INFORMATION SYSTEMS IN MANAGEMENT

Information Systems in Management (2019) Vol. 8 (1) 69−81
DOI: 10.22630/ISIM.2019.8.1.7

CLUSTER ANALYSIS AS A PRELIMINARY PROBLEM IN NEURAL MODELLING OF THE POLISH POWER EXCHANGE

JERZY TCHÓRZEWSKI [a)], JAKUB JEZIERSKI [b)]

[a)] *Computer Science Institute, Faculty of Sciences, UPH in Siedlce*
[b)] *Member of Computer Science Branch, Student, Faculty of Sciences, UPH in Siedlce*

The work focuses on cluster analysis as a preliminary problem in neural modelling based on the data quoted on the Day Ahead Market of the Polish Power Exchange as a subsystem of the system of Towarowa Giełda Energii S.A. [Polish Power Exchange]. The paper contains the results of literature research related to cluster analysis methods, description of possible applications of artificial neural networks SOM for mapping information on the volume of electrical power sold and prices obtained, description of possible applications of MATLAB and Simulink environment, and especially Neural Network Toolbox for mapping knowledge, and cluster analysis performed for selected data.

Keywords: artificial self-organizing neural networks, business intelligence, cluster analysis, neural modelling, Day Ahead Market, Polish Power Exchange.

1. Introduction

Due to immense increase in information and data towards the end of the 20th century, a need for the automated preliminary analysis and visualization of the obtained results arose. Among numerous tools for knowledge visualization, especially for cluster analysis, numerous environment tools such as MATLAB and Simulink, including Neural Network Toolbox were developed [2]. First works on the subject demonstrated that artificial neural networks, including SOM (Self-Organizing Map) played a significant role in the process of cluster analysis [11, 16-20, 22-23]. Due to

the fact that the number of transactions concluded on the Day Ahead Market (DAM) of the Polish Power Exchange Market (PPEM), and, consequently, the amount of data related to the volume of electrical power (ep) delivered and sold [MWh] and prices obtained for ep sold [PLN/MWh] is rising, the need for preliminary analysis of the data arises [12]. Relatively new, though reliable method used for analysis of large datasets is cluster analysis, thoroughly described in works by Wierzchoń S. and Kłopotek M. [24-25].

Therefore, the main objective of the conducted research was to conduct cluster analysis using the Neural Network Toolbox in the MA-TLAB and Simuink environment as a preliminary issue prior to the use of Artificial Neural Networks (SSN) to model the TGEE system, that is to model neural TGEE. The basic problem to solve was to highlight in the data set specific clusters of information with similar features and properties using the appropriate cluster analysis method. The obtained data presented in the cluster analysis are called classes that differ in terms of creating specific, more or less explicit clusters of data projected, for example, on a plane. Cluster analysis assumes the existence of a sequence of objects, which is described by an n-dimensional vector.

In order to identify similarities between pairs of objects, their similarity or dissimilarity measures are introduced. The two measures depend on each other, i.e. if the value of dissimilarity decreases, the value of similarity increases. Clustering algorithms are used for comparing objects, which algorithms are thoroughly described e.g. in works [1, 3-5, 8-10, 21]. The notion of cluster analysis is also referred to as data grouping or data clustering, and was borrowed from data exploration and machine learning as a specific non-pattern classification [1, 4-5, 7-9, 21]. Therefore, cluster analysis is a method of classification without supervisor (unsupervised learning), i.e. obtained by employing a method grouping elements into homogeneous classes or data clusters, with the rule of grouping being specific similarity between the elements, determined using probability function called metric, appropriate for the problem being analysed [3-5, 21].

2. Cluster analysis methods

Among cluster analysis methods based on artificial neural networks (ANN), the following methods may be distinguished, related to, i.a. [3-6]:

- classification, i.e. division of the measurements results being analysed by means of ANN into classes based on their specific features,
- regression, i.e. assigning data, input to the artificial neural network in the form of known features of the system to results of observation,
- clustering, i.e. cluster analysis involving grouping training pairs (so called samples) according to similarity between them,

– anomaly detection, i.e. detection of clusters similar to the pattern that is completely different from the expected pattern,
– association rules, i.e. rules related to determination of relations between the features and attributes in the matrix of values of training (here: input) pairs to the artificial neural network.

At present, the methods of data analysis, including preliminary analysis of data are more advanced (e.g. systems for classifying images by Microsoft), and have much lower error rate than those conducted by people classifying objects or elements.

In order to improve the quality of data being processed, in artificial intelligence algorithms, including artificial neural networks, input data undergo appropriate preparation. The process of preparation of the database (input matrices) comprises the following stages:

– preliminary transformation of data related, e.g. to scaling individual numeric data in order to obtain appropriate values, often combined with supplementing the missing values and characteristics construction involving formation of new characteristics, describing processes occurring in the system from the point of view of the research goal; such analyses are performed using e.g. automated algorithms, including methods that learn so called representation,
– preliminary transformation of data related e.g. to scaling selection of characteristics involving verification - whether they describe the phenomenon being analysed correctly and exhaustively from the point of view of selection of the most significant relations, including assignment of weights to them or use of other methods of classification (e.g. ranking, tournament, assigning ranks, etc.).

Classification can be defined as assignment of an object or an element to a specific input class of the sample, whose class is unknown. Assignment to the class is performed by a trained model called classifier, which acts based on the value of known characteristics or similarity of characteristics.

It is taught using sequence of samples e.g. in the form of vectors or matrices, for which both the class and the values of characteristics are known. In this case, supervised teaching is applied. There is also a possibility of using various classification algorithms. Therefore, it is necessary to select the best algorithm for the classification problem being solved in order to obtain the best efficiency and effectiveness. The most frequently used algorithms include logic regression, decision trees, random forests, support vector machine, etc. They are characterized by both low complexity and high efficiency and effectiveness of classification. Logic regression is one of many algorithms, which are used for classification when a sample is assigned to one of two classes. A significant problem to solve is then selection of the appropriate method for determination of coefficients such as the least square method, decomposition method, Cholesky factorization (decomposition), or gradient method, etc.

Decision tree is a classifier, usually represented by a binary tree, in which nodes describe individual pieces of data in terms of the values of a selected characteristic, and leaves describe individual samples that belong to a given class.

Another method is random forest as a set of classifiers (ensemble classifier) in which each single classifier is a decision tree trained continuously without halts until a leaf contains only samples belonging to the same class. Each classifier in the random forest is trained using a random data sample selected for that particular classifier. Such a technique of data generation is called bagging or bootstrap aggregating [21, 24-25].

Another algorithm used in classification is Support Vector Machine algorithm, which determines a hyperplane that separates samples from various classes with a maximum margin between them. In order to ensure maximum width of the margin between the samples from various classes, minimization of the module of the vector $||\mathbf{w}||$ is performed, at the same time maintaining class welding. A separate problem is the assessment of classifier performance, conducted in order to establish its quality. For this purpose, two sets of data are required, the first one for training the classifier, i.e. so called training set, and the other one, used for testing the classifier. More detailed assessment of classifier may be obtained by *k*-fold testing, which is often performed as data validation. Various metrics such as permutation test are used for the assessment of the classifier. Clustering algorithms use different distance measures, the most important of which are presented in table 1. They include, i. a., the Minkowski distance, the Mahalanobis distance, the cosine distance, the power distance and the Bregman divergence. Measures and models corresponding to the above distances are presented in table 2. They are described in detail, i.a. in works [21-25].

In the literature on the subject, cluster analysis algorithms are divided into the following basic categories [3-5, 7-11, 19-21, 23-25]:

- hierarchical methods based on development of classification hierarchy for a set of objects, starting with such division in which each object constitutes a cluster, and finishing with a division, in which all objects belong to one cluster,
- grouping using k-means method – grouping that involves preliminary division of the population into a predefined number of clusters, followed by the improvement of the division by moving certain elements to other groups,
- fuzzy clustering method, including the best known c-means method, in which an element is assigned to more than one category, which allows for using fuzzy clustering algorithms in tasks involving classification of elements to one or more categories,
- methods that use Self-Organizing Maps, which allow for so called knowledge mapping and self-formation of clusters of data or information about data.

Table 1. List of distance measures

No and Name	Measure	Special cases	Application	Notes
1	2	3	4	5
1. Minkowski distance	Generalised measure of distance between points in Euclidean space.	P=1 city block (Mahnattan) distance P=2 Euclidean distance P=∞ Tchebychev distance is obtained.	Exact Sciences, Psychology, Design.	With the increase of dimensionality of the problem, the difference between close and far points dwindles The value of Minkowski distance is dominated by characteristics measured on a scale with the greatest span.
2. Mahalonobis distance	It is assumed that the characteristics are not correlated with one another.	When co-variance matrix is a diagonal matrix. The distance becomes weighted Euclidean distance.	When characteristics are not correlated For identification of outliers.	For $y_i = -\sum ½ x_i$ Is transformed to Euclidean distance
3. Cosine distance	Uses cosine of the angle, and is basic measure used in systems for finding information.	-	For measuring similarity between documents.	-
4. Power distance	Increase or decrease of increasing weight.	P = r distance is equal Minkowski distance.	Changing weight in the selected dimension	-
5. Bregman divergence	Considers more complex relations between the compared vectors.	-	In problems related to signal compression.	Is often transformed to other distances.

Source: own compilation on the basis of works [5, 21, 24-25]

Table 2. Mathematical models of distance measurement

No.	No	Mathematical model
1.	Minkowski distance	$d_p(xi,xj) = [\sum_{i=1}^{n} \lvert xil - xjl \rvert^p]^{1/p}$
2.	Mahalonobis distance	$d_{\sum}(xi,xj) = \sqrt{(x_i - x_j)^T \sum^{-1} (x_i - x_j)}$
3.	Cosine distance	$d_{cos}(xi,xj) = 1 - \frac{\sum_{i=1}^{n} xil\,xjl}{\lvert\lvert x_i \rvert\rvert \, \lvert\lvert x_j \rvert\rvert}$
4.	Power distance	$(d_{p,r}(x_i, x_j) = \sum_{i=1}^{n} \lvert xil - xjl \rvert^p)^{1/r}$
5.	Bregman divergence	$d_\varphi(x,y) = \varphi(x) - \varphi(y) - (x-y)^T \nabla\varphi(y)$

Source: own compilation on the basis of works [5, 21, 24-25]

Moreover, three different techniques of fuzzy clustering can be distinguished: partitioning algorithms, which involve finding an optimum division of a set of examples into a specified number of clusters (groups), hierarchical algorithms, which involve hierarchical attempt at discovering cluster structure, density-based algorithms, which divide sets of examples using probabilistic model for base clusters.

3. Polish Power Exchange

Towarowa Giełda Energii S.A. (TGE S.A.,) [Polish Power Exchange, PPE] is a joint stock company established in 1999 [12, 16, 18]. PPE offers its participants access to uniform market data, which can be accessed through the Company's platform, the access is open, and rules for all commercial transactions are uniform.

PPE is a subsystem of PE with transactions quotation e.g. Day Ahead Market (DAM). The data allow for finding the best offer on the market and managing prices and volume of electrical power required by customers. The main areas of activity of the Exchange include commodity transactions on the following markets: the Day Ahead Market, the Commodity Forward Instruments Market with Physical Delivery, Property Rights Market, Emission Allowances Market, etc.

Polish Power Exchange initiated introduction of new solutions related to trade in electrical power. It runs markets in which the biggest companies in power industry participate. PPE is constantly developing and adding new markets. At present Polish Exchange Market conducts activity related to trade in electricity and liquid and gaseous fuel, limiting emission of pollution, limiting the volume of production, property rights, etc.

The Day Ahead Market is a spot market for electrical power. Its main goal is to create prices of power for contracts concluded on the power market in Poland. Another goal of the DAM is to balance contract positions, assessment of companies value, generation of investment signals related to development of new power capacities, etc. The DAM on the TGE S.A. consists of 24-hour day markets, where specific hour contracts are quoted. Moreover, the DAM offers three types of block contracts, i.e. BASE – delivery of 1 MWh of electrical power per hour, PEAK – delivery of 1 MWh of electrical power between hours 7:00 – 22:00, Offpeak – delivery of 1MWh of electrical power outside peak hours e.g. 0:00-7:00 and 22:00-24:00.

TGE S.A. also permits submission and settle over-the-counter transactions (OCT) based on standards contracts quoted on the DAM session, and the trading volume is calculated with the accuracy of 0.01 zł/MWh (minimum volume equals 0.1 MWh). Quotations on the exchange take place daily in electronic form, by means of a special platform developed for this purpose. Selection of a particular day takes place following the selection of this day in the calendar together with the information, which, due to public access may be used in scientific and research

experiments. Numeric data from PPE is available to anyone. The values used in the calculations cover the period of three months (01.07.2010 - 30.09.2010). The values present electrical power delivered and sold [MWh] and the average weighted by the volume of electrical power price obtained at a given hour of the 24-hour day [PLN/MWh]. Due to relatively big amount of data for an example research experiment, i.e. the dimension of 24 x 92 (24 hours of the 24-hour day, 92 samples from the above mentioned three months), only the data related to the volume of electrical power [MWh] will be shown. In order to show how the data increase and decrease in particular hours of the 24-hour day and in particular days, the visualisation was performed for particular volumes of sold and delivered electrical power for appropriate hours and days for the average price obtained for electrical power sold.

4. Preliminary preparation of data for the experiment

Data was obtained in the form of two matrices 24 x 92 (for the period of 01.07.2010 - 30.09.2010) from PPE quoted on the DAM, i.e. matrix of ep volume and matrix of prices [12]. Therefore, the first matrix presents 24 inputs corresponding to the volume of electrical power delivered and sold [MWh] in particular hours of the 24-hour day, quoted on the DAM in the period of 3 months. The second matrix contains averages weighted by price volume obtained from power sold to customers [PLN/MWh] in particular hours of the 24-hour day in the above mentioned period. In order to improve the process of training artificial neural network, the data was normalized and the course was presented in fig. 1. In order to input data to Matlab Workspace two separate matrices were created, namely: **U** – matrix of ep volume **Y –** matrix of weighted averages of ep prices.

5. Research experiment

In order to create Self-Organizing Maps, newsom function was used, which creates a self-organizing map with two inputs and a network of neurons on the hexagonal map with dimensions x by y, as follows: [2]:

$$\mathrm{net} = \mathrm{newsom}([0\ 1; 0\ 2], [2\ 4]) \tag{1}$$

To create plots of a generated map, a standard function, e.g. plot may be used:

$$\mathrm{plot}(\mathrm{net}, '+\mathrm{r}'). \tag{2}$$

The values of weights of a learning artificial neural network SOM, following initiation are further determined based on the winner takes all rule (WTA) [2, 6, 10, 13, 16, 22,26]:

$$\|u(k) - w_c(k)\| = \min\{\|u(k) - w_i(k)\|\}, \tag{3}$$

where:

u(k) – input vector,

$\mathbf{w}_c$(k) – vector of winner neuron weights,

$\mathbf{w}_i$(k) – weight vector of i-th neuron.

In comparison with competitive networks, Kohonen networks not only adapt their weights as weights related to the winner neuron but also weights of all their neighbours within specified neighbourhood radius according to the following rule [6, 10]:

$$w \begin{cases} (1-\alpha)\, w_i(k) + \alpha_i(k), & for\ i \in \Omega \\ w_i(k), & in\ other\ wise \end{cases}, \tag{4}$$

where:

Ω- neighbourhood of the winner neuron,

α- learning step.

At the beginning of learning, a neuron predisposed to be the winner, is selected based on minimization of the difference between the weights of neurons, while the components of the input vector are determined according to the dependence [6, 24-25]:

$$d(x, W_w) = \min_{1<j<N} d(x, w_i) \tag{5}$$

where: $d(x, W_w)$ – distance function between the input vector x and weights of the neuron.

The most common standard distance functions used include:

– scalar product

$$d(x, W_i) = \|x\|\, \|W_i\| \cos(x\, W_i), \tag{6}$$

– Euclidean distance

$$d(x, W_i) = \|x - W_i\| = \sqrt{\sum_{j-i}^{N} (x_i - W_j^{(i)})^2}, \tag{7}$$

– Manhattan norm

$$d(x, W_i) = \sqrt{\sum_{j-1}^{N} \|x_j - W_j^{(i)}\|}\ , \tag{8}$$

and many other methods, including those presented in table 1 and in table 2 [3-5, 21-25].

In order to perform cluster analysis, Artificial Neural Network (SOM) was used, which was designed using Neural Network Toolbox and SOM Toolbox. The experiment involved 24 input quantities related to the volume of electrical power sold in particular hours of the 24-hour day [MWh] and 24 output quantities related to the average price obtained in particular hours of the 24-hour day for electrical energy sold [PLN/MWh]. Therefore, the input matrix and the output matrix have the dimensions of 24 x 92. This allowed to define a system of MIMO type (Multi Input i Multi Output) as a real fragment of a system, which is PPE quoted on the DAM. A model of the PPE was designed in the form of Artificial Neural Network of SOM type with 24 inputs, which was used for mapping knowledge related to regularities found in the real system. Competitive learning method was used, which involves competition of neurons on a grid with the dimensions 6 x 4 as in figure 1, i.e. values of weights of 24 neurons are projected onto a plane in order to show data clusters that show self-organization – figure 2.

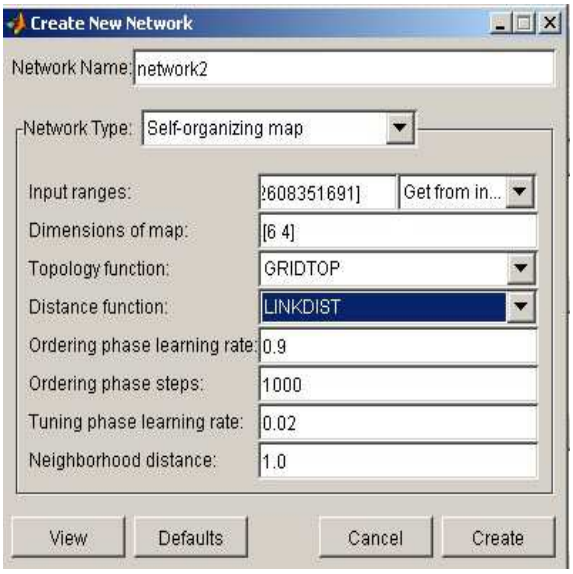

Figure 1. Designing ANN SOM. Denotations: LINKDIST – layer distance function, GRIDTOP – type of network topology, Get from Input – getting input data matrix. *Source*: elaboration using Neural Network Toolbox [5]

The figure shows four forming classes of neurons, i.e. four price categories set and quoted on the Day Ahead Market for electrical energy delivered and sold in the period being examined (in 24 hours of the 24-hour day). All the above mentioned classes are characterized by different strength of the relationship between the vol-

ume and the price of electrical energy, with the relationship being the strongest for class 1 and the weakest for class 4, respectively.

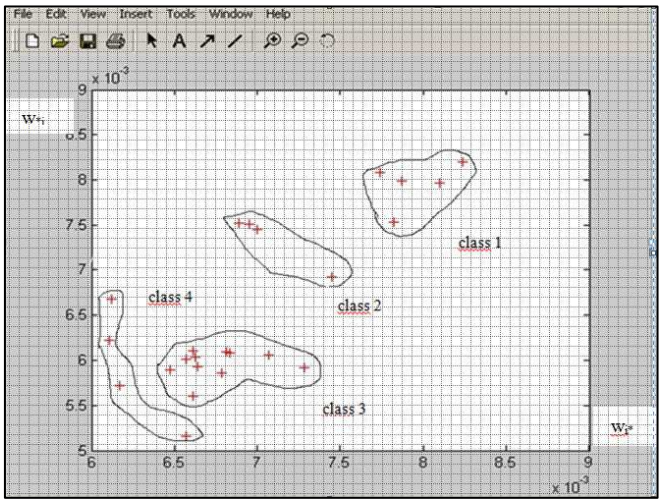

Figure 2. Mapping final values of 24 weights onto a plane using Artificial Neural Network SOM, available in SOM Toolbox. Denotations: x axis – values of weights related to neuron 1, i.e. weights w_{i*}, y axis – values of weights related to neuron 2, i.e. weights $w_{*j.}$ Elaboration using SOM Toolbox [5]

6. Conclusion and further research

Cluster analysis was performed as a preliminary problem of neural modelling of Polish Power Exchange. This was related to, i.a., performing literature research on cluster analysis methods in order to examine possibility of projecting knowledge in relation to PPE quoted on the DAM, selection of the artificial neural network (SOM) and its use for mapping information about the volume of energy sold and prices obtained on PPE DAM.

Possibilities of MATLAB and Simulink environment, especially Neural Network Toolbox were examined in the scope of the research, data was prepared and cluster analysis was performed using the prepared data, with no predefined cluster analysis method, e.g. k-means. Instead, SOM (Self-Organizing Maps) were used. Four classes were obtained indicating the occurrence of four categories of prices on PPE quoted on the DAM.

Thus, the fundamental goal of the research was fulfilled as a result of designing and conducting preliminary analysis of real data concerning PPE listed on the DAM. It turned out that there is a possibility of separating four classes covering

points on the knowledge map (fig. 2), which can be further used to carry out in-depth research in assigning to individual neurons the numbers of hours and possible tracking of changes in their position in separate hours 24 hours a day for the entire examined period, which sets a new direction for research.

REFERENCES

[1] Cichosz P. (2000) *Systemy uczące się* [in Eng. Learning systems], WNT, Warszawa, Poland, [in Polish].

[2] Demuth H., Beale M. (1997) *Matlab Neural Network Toolbox*, The Math Works. Inc.

[3] Everitt B.S., Landau S., Leese M. (2001) *Cluster analysis*, London: Arnold; New York: Oxford University Press, UK-USA.

[4] Gordon A.D. (1999) *Classification.* Chapman & Hall, London New York Washington, USA.

[5] Jezierski J. (2018) *Analiza skupień jako zagadnienie wstępne w modelowaniu neuronalnym Towarowej Giełdy Energii Elektrycznej* [in English: Cluster analysis as a preliminary issue in neural modeling of the Polish Power Exchange]. Praca inżynierska napisana w Pracowni Sztucznej Inteligencji Instytutu Informatyki na Wydziale Nauk Ścisłych UPH w Siedlcach pod kierunkiem dr hab. inż. Jerzego Tchórzewskiego, prof. UPH.

[6] Kohonen T. (1997) *Self-organizing Maps*, Springer Verlag, Berlin. Germany.

[7] Koronacki J., Ćwik J. (2008) *Statystyczne systemy uczące się* [in English: Statistical learning systems]. Wydanie drugie. AOW EXIT Andrzej Lang, Warszawa, Poland, [in Polish], p. 327.

[8] Larose D. T. (2006) *Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych* [in English: Discovering knowledge from data. Introduction to data mining]. Wydawnictwo Naukowe PWN, Warszawa [in Polish], p.227.

[9] Morzy T. (2018) *Eksploracja danych. Metody i algorytmy* [in English: Data mining. Methods and algorithms]. WN PWN, Warszawa, Poland, [in Polish].

[10] Osowski S. (2006) *Sieci neuronowe do przetwarzania informacji* [in English: Neural networks for information processing]. OW PW. Warszawa, Poland [in Polish].

[11] Osowski S., Siwek K. (1999) *The self-organizing neural network approach to load forecasting in the power system*. IJCNN'99. Proceedings of International Joint Conference on Neural Networks. Vol. 5, IEEE XPlore Digital Library, pp. 3401-3404.

[12] Strona internetowa TGE S.A. [in English: Website of the Polish Electricity Exchange]. https://tge.pl/ dostęp on-line.

[13] Tadeusiewicz R., Szaleniec M. (2015) *Leksykon sieci neuronowych* [in English: Lexicon of neural networks]. Wydawnictwo Fundacji Projekt Nauka, Wrocław, Poland, [in Polish], p. 134.

[14] Tchórzewski J. (2009) *Identification of the Electrical Energy Stock Exchange and creating knowledge maps using MATLAB environment with SIT and NNT Toolboxes.* Energy Market. EEM 2009 6th International Conference on the European, Leuven, IEEE XPlore Digital Library, pp. 1-6.

[15] Tchórzewski J., Kłopotek M. (2002) *The Concept of Discoveries in Evolving Neural Net.* [in] Advances in Soft Computing, tom 17, Wydawca IPI PAN, Warszawa, Poland [in Polish], str. 165-174.

[16] Tchórzewski J., Korzeniowski M. (2006) *Wykorzystanie sieci neuronowych SOM do prowadzenia odkryć na rynku energii elektrycznej w środowisku MATLABA i Simulinka* [in English: The use of SOM neural networks to conduct discoveries on the electricity market in the MATLABA and Simulinka environment]. Materiały Konferencyjne XI Konferencji nt. Zastosowania Komputerów w Elektrotechnice, Wydawca IEP PP, KE PAN, IEEE PS, str. 51-52.

[17] Tchórzewski J., Buziak R., Suszczyński P. (2005) *Model and Implementation of Self-Organising Neural Network for Searching Discovery in Databases*. Studia Informatica. Systemy i technologie informacyjne. Vol. 1(5)2005, pp. 35-48.

[18] Tchórzewski J., Zarzycki I., Soćko M. (2005) *Poszukiwanie odkryć w rozwijającej się elektroenergetycznej sieci przesyłowej w środowisku MATLABA i Simulinka* [in English: Searching for discoveries in the developing power transmission network in the MATLABA and Simulink environments]. Śląskie Wiadomości Elektryczne. Nr 1. SEP O/Gliwice. [in Polish], pp. 19-28.

[19] Tchórzewski J., Kwiczak I. (2004) *Mapping information from knowledge bases using self-organizing neural network on example of state office*. Warsaw International Seminar Information System. IBS PAN, Warszawa. Poland.

[20] Tchórzewski J., Glejzer K., Szarafin L. (2004) *Zastosowanie sztucznych sieci neuronowych samoorganizujących się do wizualizacji wiedzy* [in English: The use of artificial neural networks self-organizing to visualize knowledge]. Materiały Konferencji Naukowej nt. Sztuczna Inteligencja w Inżynierii Biomedycznej. Wydawnictwa AGH, Kraków. Poland [in Polish].

[21] Tchórzewski J., Kłopotek M., Kujawiak M. (2004) *Comparative Study of Methods for Making Discovery*. Proceedings of Artificial Intelligence Studies. Vol. 1 (23), No 2, pp. 137-150.

[22] Tchórzewski J. (2004) *Wizualizacja wiedzy za pomocą sieci neuronowych SOM w zakresie informatyzacji urzędów państwowych* [in English: Visualization of knowledge using neural network SOM in the field of computerization of state offices]. Badania Operacyjne i Systemowe na Drodze do Społeczeństwa Wiedzy. Problemy Współczesnej Nauki. Teoria i zastosowania. AOW EXIT, str. 113-136.

[23] Trajer J. J., Janaszek-Mańkowska M., Mańkowski D. R. (2916) *Komputerowa analiza danych w badaniach naukowych* [Computer data analysis in scientific research]. Wydawnictwo GGGW, Warszawa, Poland, [in Polish], p. 131.

[24] Wierzchoń S., Kłopotek M. (2015) *Algorytmy analizy skupień* [in English: Cluster analysis algorithms]. Wydawnictwo WNT. Warszawa, Poland, [in Polish], s. 396.

[25] Wierzchoń S., Kłopotek M. (2015) *Algorithms of Cluster Analysis*, Monograph Series. Information Technologies: Research and their Interdisciplinary Applications, IPI PAP, Warsaw, Poland., p.308.

[26] Żurada J., Barski M., Jędruch W. (1996) *Sztuczne sieci neuronowe. Podstawy teorii i zastosowania* [in English: Artificial neural networks. The basics of theory and application]. Wydawnictwo PWN, Warszawa, Poland [in Polish].