UDK 004.8

Olena O. Arsirii[1], Doctor of Technical Sciences, Professor, Head of the Department of Information Systems, E-mail: e.arsiriy@gmail.com, ORCID: https://orcid.org/0000-0001-8130-9613
Olga S. Manikaeva[1], Teaching assistant of Information Systems Department, E-mail: manikaeva@gmail.com, ORCID: http://orcid.org/0000-0002-0631-8883
Anatolii O. Nikolenko[1], Candidate of Technical Sciences, Associate Professor of Information Systems Department, E-mail: anatolyn@ukr.net, ORCID: https://orcid.org/0000-0002-9849-1797
Oksana Yu. Babilunha[1], Candidate of Technical Sciences, Associate Professor of Information Systems Department, E-mail: babilunga.onpu@gmail.com , ORCID: https://orcid.org/0000-0001-6431-3557
[1]Odessa National Polytechnic University, Avenue Shevchenko, 1, Odessa, Ukraine, 65044

Heuristic models and methods for application of the kohonen neural network in the intellectual system of medical-sociological monitoring

Annotation. For increasing the speed of making classification decisions, the possibilities of supervised deep neural networks training are used in systems of medical and sociological monitoring. But the data obtained during the monitoring process are semi-structured and -labelled. This greatly reduces speed and increases supervised machine-training error. The paper proposes heuristic models and methods of using the Kohonen network to increase the training rate on the semi-supervised mode without losing the accuracy level of data classification in intelligent systems of medical and sociological monitoring. More specifically, proposed tools include improved models for the presentation of semi-structured and -labelled medical and sociological monitoring data in the spaces of properties and features due to formalization of procedure of the detailed data aggregation, as well as the semi-supervised training of the Kohonen network using the method of matching existing pre-labelled data with the obtained labels of features clusters. In addition, the method of adjusting weight coefficients in the process of training the Kohonen neural network through the use of a modified genetic algorithm to classify medical and sociological monitoring data was further developed. The proposed modification of the genetic algorithm consists in using the method of matching labels of classes and clusters in assessment the adaptability of chromosomes in the population generated for each example from training sample of medical and sociological monitoring data. It is proposed for evaluating the adaptability of chromosomes to calculate the completeness of the classification as the ratio of the number of found representatives of the positive class to all representatives of the positive class. When performing the selection procedure on the basis of the obtained adaptability score, the two most adapted ("winner" and "vice-winner") chromosomes are selected and crossed, whereas in contrast, the two least adapted ("loser" and "vice-loser") chromosomes are deleted. The crossing is implemented as the exchange of genes between the winner and vice-winner chromosomes, taking into account a randomly selected crossing-over point. Upon mutation, a random number distributed according to a uniform law is added to each chromosome gene, which is an auxiliary method for creating a new chromosome to prevent degeneration of the population. The proposed heuristic models and methods are the basis for the implementation of a separate module as part of the intellectual system of medical and sociological monitoring. Testing the modified intellectual system using well-known test examples from the machine training database and real medical and sociological data showed an increase in the training speed of the Kohonen network with a given level of classification accuracy. Thus, the introduction of the proposed tools allows increasing the efficiency of classification decisions without losing their level of accuracy in intelligent systems of medical and sociological monitoring.

Keywords: Kohonen neural networks; genetic algorithms for training neural networks; medical and sociological monitoring

For citation:Arsirii O.O.., Manikaeva, O.S.., Nikolenko, A.O.. i Babilunha, O.Y.. Heuristic Models and Methods for Application of the Kohonen Neural Network in the Intellectual System of Medical-Sociological Monitoring. Applied Aspects of Information Technology. Publ. Nauka i Tekhnika. Odessa: Ukraine. 2020, Vol.3, No.1, 395–405. DOI:https://doi.org/10.15276/hait.01.2020.6.

Introduction. The computational capabilities of modern computer systems make it possible to collect, accumulate and analyze data on the state of public health in the mode of medical sociological monitoring (MSM). In MSM systems, to increase the efficiency of making classification decisions, the supervised training of deep neural network (DNN) is used. The creation of such a system is associated with the need to automate the intellectual activity of an expert sociologist in the formation of research hypotheses about the composition of the target audience and the adoption of a classification decision about each respondent after processing quantitative metrics or qualitative assessments of socio-demographic, medical, psychophysical, behavioural, geographical, or any other obtained characteristics [1-5].

© Arsirii, O., Manikaeva, O., Nikolenko, A., Babilunha, O., 2020

Analysis of existing research and publications

Studies show that the input data of MSM are usually semi-structured, as they are quantitative and qualitative, collected from different sources, interpreted using different and not always connected scales, and often contradict each other. [6-7] In addition, the process of making classification decisions based on the processing of large arrays of semi-structured data requires long-term work of qualified sociological experts. Therefore, for the input data of MSM, which are periodically updated and replenished, obtaining a sufficient number of such solutions (labels) in manual mode is impossible. Therefore, in practice, MSM data are *semi-labelled* (tagged) [8-9].

To quickly obtain classification decisions in the processing of such data, *semi-supervised training* of

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/deed.uk)

neural networks is used, which, in addition to data labelled for training, also uses a relatively large amount of unlabelled data [9-10]. Studies show that one of the most common ways to implement a semi-supervised machine training is to use Kohonen's self-training (self-organization) neural network with a competing layer on unmarked data to be added to the training set. But in this case, it is necessary to solve the *problem of matching the existing pre-labelled data with the so-called cluster labels* [9-10] obtained after testing the Kohonen self-organizing network. In addition, it is known that the self-organization of neurons in the Kohonen network occurs according to the WTA algorithm (*Winner Takes All*) [10-11], which means that the winner is the neuron with the smallest Euclidean distance between the input vector and the vector of its weight coefficients. And the use of multidimensional and semi-structured MSM data in the formation of the training sample significantly reduces the possibility of quick and successful (in terms of confidence level) training of the Kohonen network using the WTA algorithm with gradient optimization. On the other hand, it is known that when large spaces of admissible values of the parameters of neural networks are studied, while the surface of the training error has a complex relief [11], the usage of *genetic algorithms* with their possibility of random mutation for training neural networks can reduce the negative impact of erroneous decisions the general level of classification reliability in the intellectual system of MSM [12-16].

The purpose and objectives of the research Thus, to solve the problem of *increasing the efficiency of making classification decisions with a given level of reliability* of the intelligent MSM system when processing semi-structured and -labelled data, the development of following models and methods is essential:

1) models for the presentation of semi-structured and semi-labelled MSM data in the spaces of properties and features;

2) method for matching existing pre-labelled data with cluster labels obtained;

3) method for adjusting weight coefficients during the training of the Kohonen neural network using the modified genetic algorithm.

The main research material

1. In previous studies [17-18], the authors described in detail the results of improving the model for representing semi-structured and – labelled MSM data in the spaces of properties and features, taking into account the values, types, formats, sources, quality assessments and aggregation procedures for detailed data depending on the type of property, which allowed to formalize the process of creating educational and test samples to automate the adoption of classification decisions. We briefly present the following results.

Detailed data D_{R_i} of MSM of R_i respondents of the TA target audience in the m-dimensional space of properties is proposed to be set using the following model:

$$D_{R_i} = \langle \{V_\varphi\}_j, T_j, F_j, S_j, Q_j, Mt_j \rangle, j = 1, m, \qquad (1)$$

where: $\{V_\varphi\}_j, T_j, F_j, S_j, Q_j, Mt_j$ is the set of discrete (qualitative or quantitative) values: type, format, source, quality assessment, procedure for transforming data of the *j*-th property, and the number of respondent's properties – *m* is significantly less than the number of respondents analyzed – *n* ($m \ll n$).

In a matrix form, the training sample of MSM data is a set of aggregated data of R_i respondents of the TA target audience and consists of *labelled training set* and *unlabelled training set* $A_R^{train} = A_R^l \cup A_R^u$:

$$A_R^l = \langle X, T \rangle = \langle \begin{bmatrix} x_{11} & \dots & x_{1q} \\ \dots & \dots & \dots \\ x_{l1} & \dots & x_{lq} \end{bmatrix}, \begin{bmatrix} t_1^1 \\ \dots \\ t_l^p \end{bmatrix} \rangle,$$
$$A_R^u = \langle X \rangle = \langle \begin{bmatrix} x_{l+11} & \dots & x_{l+1q} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nq} \end{bmatrix} \rangle, \qquad (2)$$

where: the matrix X is a set of characteristics of the R_i respondents in the *q*-dimensional space of characteristics, and the vector T is the finite set of numbers of classes (names, labels, markers).

It is shown that the creation of the training sample A_{R_i} is affected by the subjective decisions of the respondents, which depend on a hypothesis previously advanced by an expert sociologist about the power of many classes $|T|$, and/or the decision of the expert sociologist about the respondent's attitude to a particular class. It has been established that the process of making such decisions is ambiguous and time-consuming, requires the intellectual work of an expert sociologist and depends on the one's qualifications. Then, in the MSM intelligent system for constantly updated data, when constructing the rule $a: X \to T$ in the automated mode for making classification decisions, it is not possible to obtain a sufficient amount of labelled data A_R^l.

To classify MSM data with semi-supervised training, the authors propose first to conduct a cluster analysis using the Kohonen neural network, which solves the problem of partitioning the training

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

sample A_R^{train} of MSM data into non-intersecting clusters from the Z set.

Then the resulting set of clustered data $A_R^{train^*}$ has the form:

$$A_R^{train^*} = \langle X, Z \rangle \langle \begin{bmatrix} x_{11} & \dots & x_{1q} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nq} \end{bmatrix}, \begin{bmatrix} z_1^1 \\ \dots \\ z_n^p \end{bmatrix} \rangle. \quad (3)$$

2. In a previous study by the authors [19-23], it has been shown in detail that the improvement of the Kohonen network semi-supervised teaching method due to the proposed method of matching class and clusters labels made it possible to obtain additional labels when creating a training and test sample to increase the efficiency of further data classification medical and sociological monitoring without reducing its confidence level.

Given the models for the presentation of semi-structured and -labelled MSM data (1)-(3), it is proposed to automate the training process of the Kohonen network using semi-labelled A_R^{train} data using a method that includes the following five steps: *initialization* of the Kohonen network, *self-organization* of competing neurons layer using the WTA algorithm, *obtaining labels of clusters Z* as elements of the original vector $A_R^{train^*}$, *matching* existing *class labels* with the labelled part of the training set A_R^l and with the obtained *labels of clusters*, as well as the final *labelling of the aggregated data* of the training set A_R^{train}. Here is a more detailed description of the method of matching class and clusters labels, which requires 4-steps:

Step 1. For all examples i of the marked part A_R^l of the training set, we form a two-dimensional histogram D_{mk} (square matrix $m, k = \overline{1, p}$) of pairwise coincidences of the values of the existing labels of classes T and the obtained labels of the clusters Z, $D_{mk} = D_{mk} + 1$, if $z_i^m = t_i^k$, where $m, k = \overline{1, p}$ (Fig. 1).

Step 2. In cycles (row and column), we assign the value of zeros to the elements of the D_{mk} histogram, leaving only those that satisfy the rule:

$$D_{mk}^* = \begin{cases} 0 \\ D_{mk}, \left\langle \begin{matrix} \max_m(D_{mk}) = \\ = \max_k(D_{mk}) \end{matrix} \middle| m = k \middle| \max(D_{mk}) <> 0 \right\rangle \end{cases}. \quad (4)$$

The correction of the D_{mk} histogram is performed until only one non-zero value remains in each row and column.

Step 3. We turn the adjusted D_{mk}^* histogram into the correspondence matrix of the values of class and clusters labels $M_{(1,2)m}$ for all examples of the training sample A_R^l (Fig. 1).

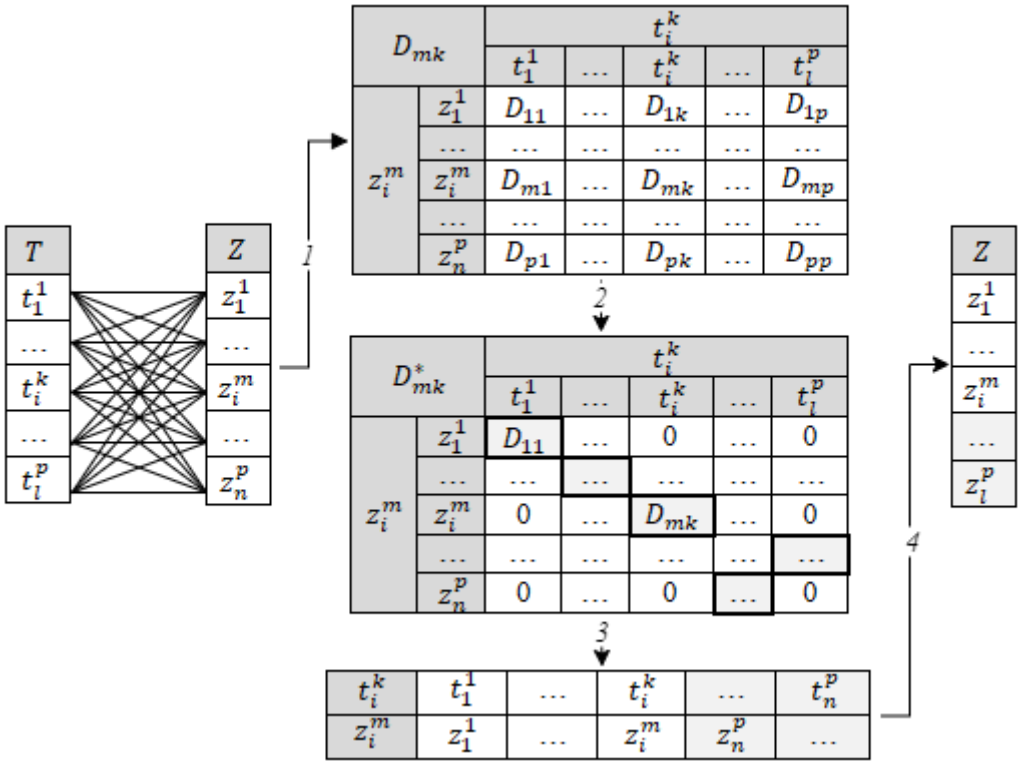


Fig. 1. Results of the steps of the method for matching class and clusters labels

Step .4. We correct the vector of cluster labels Z (3) using the correspondence matrix $M_{(1,2)m}$ and make a classification solution for the labelled part of the training sample in the form:

$$A_R^{l\,*} = \langle X, T, Z \rangle \langle \begin{bmatrix} x_{11} & \dots & x_{1q} \\ \dots & \dots & \dots \\ x_{l1} & \dots & x_{lq} \end{bmatrix}, \begin{bmatrix} t_1^1 \\ \dots \\ t_l^p \end{bmatrix}, \begin{bmatrix} z_1^1 \\ \dots \\ z_l^p \end{bmatrix} \rangle \quad (5)$$

and pass it on to the sociologist expert for further evaluation and interpretation.

The testing of the Kohonen network training method using the WTA algorithm with gradient optimization in the training samples of semi-structured and -labelled MSM data showed that the automated decision-making of classification decisions was more efficient than the manual decision of an expert sociologist, but the classification accuracy, especially in cases of similar classes, is less than 90 %.

Therefore, to increase the level of classification reliability, it is proposed to *use genetic algorithms* with the inherent possibility of random mutation for training the Kohonen network.

3. Development of a method for adjusting weight coefficients using a modified genetic algorithm. To ensure a given level of confidence of the classification of MSM data, it is proposed to improve the *self-organization stage* of the Kohonen network by using instead of the classical iterative WTA algorithm a method of matching class and clusters labels to assess the adaptability of the chromosome in the population of the genetic algorithm for each example with the training sample $\boldsymbol{A_R^{train}}$ in the form (5) (Fig. 2)

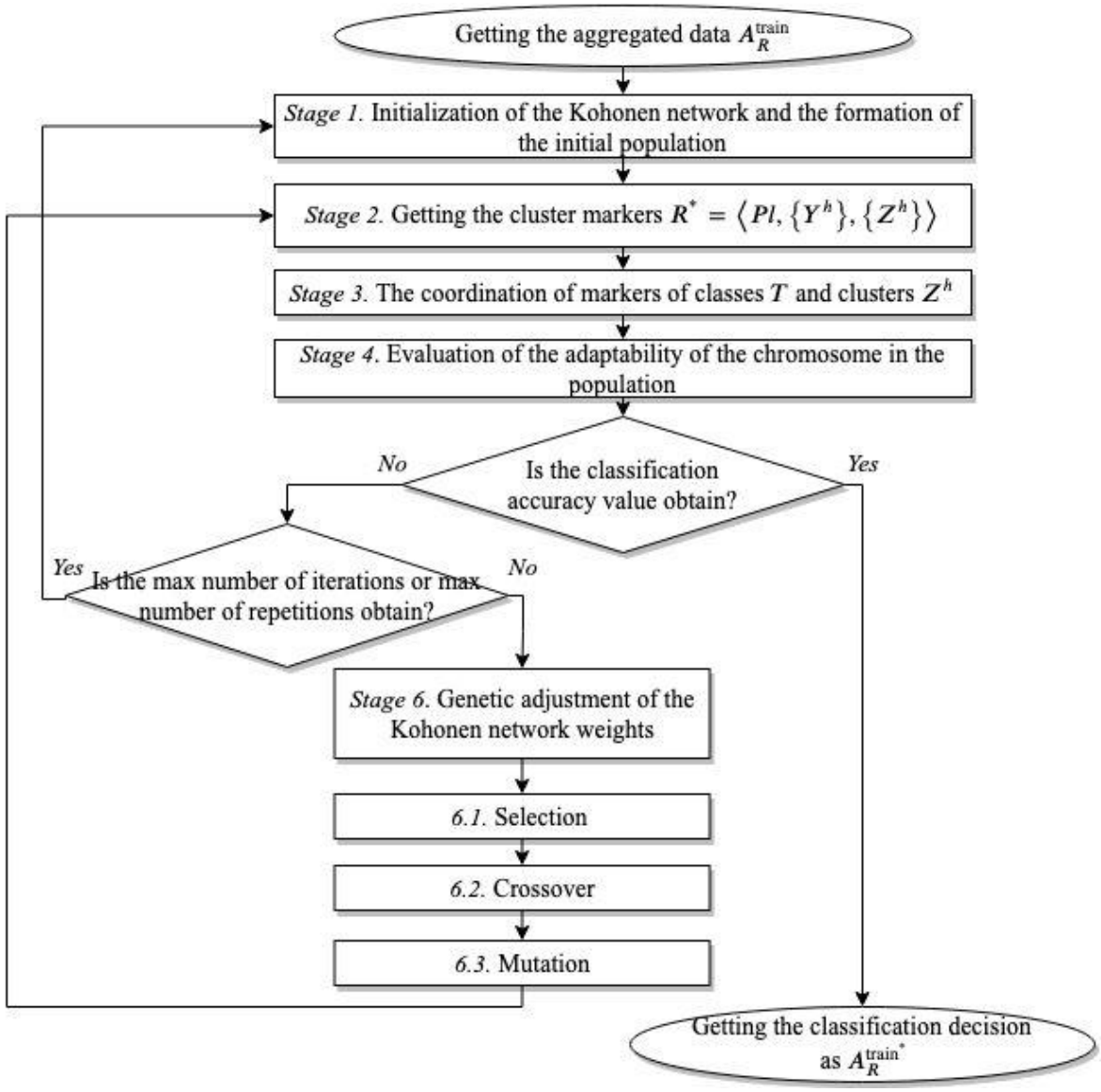


Fig. 2. Method for adjusting weight coefficients in the process of training the Kohonen neural network

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

Stage 1. Initialization of the Kohonen network and the formation of the initial population. In accordance with the dimensions of the features space q and the power of the set of classes p, a competing layer of the Kohonen network is created (Fig. 3) and, taking into account the values of A_R^{train} (2), the matrix of weight coefficients W of the neurons of the competing layer and their displacements B (6).

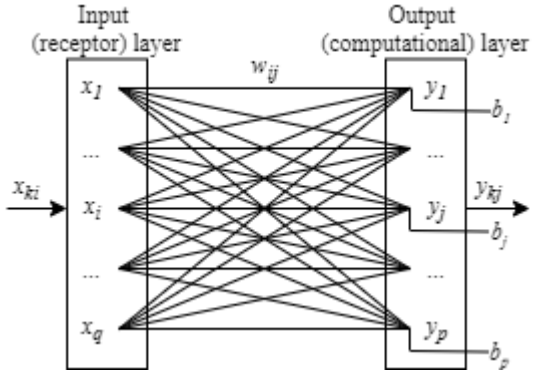


Fig. 3. Structural model of the Kohonen network

$$\langle W, B\rangle = \langle \begin{bmatrix} w_{11} & \dots & w_{1q} \\ \dots & \dots & \dots \\ w_{p1} & \dots & w_{pq} \end{bmatrix}, \begin{bmatrix} b_1 \\ \dots \\ b_p \end{bmatrix} \rangle, \quad (6)$$

where: $i = \overline{1,q}$, $j = \overline{1,p}$.

The formation of the initial Pl population consists in the creation of chromosomes in the number of N_p.

Each chromosome $CH_h \epsilon$ is a vector of random real numbers, which corresponds to individual weighting factors and displacements: $CH_h = < W, B >^h$, $h = \overline{1,N_p}$:

$$CH_h = \{w_{11}^h \dots, w_{1q}^h, \dots, w_{p1}^h, \dots, w_{pq}^h, b_1^h, \dots, b_p^h\} \quad (7)$$

If we represent CH_hy as the set of genes $\{g_1, \dots, g_G\}$, where $G = p \times q + p$ is the power of CH_h, then the population Pl looks as follows:

$$Pl = \begin{bmatrix} CH_1 \\ \cdots \\ CH_{N_p} \end{bmatrix} = \begin{bmatrix} g_{11} & \dots & g_{1G} \\ \dots & \dots & \dots \\ g_{N_p 1} & \dots & g_{N_p G} \end{bmatrix}. \quad (8)$$

Stage 2. Getting the cluster markers (labels) $R^* = \langle Pl, \{Y^h\}, \{Z^h\}\rangle$. For solutions in the form of cluster labels, the vectors $\{x_{ki}\}$ from the training sample A_R^{train} are sequentially fed to the input of the competing layer in the form of the population Pl (8), and the response as binary vector $\{y_{kj}\}^h, k = \overline{1,n}, j = \overline{1,p}, h = \overline{1,N_p}$, is calculated for each chromosome CH_h. The result of the stage is the output vector $\{z_{kj}\}^h$, and the k-element value of which is assigned the sequence number of the winner-neuron j of the binary vector $\{y_{kj}\}^h$.

Thus, with respect to each chromosome $CH_h h = \overline{1,N_p}$, for a population Pl we obtain a set of preliminary solutions in the following form:

$$R^* = \langle Pl, \{Y^h\}, \{Z^h\}\rangle = \\ = \langle \begin{bmatrix} x_{11} & \dots & x_{1q} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nq} \end{bmatrix}, \begin{bmatrix} y_{11} & \dots & y_{1p} \\ \dots & \dots & \dots \\ y_{n1} & \dots & y_{np} \end{bmatrix}^h, \begin{bmatrix} z_{11} \\ \dots \\ z_{np} \end{bmatrix}^h \rangle, \quad (9)$$

where: X is the set of characteristics of the respondents R_i, Y^h is the binary response matrix of the competing layer, Z^h is the set of cluster labels.

Stage 3. The coordination (matching) of markers (labels) of classes T and clusters Z^h is performed for each chromosome $CH_h h = \overline{1,N_p}$ according to the developed method (Fig. 1).

Then we get:

$$R^{**} = \langle Pl, X, T, \{Z^h\}\rangle = \\ = \langle \begin{bmatrix} x_{11} & \dots & x_{1q} \\ \dots & \dots & \dots \\ x_{l1} & \dots & x_{lq} \end{bmatrix}, \begin{bmatrix} t_1^1 \\ \dots \\ t_l^p \end{bmatrix}, \begin{bmatrix} z_1^1 \\ \dots \\ z_l^p \end{bmatrix}^h \rangle \quad (9)$$

Stage 4. Evaluation of the adaptability of the chromosome in the population. In a cycle, based on the existing labels of the classes T and the matched markers of the clusters Z^h, for each chromosome CH_h we calculate the accuracy of the classification (*true positive rate TPR*):

$$TPR^h = \frac{TP}{TP+FN} \cdot 100\ \%, \quad (10)$$

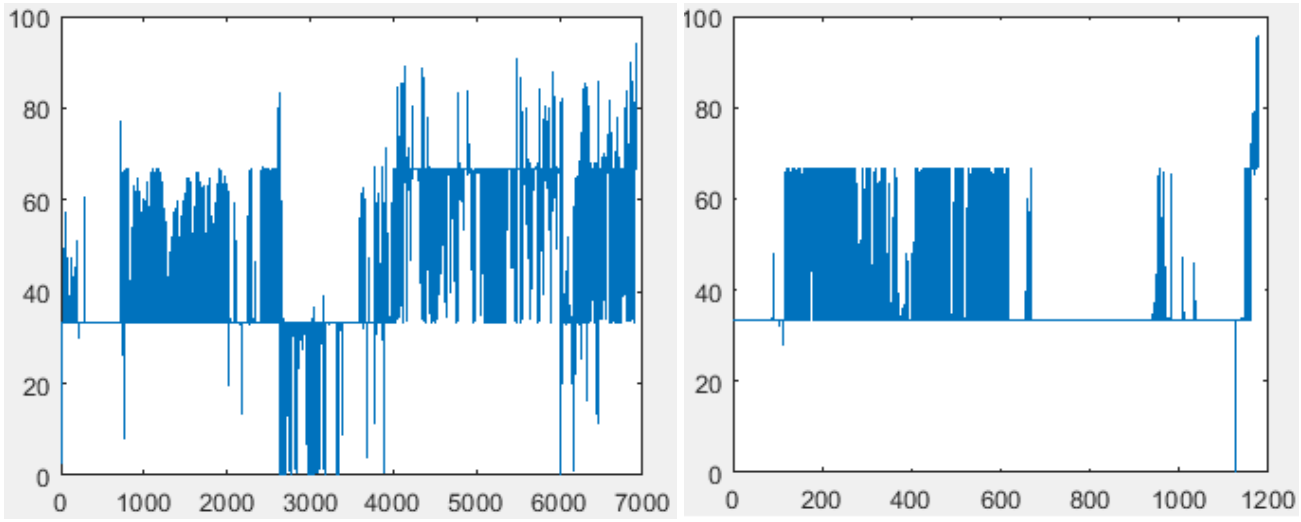
where: *TP* are true positive solutions; *FN* – positive solutions, classified as negative (Type II expert error). [24-25]

Stage 5. Checking the algorithm stop criteria. The stopping criteria of the algorithm are:

– successful completion of training involves obtaining the value of classification accuracy and issuing of the classification decision;

– unsuccessful completion of training is the exceeding the maximum number of iterations or the maximum number of repetitions of the value of "unsatisfactory" classification accuracy on the differing number of the training epochs – 7000 epochs (Fig. 4a) and 1200 epochs (Fig. 4b), which means degeneration of the population – the transition to re-initialization of Kohonen network is performed.

If the stopping criteria are not met, the training is not completed and we proceed to Stage 6.



a b

Fig. 4. Examples of population degeneration (a – 7000 training epochs, b – training 1200 epochs; along the X and Y axes there are the number of the training epoch and TPR values respectively)

Stage 6. Genetic adjustment of the Kohonen network weights consists of performing classic *selection*, *crossover*, and *mutation* procedures. When performing the selection, the population is sorted in descending order of estimation of the classification accuracy TPR^h obtained for each chromosome CH_h. For further crossing, we determine the *winner* CH_1^w and the *vice winner* CH_2^{vw} by the criteria of classification accuracy, and $CH_{N_p-1}^{vl}$ (*vice loser*) and $CH_{N_p}^l$ (*loser*) for subsequent removal last ones from the population. To perform the crossing operation, a crossover point is determining which randomly divides the chromosome into two parts and provides the exchange of the chromosome genes. In our case, the new chromosome $CH_{N_p}^l$ (*loser*) will consist of the first part of the genes CH_1^w and the second part of the CH_2^w chromosome genes, and the new chromosome $CH_{N_p-1}^{vl}$, vice versa, of the second part of the genes CH_1^w and the first part of the genes CH_2^w. (Fig. 5). Mutation procedure is an auxiliary method for creating a new chromosome. To each gene of chromosome CH_h there is added a random number distributed according to a uniform law:

$$CH_h^{new} = CH_h + \\ +rand(-0{,}05\max(CH_h), +0{,}05\max(CH_h)). \quad (11)$$

Thus, a new Pl^{new} population is formed, which then passed as input data to perform the actions of Stage 2.

The results of testing the proposed method for adjusting weighting coefficients using a modified genetic algorithm on a training set, generated from the data of the classical “Fisher's Iris” problem, are shown in Fig. 6. It is known that classes 2 and 3 intersect, and the reliability of their classification exceeds 95 % and 90 %.

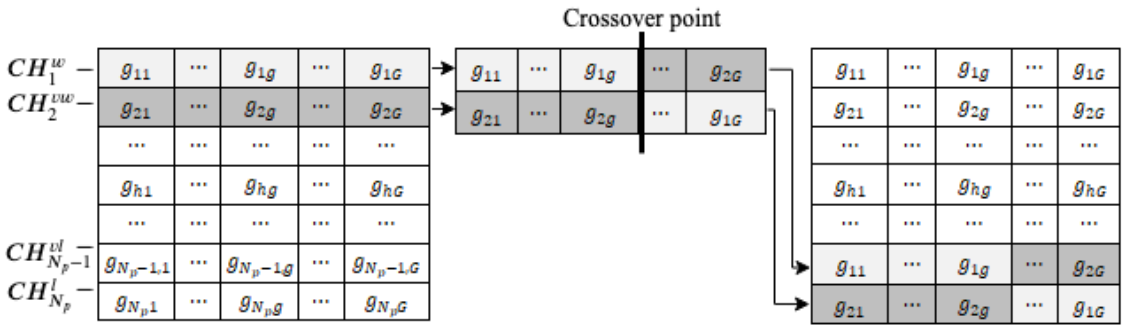
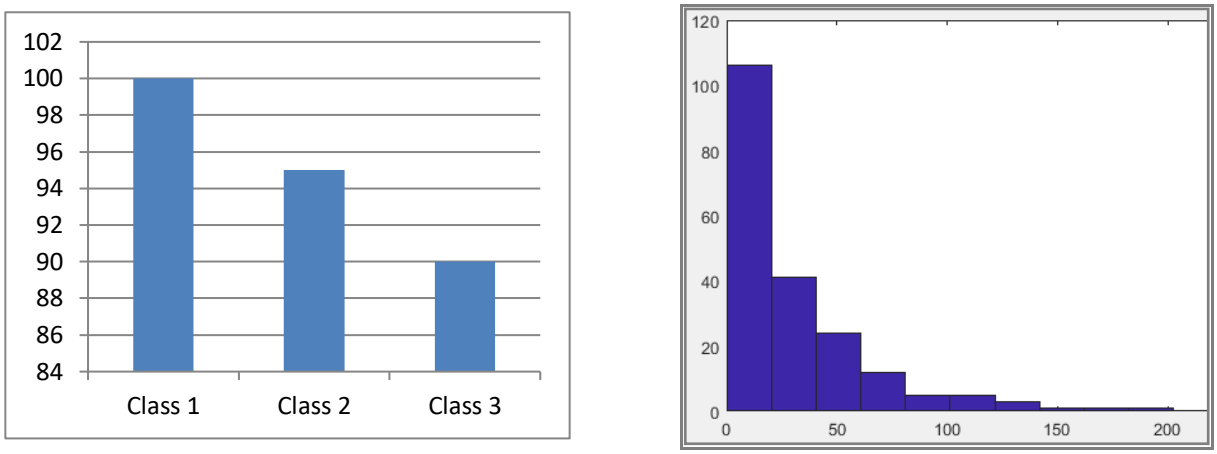


Fig. 5. The crossing operator procedure

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)



а б

Fig. 6. The result of testing the method for adjusting the weight coefficients of the Kohonen network using a modified genetic algorithm
(a – TPR; b – training diagram)

Table 1 shows the comparative results on the efficiency and accuracy of making classification decisions according to the two proposed methods for adjusting the weight coefficients of the Kohonen network in the form of modified WTA algorithms with gradient optimization and a modified genetic algorithm (GA). When implementing these algorithms, the method of matching existing pre-labelled data with the obtained cluster labels was used.

Table 1 Efficiency of training the Kohonen network and the accuracy of classification

Efficiency of training and the accuracy of classification		WTA	GA
The accuracy of classification	Training	88,96 %	94,99 %
	Testing	81,08 %	86,06 %
The number of iterations in the training of each feature		2650	350
Time of working	Training	00:25:41	00:21:27
	Testing	00:17:22	00:15:13

Fig. 7 shows the results of calculations of the classification accuracy according to the data of medical and sociological monitoring, obtained in a study of 1143 respondents on 114 factors within the topic “Ukraine – lifestyle” [26]

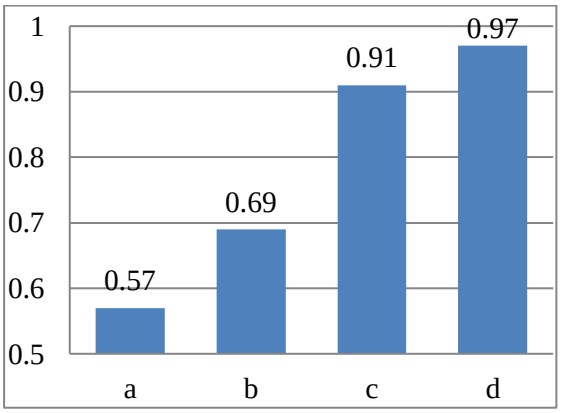


Fig. 7. The accuracy of the classification

The results indicate an increase in the classification reliability in the automated mode by 57 % (a) as compared to the manual mode, according to the respondent’s own decision – 69 % (b) and according to the decision of an expert sociologist, as well as with the automated mode based on the Kohonen network training using the WTA algorithm by 91 % (c), and according to the modified genetic algorithm by 97 % (d).

Conclusions. The heuristic models and methods proposed in the study were used to develop the intellectual MSM system, which consists of four subsystems: *generating research hypotheses* on the composition of the target audience (social environment), *preparing and storing* MSM data, *making classification decisions* for confirmation/refutation the research hypotheses, and *interpretation/assessment* of the classification results obtained. [17-23; 26].

The basis for the development of subsystems for the *preparation* and *storage* of MSM data includes models for the presentation of semi-structured and -labelled MSM data in the spaces of properties and features, taking into account the values, types, formats, sources, quality assessments and aggregation procedures as applied to the features of detailed data, which allowed to formalize the process of creating an training and testing samples for further use of the Kohonen network.

The creation of a classification *decision making subsystem* for confirming or refuting hypotheses for studying MSM data was based on machine training of the Kohonen network, which was implemented using two developed methods: matching existing pre-labelled data with obtained cluster labels and adjusting weight coefficients in the process of training the Kohonen neural network by using modified genetic algorithm.

The research results indicate that the use of the heuristic method of adjusting weight coefficients allows increasing the learning speed of the Kohonen network by 16-20 % while maintaining the specified level of classification accuracy in the context of the data from machine training archives, depending on the power of the sample. At the same time, the classification accuracy level increased by 6 % in the training sample and by 5 % in the test one [27]. Testing on MSM data showed that the use of the heuristic method for adjusting the weight coefficients of the Kohonen network using a modified genetic algorithm with a training sample of 1200 examples allows increasing the learning speed by 16 % and the classification accuracy by 6 % compared to using the modified WTA algorithm.

References

1. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). “CRISP-DM 1.0: Step- by-Step Data Mining Guide”. *SPSS*, Copenhagen.

2. Merkert, J., Mueller, M. & Hubl, M. A (2015). “Survey of the Application of Machine Learning in Decision Support Systems”, *Twenty-Third European Conference on Information Systems (ECIS 2015)*, Münster, Germany, pp. 1-15.

3. Kislova, O. N. (2005). “Intellektual'nyj analiz dannyh: vozmozhnosti i perspektivy ispol'zovanija v sociologicheskih issledovanijah” [Intellectual data analysis: Opportunities and prospects for use in sociological studies]. *Methodology, theory and practice of sociological analysis of modern society: Collection of scientific works*. Kharkiv, Ukraine, pp. 237-243 (in Russian).

4. Praveen, S. & Chandra, U. (2017). “Influence of Structured, SemiStructured, Unstructured data on various data Models”, *International Journal of Scientific & Engineering Research*, Vol. 8, Issue 12, pp. 67-69.

5. Corbetta, P. (2011). “Social Research: Theory, Methods and Techniques”, London: *Sage*, 328 p.

6. Semenov, V. E. (2009). “Analiz i interpretatsiya dannyih v sotsiologii: uchebnoe posobie”. [Analysis and interpretation of data in sociology: a training manual], Vladimir state. Un-t. Vladimir, Russian Federation, *Publishing House of VlSU*, 131 p. ISBN 978-5-89368-916-7 (in Russian).

7. Efimenko, S. A. (2019). “Mediko - sotsiologicheskiy monitoring kak instrument sovremennyih tehnologiy v upravlenii zdorovem patsientov”. [Medical and sociological monitoring as an instrument of modern technologies in managing patient health], *Internet conference Health: problems of organization, management and levels of responsibility* [Electronic resource]. – Available at: http://ecsocman.hse.ru/text/16207043/. – Active link 01.03.2020 (in Russian).

8. Triguero, Isaac. (2013). “Self-labeled Techniques for Semi-supervised Learning: Taxonomy, Software and Empirical Study”, *Knowledge and Information Systems: journal*, Vol. 42, No. 2, pp. 245-284. ISSN 0219-1377. DOI:10.1007/s10115-013-0706-y.

9. Fazakis, Nikos. (2015). “Self-Trained LMT for Semisupervised Learning. Computational” *Intelligence and Neuroscience 2016*, Vol. 2016, pp. 1-13, doi:10.1155/2016/3057481.

10. Kohonen, T. (2008). “Samoorganizuyu-shhiesya karty”. [Self-organizing maps]: trans. from English 3rd ed. Moscow, Russian Federation, *BINOM Laboratory of Knowledge*, 655 p. (in Russian).

11. Khajkin, S. (2006). “Nejronny`e seti: polny`j kurs”: 2-e izd., ispr. [Neural networks: the full course]: trans. from English: 2nd ed., Rev-Moscow, Russian Federation, *LLC “I. D. Williams”*, 1104 p. (in Russian).

12. Holland, J. H. (1975). “Adaptation in Natural and Artificial Systems: An Introductory Analysis with Application to Biology, Control and Artificial Intelligence”, *University of Michigan Press*, pp. 439-444.

13. Holland, J. H. (1992). “Genetic Algorithms”, *Scientific American*, pp.114-116.

14. Rutkovskaya, D., Pilin`skij, M., Rutkovskij, L. (2004). “Nejronny`e seti, geneticheskie algoritmy` i nechetkie sistemy”. [Neural networks, genetic algorithms and fuzzy

systems]. Moscow, Russian Federation, *Hot line–Telecom*, 452, 15 p.

15. Ettaouil M., Ghanou Y., Elmoutaouakil K. & Lazaar M. (2011). "A New Architecture Optimization Model for the Kohonen Networks and Clustering". *Journal of Advanced Research in Computer Science (JARCS*), Volume 3, Issue1, pp. 14-32.

16. Chambers, L. (2000). "The Practical Handbook of Genetic Algorithms Applications. 2nd ed." Chapman & Hall/CRC, 544 p.

17. Arsirii, O. & Manikaeva, O. (2019) "Models and Methods of Intellectual Analysis for Medical-Sociological Monitoring Data Based on the Neural Network with a Competitive Layer", *Applied Aspects of Information Technology,* Vol. 2, No. 3, pp 173-185. DOI: 10.15276/aait.03.2019.1.

18. Arsirii, O., Babilunha, O., Manikaeva, O. & Rudenko, O. (2018). "Automation of the Preparation Process Weakly-Structured Multi-Dimensional Data of Sociological Surveys in the Data Mining System", *Herald of Advanced Information Technology,* Vol. 1, No. 1, pp. 11-20. DOI: 10.15276/hait.01.2018.1.

19. Arsiri, O., Antoshchuk, S., Babilunha, O., Manikaeva, O. & Nikolenko, A. (2019). "Intellectual Information Technology of Analysis of Weakly-Structured Multi-Dimensional Data of Sociological Research", *International Scientific Conference "Intellectual Systems of Decision Making and Problem of Computational Intelligence" ISDMCI 2019*, Lecture Notes in Computational Intelligence and Decision Making, pp. 242-258.

20. Arsirij, E. A. Antoshhuk, S. G. & Manikaeva O. S. (2015). "Avtomatizacziya predstavleniya i izvlecheniya deklarativnykh znanij v czifrovom makete predpriyatiya pri iniczializaczii proektov okhrany` truda". [Automation of the presentation and extraction of declarative knowledge in a digital layout of an enterprise during the initialization of labor protection projects]. *Electrotechnical and computer systems*, 19 (95), pp. 266-271 (in Russian).

21. Manikaeva, O. S. Arsirii E. A. & Vasilevskaya A. P. (2015). "Mashinnoe obuchenie samoorganizuyushhegosya sloya Kokhonena v sistemakh nejrosetevogo raspoznavaniya obrazov po statisticheskoj informaczii". [Machine learning of a self-organizing Kohonen layer in neural network pattern recognition systems based on statistical information], *Systems Technology. Regional interuniversity collection of scientific works*, VIP. 5(100), Dnipropetrovsk, Ukraine, pp. 93-105 (in Russian).

22. Arsirii, E. A., Manikaeva O. S. & Vasilevskaya A. P. (2015). "Razrabotka podsistemy` podderzhki prinyatiya reshenij v sistemakh nejrosetevogo raspoznavaniya obrazov po statisticheskoj informaczii". [Development of the subsystem of decision support in the systems of neural network image recognition by statistical information], *Eastern-European Journal of Enterprise Technologies*, No. 6/4(78), pp. 4-12 (in Russian).

23. Manikaeva, O. & Federico Barber Sanchís (2017). "Development of Optimization Method to train Kohonen Neural Network based on Genetic Algorithm", *V Ukrainian-German "Conference. Informatics. Culture. Technology"*. Odessa, Ukraine, pp. 13-14.

24. Powers, David M W (2011)."Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation", *Journal of Machine Learning Technologies.* 2 (1), pp.37-63.

25. Fawcett, Tom (2006). "An Introduction to ROC Analysis", *Pattern Recognition Letters.* 27 (8): pp. 861-874. DOI:10.1016/j.patrec.2005. 10.010.

26. The data of the sociological research "Ukraine-style of life" [Digital resource]. – Available at: http://edukacjainauka.pl/ limesurvey/index.php/lang-pl. – Active link 01.03.20.

27. Center for Machine Learning and Intelligent Systems [Digital resource] – Available at: http://archive.ics.uci.edu/ml/datasets.html limesurvey/index.php/lang-pl. – Active link 01.03.20.

Received 31.01.2020
Received after revision 18.02. 2020
Accepted 20.02.2020

УДК 004.8

[1]**Арсірій, Олена Олександрівна,** доктор техніч. наук, професор, зав. кафедрою інформаційних систем, E-mail: e.arsiriy@gmail.com, ORCID: https://orcid.org/0000-0001-8130-9613
[1]**Манікаєва, Ольга Сергіївна,** асистент кафедри інформаційних систем,
E-mail:manikaeva@gmail.com, ORCID: http://orcid.org/0000-0002-0631-8883
[1]**Ніколенко, Анатолій Олександрович,** кандидат техніч. наук, доцент кафедри інформаційних систем, E-mail: anatolyn@ukr.net, ORCID: https://orcid.org/0000-0002-9849-1797
[1]**Бабілунга, Оксана Юрьєвна,** кандидат техніч. наук, доцент кафедри інформаційних систем,

E-mail: babilunga.onpu@gmail.com, ORCID: https://orcid.org/0000-0001-6431-3557
[1]Одеський національний політехнічний університет, пр. Шевченка,1, м. Одеса, Україна, 65044

ЕВРІСТИЧНІ МОДЕЛІ ТА МЕТОДИ ЗАСТОСУВАННЯ НЕЙРОННОЇ МЕРЕЖІ КОХОНЕНА В ІНТЕЛЕКТУАЛЬНІЙ СИСТЕМІ МЕДИКО-СОЦІОЛОГІЧНОГО МОНІТОРИНГУ

Анотація. *В системах медико-соціологічного моніторингу для збільшення оперативності прийняття класифікаційних рішень використовують можливості навчання глибинних нейронних зі вчителем. Але данні, що накопичуються в процесі моніторингу є слабко структурованими та маркованими. Це значно знижує швидкість та підвищує помилку машинного навчання зі вчителем. В роботі запропоновані евристичні моделі та методи застосування мережі Кохонена для збільшення швидкості навчання з частковим залученням вчителя без втрати рівня достовірності класифікації даних в інтелектуальних системах медико-соціологічного моніторингу. А саме удосконалені моделі представлення слабко структурованих та маркованих даних медико-соціологічного моніторингу в просторах властивостей та ознак за рахунок формалізації процедури агрегації деталізованих даних, а також метод навчання мережі Кохонена з частковим залученням вчителя з використанням методу узгодження наявних попередньо маркованих даних із отриманими маркерами кластерів ознак. Крім того отримав подальший розвиток метод корегування вагових коефіцієнтів в процесі навчання нейронної мережі Кохонена за рахунок використання модифікованого генетичного алгоритму для класифікації даних медико-соціологічного моніторингу. Запропонована модифікація генетичного алгоритму полягає у використанні методу узгодження маркерів класів та кластерів при розрахуванні оцінки пристосованості хромосом в популяції, яка генерується для кожного прикладу із навчальної вибірки даних медико-соціологічного моніторингу. Для оцінки пристосованості хромосом запропоновано розраховувати повноту класифікації, як відношення кількості представників позитивного класу, яких знайдено до усіх представників позитивного класу. При виконанні процедури селекції на основі отриманої оцінки пристосованості обирається по дві найбільш («переможець» та «віце-переможець») пристосовані хромосоми, які схрещуються та найменш («лузер» та «віце-лузер») пристосовані хромосоми які видаляються. Схрещування реалізовано як обмін генами між хромосомами «переможець» та «віце-переможець» з врахуванням випадково обраної точки кросинговеру. При мутації до кожного гену хромосоми додається випадкове число, розподілене за рівномірним законом, що є допоміжним методом для створення нової хромосоми для запобігання виродження популяції. Запропоновані евристичні моделі та методи є основою для реалізації окремого модуля у складі інтелектуальної системи медико-соціологічного моніторингу. Апробація модифікованої інтелектуальної системи на відомих тестових прикладах із бази даних машинного навчання та реальних медико-соціологічних даних показала підвищення швидкості навчання мережі Кохонена із забезпеченням заданого рівня достовірності класифікації.*

Ключові слова: *нейронні мережі Кохонена; генетичні алгоритми для навчання нейронних мереж; медико-соціологічний моніторинг*

УДК 004.8

[1]**Арсирий, Елена Александровна,** доктор технич. наук, профессор, зав. кафедрой информационных систем, E-mail: e.arsiriy@gmail.com, ORCID: https://orcid.org/0000-0001-8130-9613
[1]**Маникаева, Ольга Сергеевна,** ассистент кафедры информационных систем, E-mail:manikaeva@gmail.com, ORCID: http://orcid.org/0000-0002-0631-8883
[1]**Николенко, Анатолий Александрович,** кандидат технич. наук, доцент кафедры информационных систем, E-mail: anatolyn@ukr.net, ORCID: https://orcid.org/0000-0002-9849-1797
[1]**Бабилунга, Оксана Юрьевна,** кандидат технич. наук, доцент кафедры информационных систем, E-mail: babilunga.onpu@gmail.com, ORCID: https://orcid.org/0000-0001-6431-3557
[1]Одесский национальный политехнический университет, пр. Шевченко, 1, г. Одесса, Украина, 65044

ЭВРИСТИЧЕСКИЕ МОДЕЛИ И МЕТОДЫ ПРИМЕНЕНИЯ НЕЙРОННОЙ СЕТИ КОХОНЕНА В ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЕ МЕДИКО-СОЦИОЛОГИЧЕСКОГО МОНИТОРИНГА

Аннотация. *В системах медико-социологического мониторинга для увеличения оперативности принятия классификационных решений используют возможности обучения глубинных нейронных с учителем. Но данные, которые накапливаются, в процессе мониторинга являются слабо структурированными и маркированными. Это значительно снижает скорость и повышает ошибку машинного обучения с учителем. В работе предложены эвристические модели и методы применения сети Кохонена для увеличения скорости обучения с частичным привлечением учителя без потери уровня достоверности классификации данных в интеллектуальных системах медико-социологического мониторинга. А именно усовершенствованные модели представления слабо структурированных и маркированных данных медико-социологического мониторинга в пространствах свойств и признаков за счет формализации процедуры агрегации детализированных данных, а также метод обучения сети Кохонена с частичным привлечением учителя с использованием метода согласования имеющихся предварительно маркированных данных с полученными маркерами кластеров признаков.*

Кроме того получил дальнейшее развитие метод корректировки весовых коэффициентов в процессе обучения нейронной сети Кохонена за счет использования модифицированного генетического алгоритма для классификации данных медико-социологического мониторинга. Предложенная модификация генетического алгоритма заключается в использовании метода согласования маркеров классов и кластеров при расчете оценки приспособленности хромосом в популяции, которая генерируется для каждого примера с обучающей выборки данных медико-социологического мониторинга. Для оценки приспособленности хромосом предложено рассчитывать полноту классификации, как отношение числа представителей положительного класса, которые найдено ко всем представителям положительного класса. При выполнении процедуры селекции на основе полученной оценки приспособленности избирается по две наиболее («победитель» и «вице-победитель») приспособленные хромосомы, которые скрещиваются и наименее («лузер» и «вице-лузер») приспособленные хромосомы которые удаляются. Скрещивание реализовано как обмен генами между хромосомами «победитель» и «вице-победитель» с учетом случайно выбранной точки кроссинговера. При мутации к каждому гену хромосомы добавляется случайное число, распределенное по равномерному закону, что является вспомогательным методом для создания новой хромосомы для предотвращения вырождения популяции. Предложенные эвристические модели и методы являются основой для реализации отдельного модуля в составе интеллектуальной системы медико-социологического мониторинга. Апробация модифицированной интеллектуальной системы на известных тестовых примерах, из базы данных машинного обучения и реальных медико-социологических данных, показала повышение скорости обучения сети Кохонена с обеспечением заданного уровня достоверности классификации.

Ключевые слова: *нейронные сети Кохонена; генетические алгоритмы для обучения нейронных сетей; медико-социологический мониторинг*



Olena O. Arsirii, Doctor of Technical Sciences, Professor, Head of the Department of Information Systems
Research field: Information technology, decision support systems, machine learning, neural networks



Olga S. Manikaeva, Teaching assistant of Information Systems Department
Research field: artificial intelligence methods and systems, neural networks



Anatolii O. Nikolenko, Candidate of Technical Sciences, Associate Professor of Information Systems Department
Research field: Modeling systems, Digital signal and image processing, Artificial Intelligence Methods and Systems



Oksana Yu. Babilunha, Candidate of Technical Sciences, Associate Professor of Information Systems Department
Research field: Image Processing, Data analysis, Artificial Intelligence Methods and Systems