such as Thin-Plate Smoothing or Local Regression Methods [2] are significantly restricted. In such circumstances, the sparseness of data results in the unacceptably large variance of estimates ("the curse of dimensionality").

The drawbacks of both parametric (linearity) and standard nonparametric solutions (multidimensionality and interpretability) are overcome by Generalized Additive Models (GAM). Their methodology allows for estimating the additive terms individually using a univariate smoother – each input is considered independently. This addresses the issue associated with "the curse of dimensionality". Additionally, individual term's estimate directly explains the relative contribution to the response changes and thus, Generalized Additive Models (GAM) are among the most interpretable statistical models.

3. Description of solutions

Generalized Additive Models (GAM) were first proposed by Hastie and Tibshirani [9]. Their fit allows to combine:

– the flexibilities of Generalized Linear Models (GLM) – an arbitrary function of dependent variable,
– the additive assumptions that enable to explore the data nonparametrically.

Generalized Additive Models (GAM) extend Linear Models (LM) and Generalized Linear Models (GLM) to include smooth functions of explanatory variables. They are an important step forward in the generalization of Generalized Linear Models (GLM). Generalized Additive Models (GAM) do not require any transformations of the predictors to improve the fit. The different regression models might be envisioned as being nested within each other, with linear regression being the most limiting case, and Generalized Additive Models (GAM) the most general. They combine the abilities to explore the data nonparametrically simultaneously with the distributional flexibilities of Generalized Linear Models (GLM). Instead of having a single estimation coefficient for each of the predictors, Generalized Additive Models (GAM) use an arbitrary nonparametric function to approximate the association between each of the predictors and the response.

The only underlying assumption made is that the nonparametric functions are additive and that the components are smooth. Generalized Additive Models (GAM), like Generalized Linear Models (GLM), apply a monotonic link function to establish a relationship (link) between the mean of the response variable and a "smoothed" function of the explanatory

variables. They are constructed through summing up all the functions which fit the data locally. The final model closely represents the behavior of the data (data driven approach). However, apart from the nature of the response-predictor relationship, the probability distribution of the response must still be specified. In this sense, Generalized Additive Models (GAM) are more aptly referred to as semiparametric models.

3.1. Fundamentals of Generalized Additive Models (GAM)

Generalized Additive Models (GAM) combine Generalized Linear Models (GLM) and Additive Models (AM):

Generalized Linear Models (GLM) extend the response distribution of the linear model into the exponential family. Providing that Y is a response random variable and $X_j, ..., X_p$ are explanatory variables, a standard linear regression model might be expressed as $E(Y) = \beta_0 + \sum_{j=1}^{p} \beta_j X_j$. Assuming that $E(Y) = \mu$ and $\eta = g(\mu)$ where $g(\bullet)$ is a smooth monotonic differentiable (up to third order) link function, the response-predictor relationship in Generalized Linear Models (GLM) is defined by $\eta = \beta_0 + \sum_{j=1}^{p} \beta_j X_j$. The link function $g(.)$ describes how the expected value of Y is related to linear predictor $E(Y) = \mu$. Because the link function is a monotonic and invertible function, the mean can be expressed as the inversely linked linear predictor: $E(Y) \equiv \mu = g^{-1}(\eta)$ where $g^{-1}(.)$ is so called inverse link function. The form $\eta = g(\mu)$ emphasizes that Generalized Linear Models (GLM) use transformations of the mean (no transformation of the data). The second form, i.e. $\mu = g^{-1}(\eta)$ shows how predictions of the mean can be obtained following the estimation of η. The most commonly employed link functions are Normal, Exponential, Gamma, Inverse Gamma, Poisson and Binomial. For instance, for binary data, a common link function is the logit link: $g(t) = \log[t/(1 - t)]$. The mean function of Generalized Linear Model (GLM) with the assumed logit link function and one predictor can be written as: $\log\left\{\frac{\mu}{1-\mu}\right\} = \beta_0 + \beta_1 x$ and thus, $\mu = \frac{1}{1+\exp\{-\beta_0 - \beta_1 x\}}$. This is known as a logistic regression model. The response variable in Generalized Linear Models (GLM) is assumed to be a member of exponential family.

– Additive Models (AM) extend the parametric form of predictors in the linear model to nonparametric forms. Additive Model (AM) is defined as: $E(Y) = s_0 + \sum_{j=1}^{p} s_j(X_j)$ where the smoothers s_i, $i = 1, ..., p$ are smoothing splines. Please note that smooth functions have to be constrained to have zero mean.

Combining Generalized Linear Models (GLM) and nonparametric Additive Models (AM), Generalized Additive Models (GAM) might be defined as:

$$\eta = s_0 + \sum_{j=1}^{p} s_j(X_j). \tag{1}$$

where the response variable has a probability density from the exponential family. Generalized Additive Models (GAM) extend Generalized Linear Models (GLM) by replacing the form $\beta_0 + \sum_{j=1}^{p} \beta_j X_j$ with the additive form $s_0 + \sum_{j=1}^{p} s_j(X_j)$.

The form and the nature of Generalized Additive Models (GAM) are dependent on the backfitting algorithm, the local scoring method, the specified smoothing parameters and the degrees of freedom (DF) used for their computation. All of these parameters are thoroughly discussed in the literature [19]. In this paper, only a brief summary is provided:

- **Backfitting algorithm:** The backfitting and the local scoring form an interactive method to estimate the smoothers s_i, $i = 1, ..., p$. The backfitting algorithm is an algorithm that enables to fit Additive Models (AM). It might be used with different smoothers such as univariate or bivariate splines. The iterative mechanism permits to estimate each of the smoothing functions $s_k(\ )$, given estimates $\{\hat{s}_j(\ ),\ j \neq k\}$ [6].
- **Selection of smoothing parameters:** Very different types of smoothing functions could be specified in Generalized Additive Models (GAM): Cubic Smoothing Spline, Local Regression, Thin-Plate Smoothing Spline, etc. A smoother is an operator for summarizing the trend and the variability of a response measurement Y as a nonparametric function of explanatory measurements $X_1, ..., X_p$. Smoothing methodology offers a way by which nonlinear and nonparametric relationships can be handled without the restrictions of parametric models. In Generalized Additive Models (GAM), each smoother has a single unique smoothing parameter. The most commonly used methods for the selection of smoothing parameters are Cross Validation (CV) function and Generalized Cross Validation (GCV) technique. For more details, please refer to Wahba [19].

4. Empirical results

4.1. Description of data

The data set used in this paper is obtained from the National Health & Nutrition Examination Survey (NHANES). NHANES is an ongoing program designed to assess the health status of patients in the United States. The NHANES collects, among others, demographic, health history and behavioral information. This paper uses blood pressure measurements and demographic characteristics data. Blood pressure measurements were assessed during physician examinations (taken in the mobile examination centers), whereas demographic characteristics were collected during personal interviews. For the analysis purposes, the data from 2003 to 2010 is pooled. Calculations are performed using SAS Base 9.2 and R[1].

[1] SAS = Statistical Analysis System (system software provided by SAS Institute Inc., 4GL language), R = programming software and language for statistical computing developed by Development Core Team (Robert Gentleman and Ross Ihaka).

The primary objective of this real-life empirical example is to investigate:

1) the usefulness and flexibility of Generalized Additive Models (GAM) for medical scheme data,
2) the dependence of hypertension on various medical factors,
3) the patterns of hypertension and the effects of the independent variables on the response,
4) the strength of the association between independent and dependent variables.

The response measurement is either continuous derived Mean Systolic/Diastolic Blood Pressure (referred to as Mean SBP/Mean DBP) or derived binary Hypertension/Borderline Hypertension level (Table 1). Mean SBP/Mean DBP is an average of three blood pressure readings taken during physician examinations. The binary Hypertension/Borderline Hypertension response is derived based on Mean SBP/Mean DBP and takes the value of 'Yes' (Hypertension/Borderline Hypertension) if Mean SBP is greater/equal 120 mmHg or Mean DBP greater/equal 80 mmHg and the value of 'No' (No Hypertension/Borderline Hypertension) otherwise. Its derivation is intended to account for both Mean SBP and Mean DBP. The histogram with an overlaid univariate kernel density estimate for continuous response variables is presented in Fig. 1. One-Way Frequencies for their binary counterparts are shown in Table 2.

Table 1

Response variables used for fitting Generalized Additive Models (GAM)

Variable Name	Variable Explanation
HYPERTFL	Hypertension/Borderline Hypertension (1=Yes, 0=No) [Derived]
mSBP	Mean Systolic Blood Pressure (mm Hg) [Derived]
mDBP	Mean Diastolic Blood Pressure (mm Hg) [Derived]

Table 2

One-Way Frequencies for binary Hypertension/Borderline Hypertension response

Hypertension/Borderline Hypertension (HYPERTFL)	Frequency	Percent	Cumulative Frequency	Cumulative Percent
[1=Yes]	9905	44.61	9905	44.61
[0=No]	12299	55.39	22204	100.00

Table 3

Moments for Mean SBP and Mean DBP

Moments	Mean SBP	Mean DBP	Moments	Mean SBP	Mean DBP
N	22204	22204	**Sum Weights**	22204	22204
Mean	119.869	67.93787	**Sum Observations**	2661592	1508492
Std Deviation	17.521	11.81245	**Variance**	306.997	139.533
Skewness	1.136	0.04760	**Kurtosis**	**1.873**	**0.160**
Coeff Variation	14.617	17.38713	**Std Error Mean**	0.117	0.079

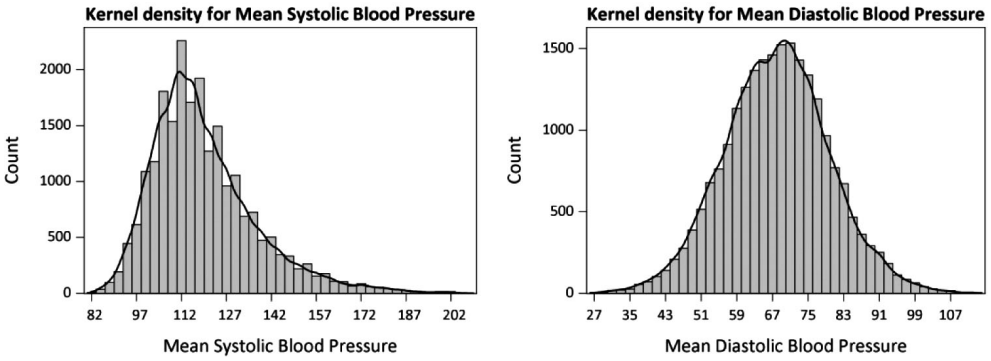


Fig. 1. Histogram for Mean SBP/Mean DBP

Rys. 1. Histogram dla Średniego Ciśnienia Skurczowego/Rozkurczowego Krwi

The predictor measurements are mainly variables representing physiological measurements, medical attributes as well as demographic and socioeconomic characteristics. Ratio of Income to Poverty compares a family's income with their appropriate poverty threshold[2]. The explanatory variables are listed in Table 4.

Table 4

Explanatory variables used for fitting Generalized Additive Models (GAM)

Variable Name	**Variable Explanation**
AGEYRS	Age at Screening (years)
GENDER	Gender (1=Male, 2=Female)
BMXBMI	Body Mass Index (kg/m²)
LBXSUA	Uric acid (mg/dL)
LBDHDDSI	HDL-cholesterol (mmol/L)
LBXSGTSI	Gamma Glutamyl Transferase (GGT) (U/L)
INDFMPIR	Family PIR (Ratio of Family Income to Poverty)
LBDSGLSI	Glucose (mmol/L)
LBDSCRSI	Creatinine (umol/L)

Model construction was preceded by outlier's detection. It involved removing extreme or missing values that might unduly influence the results of the analysis and potentially lead to incorrect conclusions. Extreme values were defined as values deviating from the expected range of 1st percentile and 99th percentile. Additionally, it is important to mention that in order to prepare the data for the analysis, it is recommended to apply the reduction of the high-dimensionality of the data set [12]. Reducing the number of variables under consideration

[2] Ratio of '1' means living right at the poverty line (income at 100% of poverty level), ratio above '1' indicates living above the official definition of poverty (i.e. a ratio of '1.5' means that income is 150% above the poverty threshold).

mitigates the effects of the curse of dimensionality and contributes to more accurate data analysis results [11].

Based on raw data and before building Generalized Additive Models (GAM), a positive association with Mean SBP/Mean DBP for Age and Body Mass Index (BMI) is noticed.

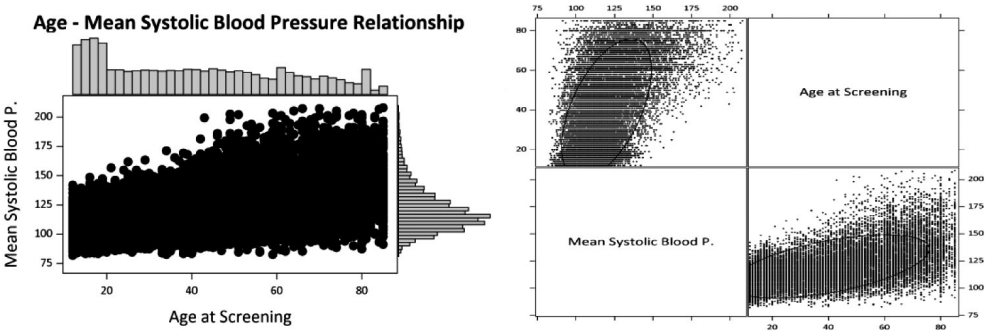


Fig. 2. Marginal Scatter Plot for Age – Mean SBP Relationship

Rys. 2. Brzegowy Wykres Rozrzutu dla Relacji Wiek – Średnie Ciśnienie Skurczowe Krwi

Knowing the positive response-predictor relationship for both explanatory variables, Age at Screening is classified into 4 Age Cohorts: 12–<23 years, 23–<40 years, 40–<57 years and 57–<85 years, and Body Mass Index (BMI) into 3 BMI Groups: 14–<25 kg/m², 25–<30 kg/m², 30–<45 kg/m². Please note that Body Mass Index (BMI) in the range of 25–<30 kg/m² could be an indicator of being overweight and Body Mass Index (BMI) > 30 kg/m² an indicator of being obese. Age Cohort, BMI Group and Gender are treated as the classification variables and constitute the parametric part of Generalized Additive Models (GAM).

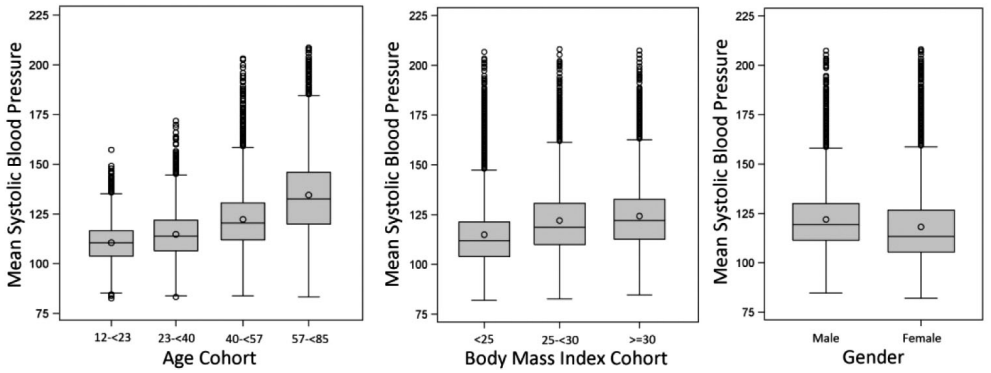


Fig. 3. Box plots of Mean SBP for classification variables

Rys. 3. Wykresy pudełkowe Średniego Ciśnienia Skurczowego dla poszczególnych zmiennych klasyfikujących

Alike the classification variables, the continuous predictors are intended to account for the nonparametric inferences. To examine the trends between the explanatory variables, all of them are put on one panel (Fig. 4). Only a weak correlation exists between explanatory variables. Thus, the impact of multicollinearity (concurvity) on parameter estimates is not a major issue. Appendix 1 presents bivariate kernel density estimates for selected explanatory variables, with contour and surface plots, in which density function is averaged across the observed data points to create a smooth approximation.

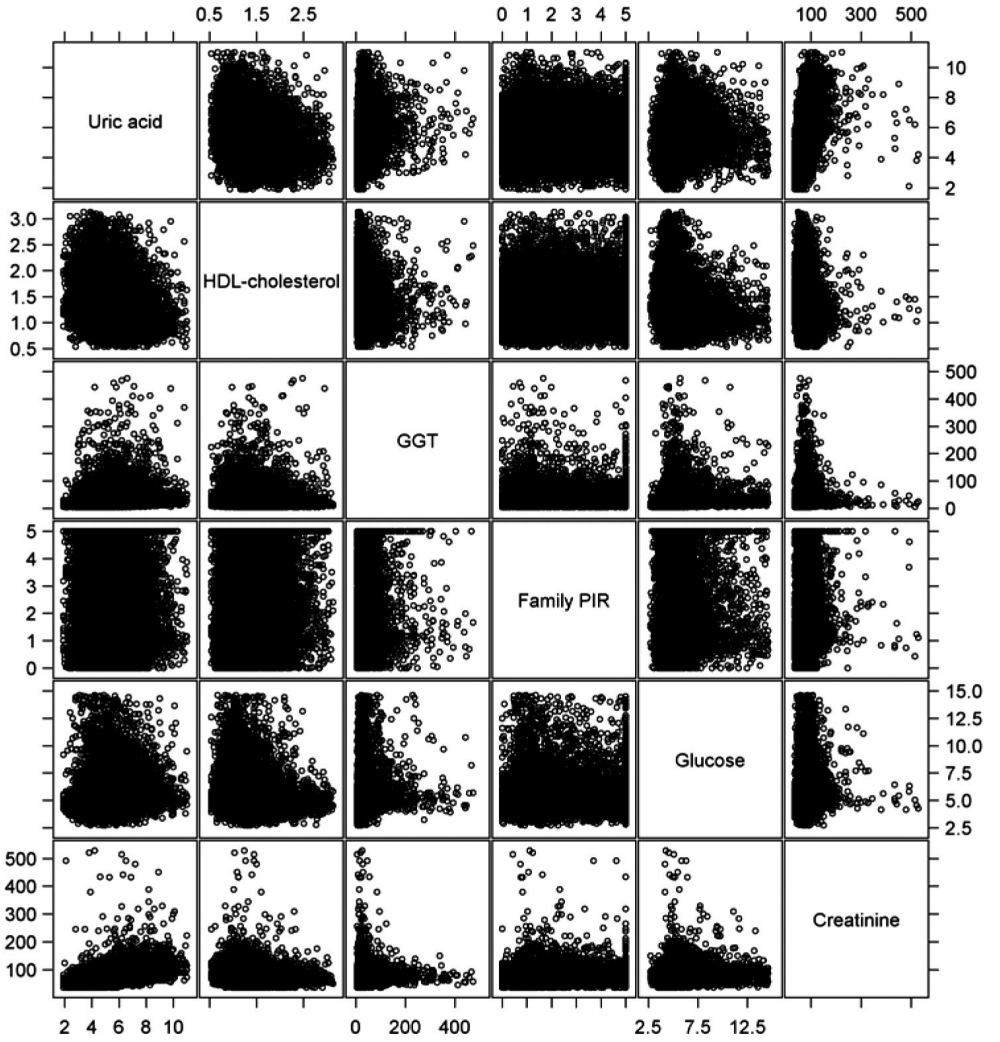


Fig. 4. Scatter Plot Matrices for Continuous Explanatory Variables

Rys. 4. Wykres Rozrzutu dla Ciągłych Zmiennych Objaśniających

4.2. Modeling binary response variable

As assessed by the scatter plot of explanatory variables (Fig. 4), there are striking features in the analyzed data set. It is really hard to determine from the plot whether the relationship between explanatory variables and the response is linear or not. Another issue has to do with the response distribution. Although histograms for continuous response variables, particularly for Mean DBP, seem to look pretty symmetric (Fig. 1), they are not of normal distribution and more importantly, do not follow exponential family features (as verified by Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling Tests).

To account for both Mean DBP and Mean SBP, as the starting point, Generalized Additive Model (GAM) for binary Hypertension/Borderline Hypertension is built (Table 5). It depends on additive predictors through a 'Logit' link function. An additive model with univariate cubic smoothing splines is requested for all continuous explanatory predictors (Uric acid, HDL-cholesterol, Gamma Glutamyl Transferase (GGT), Family PIR, Glucose and Creatinine).

Parameter Estimates for the parametric (linear) part of the model (Table 5) indicate high significance of the linear trends for each of the explanatory variables, with p-values much lower than the assumed significance level of 0.05. The Analysis of Deviance (Table 6) presents a χ^2 test comparing the deviance between the fully specified model (all explanatory variables with parametric and nonparametric part) and the model without the nonparametric component of a given variable (omitting nonlinearity). The nonparametric effects are concluded to be highly significant for each univariate smoothing splines introduced into the model.

Table 5

Parameter Estimates (Binary Regression, Link Function = Logit)

Parameter	Par. Estimate	Standard Error	t Value	Pr > |t|
Intercept	–3.28	0.132	–24.9	< 0.0001**
Age Cohort (23–<40 vs 12–<23 years)	0.60	0.048	12.4	< 0.0001**
Age Cohort (40–<57 vs 12–<23 years)	1.42	0.050	28.5	< 0.0001**
Age Cohort (57–<85 vs 12–<23 years)	2.26	0.052	43.5	< 0.0001**
Gender (Male vs Female)	0.35	0.040	8.9	< 0.0001**
BMI Group (25–<30 vs 14–<25 kg/m²)	0.25	0.040	6.1	< 0.0001**
BMI Group (30–<45 vs 14–<25 kg/m²)	0.57	0.044	12.9	< 0.0001**
LINEAR(Uric acid)	0.11	0.015	7.6	< 0.0001**
LINEAR(HDL-cholesterol)	0.20	0.045	4.5	< 0.0001**
LINEAR(Gamma Glutamyl Transferase)	< 0.01	< 0.001	9.8	< 0.0001**
LINEAR(Family PIR)	–0.03	0.001	–2.7	0.0066*
LINEAR(Glucose)	0.06	0.014	4.6	< 0.0001**
LINEAR(Creatinine)	< 0.01	0.001	4.1	< 0.0001**

Table 6

Smoothing Model Analysis – Analysis of Deviance

Cubic Spline (CS)	Generalized Cross Valid.	Sum of Squares	Chi-Square	Pr > ChiSq
CS(Uric acid)	0.538	19.550	19.550	0.0002**
CS(HDL-cholesterol)	0.394	11.222	11.222	0.0106*
CS(Gamma Glutamyl Transferase)	5.346	113.702	113.702	< 0.0001**
CS(Family PIR)	0.625	13.693	13.692	0.0034**
CS(Glucose)	3.085	27.985	27.985	< 0.0001**
CS(Creatinine)	4.717	26.990	26.990	< 0.0001**

To allow the visual judgment of the relative nonparametric effect sizes, a curvewise Bayesian confidence interval (standard-error band) to each smoothing component is used [9]. Smoothing Components Plot (Fig. 5) demonstrates the estimated smoothing spline functions with the linear effect subtracted out. It gives an idea where significant

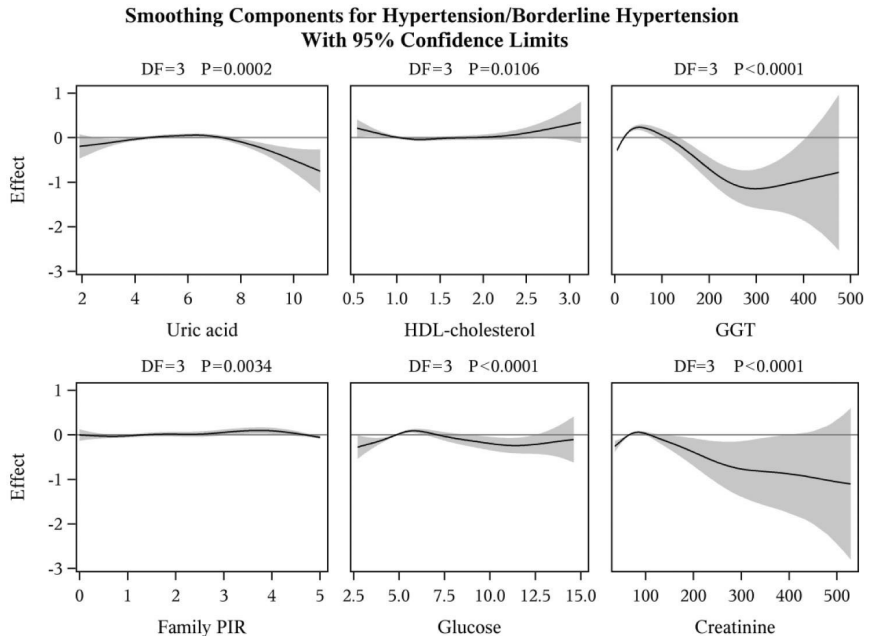


Fig. 5. Smoothing Components Plot for Hypertension/Borderline Hypertension

Rys. 5. Wykres Komponentów Wygładzania dla Nadciśnienia/Podwyższonego Ciśnienia Krwi

nonlinearities occur. The small p-values indicate that the data exhibits significant nonlinear structure. All variables are nonlinear predictors of Hypertension/Borderline Hypertension. They have a pretty pronounced complicated structure with a quadratic pattern for Uric acid and HDL-cholesterol, and even more curved pattern for Gamma Glutamyl Transferase (GGT) or Creatinine. It highlights the ability of Generalized Additive Models (GAM) to uncover nonlinear relationships and their potential in identifying patterns which are missed by standard parametric approaches.

The estimate of Generalized Additive Model (GAM) is just the sum of individual predictors' estimates plus a constant. Fig. 6 shows the partial prediction and the entire prediction effects of individual predictors (derived as the sum of the estimated linear terms – parametric part and the respective nonlinear partial predictions – nonparametric part). It further reveals the nature of the data and the overall shape of the relationship between the predictors and the response variable.

Concluding, the estimated Generalized Additive Model (GAM) for binary Hypertension/ Borderline Hypertension response reveals pronouncedly complex nonlinear patterns in the response-predictor relationships for all explanatory variables. For more in-depth investigation, the estimation of continuous Mean SBP/Mean DBP is needed. However, as verified by Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling Tests, both Mean SBP and Mean DBP do not belong to the class of exponential family distributions. Excess kurtosis is negative for both variables (Table 3) indicating platykurtic distributions. Thus, the estimation of continuous response variables requires going beyond exponential family distributions. This is accomplished by extending the standard Generalized Additive Models (GAM) to Generalized Additive Models (GAM) for Location Scale and Shape.

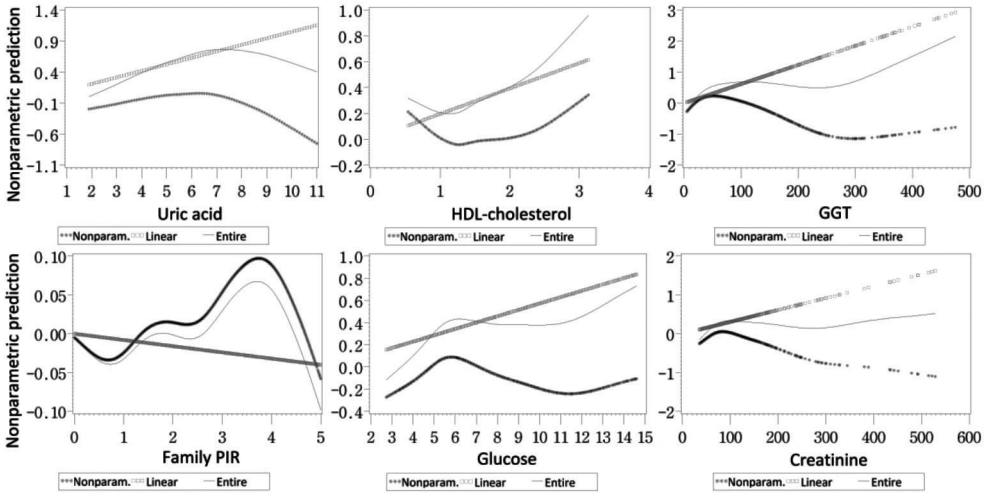


Fig. 6. Partial predictions for explanatory variables – predictors with and without linear terms

Rys. 6. Prognozy częściowe dla zmiennych objaśniających – predyktory z oraz bez składowych liniowych

4.3. Modeling continuous response variables

In Generalized Additive Models (GAM) for Location Scale and Shape the probability density function $f(y_i \mid \theta^i)$ is conditional on distribution parameter vector $\theta^i = (\theta_{1i}, \theta_{2i}, \theta_{3i}, \theta_{4i}) = (\mu_i, \sigma_i, \upsilon_i, \tau_i)$ for $i = 1, 2, ..., n$. Each of the parameters $(\mu_i, \sigma_i, \upsilon_i, \tau_i)$ may be a function of the predictors. The first two distribution parameters μ_i and σ_i are referred to as location and scale distribution parameters, whereas υ_i and τ_i are referred to as shape distribution parameters (skewness and kurtosis). The formulation of Generalized Additive Models (GAM) for Location Scale and Shape goes as follows [14]:

$$g_k(\theta_k) = \boldsymbol{\eta}_k = h_k(\mathbf{X}_k \boldsymbol{\beta}_k) + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}) \tag{2}$$

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = h_2(\mathbf{X}_1 \boldsymbol{\beta}_1) + \sum_{j=1}^{J_1} h_{j1}(\mathbf{x}_{j1}) \tag{3}$$

$$g_2(\boldsymbol{\mu}) = \boldsymbol{\eta}_2 = h_2(\mathbf{X}_2\, \boldsymbol{\beta}_2) + \sum_{j=1}^{J_2} h_{j2}(\mathbf{x}_{j2}) \tag{4}$$

$$g_3(\boldsymbol{\mu}) = \boldsymbol{\eta}_3 = h_3(\mathbf{X}_3\, \boldsymbol{\beta}_3) + \sum_{j=1}^{J_3} h_{j3}(\mathbf{x}_{j3}) \tag{5}$$

$$g_4(\boldsymbol{\mu}) = \boldsymbol{\eta}_4 = h_4(\mathbf{X}_4\, \boldsymbol{\beta}_4) + \sum_{j=1}^{J_4} h_{j4}(\mathbf{x}_{j4}) \tag{6}$$

where:

$\mathbf{y}^T = (y_1, y_2, ..., y_n)$	– the vector of response variables,
θ_k for $k = 1, 2, 3, 4$	– the distribution parameter vector,
$\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\upsilon}$, $\boldsymbol{\tau}$ and $\boldsymbol{\eta}_k$	– vectors of length n,
$\boldsymbol{\beta}_k^T = (\beta_{1k}, \beta_{2k}, ...\beta_{J_k^{'}k})$	– a parameter vector of length $J_k^{'}$,
$\mathbf{X}_k$	– a known design matrix of order $n \times J_k^{'}$,
$\mathbf{x}_{jk}$ for $j = 1, 2, ..., J_k$	– vectors of length n,
$g_k(.)$	– monotonic link functions relating the distribution parameters $(\mu_i, \sigma_i, \upsilon_i, \tau_i)$ to explanatory variables,
h_{jk}	– an unknown smoothing function of explanatory variables x_{jk},
$\mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$	– a vector evaluating the function h_{jk} at $\mathbf{x}_{jk}$,
h_k	– nonlinear parametric function of explanatory variables.

If equation (2) does not include any of the additive terms in any of the distribution parameters ($J_k = 0$) then the model defined by (2) reduces to the nonlinear parametric model $g_k(\theta_k) = \eta_k = h_k(\mathbf{X}_k\, \boldsymbol{\beta}_k)$. If additionally $h_k(\mathbf{X}_k\, \boldsymbol{\beta}_k) = \mathbf{X}_k^T\, \boldsymbol{\beta}_k$ then model (2) reduces to the linear parametric one $g_k(\theta_k) = \eta_k = \mathbf{X}_k \boldsymbol{\beta}_k$.

Equation (2) allows for modeling the distribution parameters as linear/nonlinear parametric function ($h_k(\mathbf{X}_k\boldsymbol{\beta}_k)$) and nonparametric smooth function $\left(\sum_{j=1}^{J_k} h_{jk}(x_{jk})\right)$ of explanatory variables.

The form of the response distribution $f(y_i|\mu_i,\sigma_i,\upsilon_i,\tau_i)$ in Generalized Additive Models (GAM) for Location Scale and Shape may be very general. Table 7 compares the goodness of fit based on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for 28 different continuous distributions fitted to continuous response variables. For details about these distributions, please refer to Johnson, Kotz and Kemp [10].

Table 7

Continuous distributions applied to Mean SBP and Mean DBP data – Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)

	Mean SBP			Mean DBP		
	Df	**AIC**	**BIC**	**Df**	**AIC**	**BIC**
Box-Cox Power Exponential	16	**177773**	**177901**	16	167818	167946
Box-Cox-t	16	**177809**	**177936**	16	167825	167953
Inverse Gaussian	14	177895	178007	14	169376	169488
Zero Adjusted IG	15	177897	178017	15	169378	169498
Generalized Beta Type 2	16	177904	178032	16	167940	168068
Box-Cox Cole and Green	15	178230	178350	15	167880	168000
Generalized Gamma	15	178366	178486	15	167916	168036
Generalized Inverse Gaussian	15	178484	178603	15	168446	168566
Log Normal	14	178619	178732	14	169051	169163
Log Normal (Box-Cox)	14	178620	178732	14	169051	169163
Gamma	14	178896	179008	14	168444	168556
Shash	16	178896	179024	16	167811	167939
Johnson's SU (the mean)	16	178930	179058	16	167764	167893
Skew t Type 1	16	178986	179113	16	167767	167895
Johnson's Original SU	16	179076	179204	16	168173	168301
Skew t Type 2	16	179091	179220	16	167766	167894
Skew Power Exponential Type 1	16	179152	179280	16	167764	167892
Skew Power Exponential Type 2	16	179298	179427	16	167764	167892
Generalized y	16	179765	179893	16	167762	167890
t Family	15	179768	179888	15	167764	167884
Power Exponential	15	179911	180032	15	**167761**	**167881**
Reverse Gumbel	14	180144	180257	14	171709	171821
NET	14	180166	180278	14	169169	169281

	Mean SBP			Mean DBP		
	Df	**AIC**	**BIC**	**Df**	**AIC**	**BIC**
Normal	14	181593	181705	14	167829	167942
Weibull	14	184068	184180	14	168911	169023
Gumbel	14	190253	190365	14	171757	171869
Exponential	13	256838	256942	13	231641	231745

With both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), the Box-Cox Power Exponential (BCPE) and the Box-Cox-t (BCT) distributions come as the best ones in approximating Mean DBP. For Mean SBP, both AIC and BIC favor the Power Exponential distribution. The Box-Cox Power Exponential (BCPE) and the Box-Cox-t (BCT) are continuous four parameter distributions (μ, σ, υ, τ) They generalize the Box-Cox Cole and Green distribution (BCCG) to allow for modeling kurtosis and skewness [3]. The Power Exponential distribution requires three distribution parameters (μ, σ, υ). For details about probability density functions of the Box-Cox-t/Box-Cox Power Exponential and the Power Exponential distributions please refer to Rigby and Stasinopoulos [14].

The estimation of continuous response variables assuming the Box-Cox Power Exponential distribution (BCPE) for Mean DBP and the Power Exponential distribution for Mean SBP is preceded by the selection of explanatory variables. The first distribution parameter μ (the mean of the response) is modeled assuming the inclusion of all the explanatory variables defined at the very beginning (Table 4). Table 8 checks whether the model can be simplified by potential dropping any of the terms in μ.

Table 8

Single term deletions for μ

	Mean DBP			Mean SBP		
	AIC	**LRT**	**Pr(Chi)**	**AIC**	**LRT**	**Pr(Chi)**
Age Cohort (years)	170573	2760	< 0.0001**	184047	4141	< 0.0001**
Gender (1=Male, 2=Female)	167833	17	< 0.0001**	180078	169	< 0.0001**
BMI Group (kg/m²)	167883	69	< 0.0001**	180219	312	< 0.0001**
Uric acid (mg/dL)	167861	45	< 0.0001**	179984	75	< 0.0001**
HDL-cholesterol (mmol/L)	167816	<0	0.6537	179933	23	< 0.0001**
Gamma Glut. Trans. (U/L)	167978	161	< 0.0001**	180048	139	< 0.0001**
Family PIR	167839	22	< 0.0001**	179956	47	< 0.0001**
Glucose (mmol/L)	167843	27	< 0.0001**	179962	52	< 0.0001**
Creatinine (umol/L)	167841	25	< 0.0001**	179969	60	< 0.0001**

Based on the Chi square test, HDL-cholesterol (mmol/L) does not contribute significantly to Mean DBP (no significant reduction of Akaike Information Criterion (AIC) as assessed by Likelihood-Ratio Test (LRT)). For Mean SBP, no terms can be left out, all of them will contribute to the final model.

Modeling the distribution parameters (μ, σ, υ) and τ of continuous response variables and thus selecting the best distributions for Mean DBP/Mean SBP is performed based on the linear parametric functions of the predictors (Table 7). For fitting nonlinear and nonparametric smooth functions, as the next step, additive term functions are applied and checked for the goodness of fit. Note that in this paper the modeling of the Box-Cox Power Exponential distribution (BCPE) for Mean DBP and the Power Exponential distribution for Mean SBP as the nonparametric smooth terms is restricted to Cubic Smoothing Spline Functions. Alternative additive terms, such as: Penalized Splines [4], Thin-Plate Smoothing Splines, Local Regression Splines [2], Fractional Polynomials, Power Polynomials [15], Random effects, Random coefficients [1], although very attractive, are not employed and not compared in this paper. Of particular notice are Random effects and Generalized Additive Mixed Models (GAMM) which pose very different and flexible approach to estimating Generalized Additive Models (GAM) [13].

Cubic Smoothing Spline Functions are curves which are made up of sections of joined cubic polynomials so that the functions $h(x)$ in model (2) are continuous in value and twice continuously differentiable. They are extensively covered in the literature [7, 9], thus their derivation is omitted.

Given $X = x$, Mean DBP is modeled by the Box-Cox Power Exponential distribution denoted as $BCPE(\mu, \sigma, \upsilon, \tau)$ where the distribution parameters (μ, σ, υ) and τ are modeled as smooth nonparametric functions of x, i.e.: $Y \sim BCPE(\mu, \sigma, \upsilon, \tau)$ where $g_1(\mu)$, $g_2(\mu)$, $g_3(\upsilon)$, $g_4(\tau)$ are defined by (3)–(6), respectively and for $k = 1, 2, 3, 4$, $g_k(.)$ are known link functions. The similar approach applies to Mean SBP (the Power Exponential distribution), i.e. $Y \sim PE(\mu, \sigma, \upsilon)$.

In order to establish whether smoothing terms are needed in the μ model defined by (3), all possible combinations of linear and cubic spline functions to the data are fitted. For each of the estimated models, Akaike Information Criterion (AIC) is assessed (stepwise model selection). The selection process is very time-consuming and its outputs very extensive. Thus, the full selection process of smoothing cubic splines is not included in this paper. Table 9 and Table 10 are brief summaries of the results.

Table 9

Summary of the selection process for Mean DBP – the first distribution parameter μ

From	To	Deviance	Resid. Df	Resid. Dev	AIC
LINEAR(Creatinine)	CS(Creatinine)	–181	22186	167606	167641
LINEAR(GGT)	CS(GGT)	–143	22183	167462	167504
LINEAR(Glucose)	CS(Glucose)	–70	22180	167392	167440
CLASS(Gender)		< 0	22181	167393	167439

Table 10

Summary of the selection process for Mean SBP – the first distribution parameter μ

From	To	Deviance	Resid. Df	Resid. Dev	AIC
LINEAR(GGT)	**CS(GGT)**	–141	22186	179740	179776
LINEAR (Creatinine)	**CS(Creatinine)**	–47	22183	179693	179735
LINEAR(Glucose)	**CS(Glucose)**	–31	22180	179662	179710
LINEAR(Uric acid)	**CS(Uric acid)**	–27	22177	179636	179689
LINEAR (HDL-cholesterol)	**CS(HDL-cholest.)**	–17	22174	179618	179678
LINEAR (Family PIR)	**CS(Family PIR)**	–16	22171	179603	179668

Having the model for μ determined, models for variance, skewness and kurtosis are searched for. The model selection procedure for these distribution parameters comprised of:

- Choosing link functions: for the Box-Cox Power Exponential distribution (Mean DBP), the default identity link functions are chosen for μ and υ, and log link functions are chosen for σ and τ;
- Selecting linear terms influencing the given distribution parameter: this is achieved by fitting models with linear terms;
- Applying cubic spline functions $h(x)$ to explanatory variables exhibiting nonparametric relation to the given distribution parameter: this is verified based on the Akaike Information Criterion (AIC);
- Selecting appropriate level of the "smoother" for each of the predictors modeled nonparametrically, i.e. choosing the effective degrees of freedom (DF) for smooth cubic spline functions $h(x)$ and denoted as df_μ, df_σ, df_υ and df_τ respectively: this is achieved by employing numerical optimization function to minimize the Generalized Akaike Information Criterion $GAIC(\#) = -2\hat{\ell} + \# df$ over hyper-parameters df_μ, df_σ, df_υ and df_τ in the Box-Cox Power Exponential model for Mean DBP, where $\hat{\ell}$ is the maximized log-likelihood function, # denotes penalty, *df* refers to the effective degrees of freedom (DF) and $-2\hat{\ell}$ is referred to as the global deviance. Due to limited space for this paper, the process of finding and searching for the best fit is not presented here. Note that the best model is found for hyper-parameters corresponding to penalty #= 2 and results in selecting a vector of hyper-parameters minimizing *GAIC* (2). For more details please refer to Rigby and Stasinopoulos [14].

Given that *CS* is Cubic Spline function and *df* denotes degrees of freedom (DF), the model for Mean DBP defined by four distribution parameters μ, σ, υ and τ of the Box-Cox Power Exponential *BCPE*(μ, σ, υ, τ) (for patients from Age Cohort: 57–<85 years and with Body Mass Index (BMI) Group: 30–<45 kg/m²) is given by:

$$\begin{cases} \mu = 58.09 + 6.05*Age\ Cohort + 1.33*BMI\ Group + 0.25*Uric\ acid \\ \quad + 0.04*CS(Gamma\ Glutamyl\ Transf., df = 6.48) - 0.25*Family\ PIR \\ \quad - 0.21*CS(Glucose, df = 7.25) + 0.03*CS(Creatinine, df = 9.49) \\ \log(\sigma) = -1.92 + 0.02*Age\ Cohort + 0.01*CS(Uric\ acid, df = 0.89) \\ \quad - 0.02*Family\ PIR + 0.01*CS(Glucose, df = 3.20) \\ \upsilon = 1.41 - 0.20*Age\ Cohort + 0.06*Family\ PIR \\ \quad - 0.003*CS(Creatinine, df = 2.76) \\ \log(\tau) = 0.71 - 0.13*Age\ Cohort + 0.09*Gender \end{cases}$$

Applying similar model selection procedure to Mean SBP for the first three distribution parameters, the model assuming the Power Exponential distribution PE(μ, σ, υ) is given by:

$$\begin{cases} \mu = 96.23 + 2.75*Age\ Cohort + 1.46*Gender + 4.59*BMI\ Group \\ \quad + 0.40*CS(Uric\ acid,\ df = 2.18) + 0.93*CS(HDL\text{-}cholest., df = 3.90) \\ \quad + 0.05*CS(Gamma\ Glutamyl\ Transf.,\ df = 8.11) \\ \quad - 0.20*CS(Family\ PIR,\ df = 2.39) + 0.79*CS(Glucose,\ df = 6.52) \\ \quad + 0.06*CS(Creatinine,\ df = 3.54) \\ \log(\sigma) = 2.03 + 0.76*Age\ Cohort - 0.04*Gender \\ \quad + 0.04*CS(HDL\text{-}cholest., df = 0.90) \\ \quad + 0.0008*CS(Gamma\ Glutamyl\ Transf., 2.55) - 0.02*Family\ PIR \\ \quad + 0.01*CS(Glucose, df = 1.34) + 0.001*Creatinine \\ \log(\upsilon) = 0.53 - 0.09*Age\ Cohort + 0.06*Gender \end{cases}$$

5. Discussion

The real-life example of Mean DBP/Mean SBP demonstrates the need for going beyond the exponential family distributions and thus, the usefulness of Generalized Additive Models (GAM) for Location, Scale and Shape. This overcomes the shortcomings associated with standard Generalized Additive Models (GAM) and Generalized Linear Models (GLM). In Generalized Additive Models (GAM) for Location, Scale and Shape the assumption of exponential family distributions is relaxed. It allows for modeling not only the mean associated with the location but also other distribution parameters of the response variable as additive nonparametric smoothing functions of explanatory variables. In the case of Mean DBP/Mean SBP data, the usage of the Box-Cox Power Exponential/Power Exponential distributions within Generalized Additive Models (GAM) allowed for modeling kurtotic distributions.

The estimated Generalized Additive Model (GAM) for binary Hypertension/ Borderline Hypertension response indicates that all the explanatory variables influence hypertension: Age Cohort, Gender, Body Mass Index (BMI) Group, Uric Acid, HDL-cholesterol, Gamma Glutamyl Transferase (GGT), Family PIR, Glucose and Creatinine. All explanatory variables are statistically significant with *p*-values much

lower than the assumed significance level of 0.05. It applies also to Generalized Additive Models (GAM) for Location, Scale and Shape estimated for Mean SBP. The results suggest that Mean DBP (which refers to the pressure when the heart is resting between beats) does not depend on Gender and HDL-cholesterol.

Systolic Blood Pressure seems to have stronger relationship with physiological and medical attributes than Diastolic Blood Pressure. All explanatory variables are nonlinear predictors of Systolic Blood Pressure. The models estimated for Hypertension/Borderline Hypertension and Mean DBP/Mean SBP suggest that:

- The blood pressure is higher among older subjects and subjects with elevated Body Mass Index (BMI). The risk of Hypertension/Borderline Hypertension is higher among population groups with overweight and obesity, particularly in Body Max Index Group >=30 kg/m^2. A similar trend prevails for Age Cohorts. The slopes of blood pressure are significantly higher in men than women (Table 5).
- People with elevated Uric Acid levels are at greater risk of hypertension. It is backed by other studies on Uric Acid [5]. The researchers conclude that Uric Acid may be an independent driver of high blood pressure and a marker of its prediction. A simple blood test can determine how much of it is present in the body. Effective drugs already exist which lower the level of Uric Acid and thus, offer a potential remedy for high blood pressure prevention.
- HDL-cholesterol levels influence Systolic Blood Pressure response. This is due to the association between high HDL-cholesterol and atherosclerosis (accumulation of HDL-cholesterol on the walls of arteries – hardening of the arteries). High HDL-cholesterol accelerates the progression of atherosclerosis which is thought to contribute to hypertension. This link is not significant for Diastolic Blood Pressure readings.
- There is a positive link between Gamma Glutamyl Transferase (GGT) levels, a marker of oxidative stress and blood pressure. Although the underlying mechanism of this association is still unclear, some studies confirm that higher Gamma Glutamyl Transferase (GGT) is implicated in the pathogenesis and progression of hypertension [17].
- People in low socioeconomic status environments as assessed by Income-Poverty Ratio are more susceptible to illnesses. Those with lower income tend to be at greater risk of hypertension. These findings are pretty worrying, especially in light of the fact that most researches to date are concentrated on hypertension in developed urban countries. As a result, very little is known about the problems and barriers to treatment and diagnosis outside high-income areas. This issue is even more acute knowing that the epidemic of cardiovascular disease occurring in low-income nations is largely driven by the increasing prevalence of high blood pressure.
- Glucose is able to induce Systolic Blood Pressure. Elevated Glucose level increases the likelihood of having diabetes which leads to higher Systolic Blood Pressure and heart diseases. The relationship with Diastolic Blood Pressure seems to be reversed.
- Higher blood pressure is associated with elevated Serum Creatinine level (an indicator of chronic renal disease). Creatinine draws water into the muscle what increases body weight and muscle volume. Retaining more water in the body impacts the blood volume and thus, blood pressure. Of note is the fact that half of the body's Creatinine is created naturally (produced in the liver and kidney) whereas the rest comes from the diet (the consumption of red meat and poultry). It confirms that appropriate diet plays its role in the treatment of blood pressure.

6. Conclusions

In this paper, the underlying methodology for Generalized Additive Models (GAM) has been introduced. The real-life data set example has demonstrated how one can use the nonparametric approach to model medical scheme data. This was achieved by investigating the dependencies and patterns of Hypertension/Borderline Hypertension and Mean DBP/Mean SBP on various physiological measurements, medical attributes and Income to Poverty Ratio. It is concluded that Generalized Additive Models (GAM) are a very powerful and flexible tool in an exploratory analysis, especially when you have little prior information about the data or you want to find new features that parametric tools ignore. Its nonparametric nature does not require much prior information and can also shed light on underlying parametric relationships. Generalized Additive Models (GAM) help to avoid model misspecification and provide information that might not be revealed by standard modeling techniques. The presented Generalized Additive Models (GAM) revealed pronouncedly complex nonlinear patterns in the response-predictor relationships for all predictors entered into the model. These nonlinear associations have been handled without the restrictions of parametric models, without sacrificing the interpretability and without the bias associated with the "curse of dimensionality". The built Generalized Additive Models (GAM) seem to represent the behavior of the data closer than the parametric counterparts. It underlines the importance of this class of models in detecting nonlinear dependencies and suggests potential failure of parametric solutions in capturing important features of the medical scheme data.

References

[1] Chambers J.M., Hastie T.J., *Statistical Models in S*, Chapman & Hall, London 1992.

[2] Cleveland W.S., Grosse E., Devlin S.J., *Regression By Local Fitting*, Journal of Econometrics, vol. 37, 1988, 87-114.

[3] Cole T.J., Green P.J., *Smoothing reference centile curves: the LMS method and penalized likelihood*, Statistical Modeling, vol. 11, 1902, 1305-1319.

[4] Eilers P.H., Marx B.D., *Flexible smoothing with B-splines and penalties*, Statistical Science, vol. 11, 1996, 89-121.

[5] Feig D.I, Kang D.H., Nakagawa T., Mazzali M., Johnson R.J., *Uric acid and hypertension*, Curr Hypertens Rep., vol. 8(2), 2006, 111-115.

[6] Friedman J.H., Stuetzle W., *Projection Pursuit Regression*, Journal of the American Statistical Association, vol. 76, 1981, 817-823.

[7] Green P.J., Silverman B.W., *Nonparametric Regression and Generalized Linear Models*, Chapman & Hall, London 1994.

[8] Gurven M., Blackwell A., Rodríguez A., Stieglitz J., Kaplan H., *Does Blood Pressure Inevitably Rise With Age?*, Longitudinal Evidence Among Forager-Horticulturalists, Hypertension, vol. 60(1), 2012, 25-33.

[9] Hastie T.J., Tibshirani R.J., *Generalized Additive Models*, Chapman & Hall, London 1990.

[10] Johnson N.L., Kotz S., Kemp A.W., *Univariate Discrete Distributions*, Wiley, New York 2005.

[11] Lee J.A., Verleysen M., *Nonlinear Dimensionality Reduction*, Springer, New York 2007.
[12] Łukasik S., Kulczycki P., *Using Topology Preservation Measures for High-Dimensional Data Analysis in a Reduced Feature Space*, Technical Transactions, vol. 1-AC/2012, Cracow University of Technology Press, 5-15.
[13] Pinheiro J.C., Bates D.M., *Mixed-Effects Models in S and S-PLUS*, Springer-Verlag, New York 2000.
[14] Rigby R.A., Stasinopoulos D.M., *Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis*, Statistical Modeling, vol. 6, 2006, 209-229.
[15] Royston P., Altman, D.G., *Regression using fractional polynomials of continuous covariates: parsimonious parametric modeling*, Appl. Statist., vol. 43, 1994, 429-467.
[16] Stone C.J., Hansen M., Kooperberg C., Truong Y.K., *Polynomial splines and their tensor products in extended linear modeling*, Annals of Statistics, vol. 25, 1997, 1371-1470.
[17] Shankar A., Li J., *Association between serum gamma-glutamyltransferase level and prehypertension among US adults*, Circ J., vol. 71(10), 2007, 1567-1572.
[18] Tomera J.F., Harakal C., *Multiple linear regression analysis of blood pressure, hypertrophy, calcium and cadmium in hypertensive and nonhypertensive states*, Food and Chemical Toxicology, vol. 35(7), 1997, 713-718.
[19] Wahba G., *Bayesian Confidence Intervals for the Cross Validated Smoothing Spline*, Journal of the Royal Statistical Society, vol. 45, 1983, 133150.
[20] Webpage of National Health & Nutrition Examination Survey (source of the data set used for the analysis): http://www.cdc.gov/nchs/nhanes.htm.

Table 11

Distribution parameters of Box-Cox Power Exponential distribution (BCPE) – Mean DBP

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
μ – Mu Coefficients (Mu link function: identity)				
(Intercept)	58.09	0.481	120.7	< 0.0001**
Age Cohort (23–<40 vs 12–<23 years)	5.76	0.212	27.1	< 0.0001**
Age Cohort (40–<57 vs 12–<23 years)	10.81	0.222	48.7	< 0.0001**
Age Cohort (57–<85 vs 12–<23 years)	6.05	0.240	25.2	< 0.0001**
BMI Group (25–<30 vs 14–<25 kg/m²)	< 0.01	0.184	< 0.1	0.1986
BMI Group (30–<45 vs 14–<25 kg/m²)	1.33	0.202	6.6	< 0.0001**
LINEAR(Uric acid)	0.25	0.065	3.8	< 0.0001**
CS(Gamma Glutamyl Transferase, df = 6.48)	0.04	0.003	13.1	< 0.0001**
LINEAR(Family PIR)	–0.25	0.046	–5.3	< 0.0001**
CS(Glucose, df = 7.25)	–0.21	0.067	–3.1	< 0.0001**
CS(Creatinine, df = 9.49)	0.03	0.004	6.2	< 0.0001**

	Estimate	Std. Error	t value	Pr(>|t|)
σ – Sigma Coefficients (Sigma link function: log)				
(Intercept)	–1.92	0.031	–62.5	< 0.0001**
Age Cohort (23–<40 vs 12–<23 years)	–0.07	0.015	–4.7	< 0.0001**
Age Cohort (40–<57 vs 12–<23 years)	–0.19	0.015	–12.6	< 0.0001**
Age Cohort (57–<85 vs 12–<23 years)	0.02	0.015	1.4	0.1503
CS(Uric acid, df = 0.89)	0.01	0.004	3.7	0.0002**
LINEAR(Family PIR)	–0.02	0.003	–5.9	< 0.0001**
CS(Glucose, df = 3.20)	0.01	0.004	2.8	0.0049**
υ – Nu Coefficients (Nu link function: identity)				
(Intercept)	1.41	0.117	12.0	< 0.0001**
Age Cohort (23–<40 vs 12–<23 years)	–0.39	0.101	–3.8	< 0.0001**
Age Cohort (40–<57 vs 12–<23 years)	–0.67	0.113	–5.5	< 0.0001**
Age Cohort (57–<85 vs 12–<23 years)	–0.20	0.094	–2.2	0.0246*
LINEAR(Family PIR)	0.06	0.022	2.6	0.0082*
CS(Creatinine, df = 2.76)	< –0.01	0.001	–2.9	0.0032**
τ – Tau Coefficients (Tau link function: log)				
(Intercept)	0.71	0.040	17.8	< 0.0001**
Age Cohort (23–<40 vs 12–<23 years)	–0.19	0.042	–4.5	< 0.0001**
Age Cohort (40–<57 vs 12–<23 years)	–0.18	0.043	–4.2	< 0.0001**
Age Cohort (57–<85 vs 12–<23 years)	–0.13	0.042	–3.2	0.0014**
Gender (Male vs Female)	0.09	0.030	2.9	0.0034**
Degrees of Freedom for the fit: 68.90, No. of observations in the fit: 22204				

Table 12

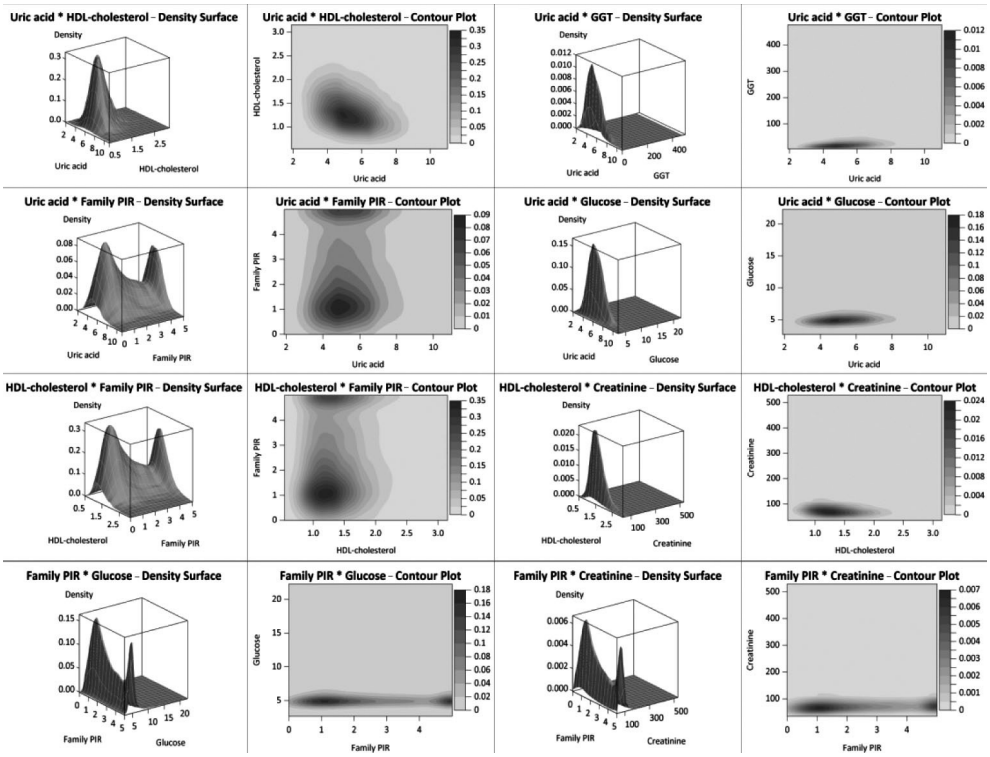
Distribution parameters of Power Exponential distribution (PE) – Mean SBP

	Estimate	Std. Error	t value	Pr(>|t|)
μ – Mu Coefficients (Mu link function: identity)				
(Intercept)	96.23	0.696	138.2	< 0.0001**
Age Cohort (23–<40 vs 12–<23 years)	6.18	0.250	24.7	< 0.0001**
Age Cohort (40–<57 vs 12–<23 years)	18.23	0.319	57.2	< 0.0001**

	Estimate	Std. Error	t value	Pr(>|t|)
Age Cohort (57–<85 vs 12–<23 years)	2.75	0.198	13.9	< 0.0001**
Gender (Male vs Female)	1.46	0.191	7.6	< 0.0001**
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
BMI Group (25–<30 vs 14–<25 kg/m²)	2.20	0.194	11.3	< 0.0001**
BMI Group (30–<45 vs 14–<25 kg/m²)	4.59	0.219	20.9	< 0.0001**
CS(Uric acid, df = 2.18)	0.40	0.077	5.2	< 0.0001**
CS(HDL-cholesterol, df = 3.90)	0.93	0.230	4.1	< 0.0001**
CS(Gamma Glutamyl Transferase, df = 8.11)	0.05	0.004	13.5	< 0.0001**
CS(Family PIR, df = 2.39)	–0.20	0.045	–4.2	< 0.0001**
CS(Glucose, df = 6.52)	0.79	0.085	9.4	< 0.0001**
CS(Creatinine, df = 3.54)	0.06	0.005	12.2	< 0.0001**
σ – Sigma Coefficients (Sigma link function: log)				
(Intercept)	2.03	0.042	48.1	< 0.0001**
Age Cohort (23–<40 vs 12–<23 years)	0.12	0.015	7.9	< 0.0001**
Age Cohort (40–<57 vs 12–<23 years)	0.48	0.016	29.2	< 0.0001**
Age Cohort (57–<85 vs 12–<23 years)	0.76	0.016	48.7	< 0.0001**
Gender (Male vs Female)	–0.04	0.013	–3.3	< 0.0001**
CS(HDL-cholesterol, df = 0.90)	0.04	0.015	3.1	< 0.0001**
CS(Gamma Glutamyl Transferase, df = 2.55)	< 0.01	< 0.001	4.1	< .0001**
LINEAR(Family PIR)	–0.02	0.003	–6.7	< 0.0001**
CS(Glucose, df = 1.34)	0.01	0.004	2.9	< 0.0001**
LINEAR(Creatinine)	< 0.01	<0.001	4.1	< 0.0001**
υ – Nu Coefficients (Nu link function: log)				
(Intercept)	0.53	0.065	8.4	<0.0001**
Age Cohort (23–<40 vs 12–<23 years)	–0.19	0.041	–4.7	<0.0001**
Age Cohort (40–<57 vs 12–<23 years)	–0.32	0.041	–7.8	<0.0001**
Age Cohort (57–<85 vs 12–<23 years)	–0.09	0.042	–2.0	0.0430*
Gender (Male vs Female)	0.06	0.028	2.1	0.0350*
Degrees of Freedom for the fit: 63.59, No. of observations in the fit: 22204				

Appendix 1

Kernel Density for selected Explanatory Variables: Density Surface and Contour Plot



Appendix 2

Probability density functions of Box-Cox-t, Box-Cox Power Exponential and Power Exponential distributions

The probability density function $f_T(y_i|\mu_i,\sigma_i,\upsilon_i,\tau_i)$ of the Box-Cox-t distribution (BCT) is given by:

$$f_T(y\,|\,\mu,\sigma,\upsilon,\tau)=\frac{y^{\upsilon-1}f_T(z)}{\mu^{\upsilon}\sigma F_T\left(\dfrac{1}{\sigma\,|\,\upsilon\,|}\right)} \tag{7}$$

for $y > 0$, where Y is a positive random variable of the Box-Cox t distribution, $\mu > 0$, $\sigma > 0$, and $-\infty < \upsilon < \infty$ and Z is the transformed random variable given by:

$$Z = \begin{cases} \dfrac{1}{\sigma\upsilon}\left[\left(\dfrac{Y}{\mu}\right)^{\upsilon} - 1\right], & \text{if} \quad \upsilon \neq 0 \\ \dfrac{1}{\sigma}\log\left(\dfrac{Y}{\mu}\right), & \text{if} \quad \upsilon = 0 \end{cases} \tag{8}$$

where Z follows a truncated t distribution with degrees of freedom (DF), $\tau > 0$. $f_T(t)$ and $F_T(t)$ are respectively the probability density function and the cumulative distribution function of a random variable T. T has a standard t distribution with degrees of freedom (DF) $\tau > 0$.

The probability density function of the Box-Cox Power Exponential distribution (BCPE) is given by (7) where Z follows a truncated standard Power Exponential distribution with power distribution parameter, $\tau > 0$. The Power Exponential distribution requires three parameters. The probability density function of the Power Exponential family distribution is defined by:

$$f_Y(y \mid \mu, \sigma, \upsilon) = \frac{\upsilon \exp\left[-\left|\dfrac{z}{c}\right|^{\upsilon}\right]}{2c\sigma\Gamma\left(\dfrac{1}{\upsilon}\right)} \tag{9}$$

for $-\infty < y < \infty$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $\upsilon > 0$ and where $z = (y - \mu)/\sigma$ and $c^2 = \Gamma\left(\dfrac{1}{\upsilon}\right)\left[\Gamma\left(\dfrac{3}{\upsilon}\right)\right]^{-1}$. In this parameterization, $E(Y) = \mu$ and Var $(Y) = \sigma^2$.

CONTENTS

TREŚĆ