8. Fuzzy Approximation-Based Adaptive Backstepping Optimal Control for a Class of Nonlinear Discrete-Time Systems With Dead-Zone / Liu Y.-J., Gao Y., Tong S., Li Y. // IEEE Transactions on Fuzzy Systems. 2016. Vol. 24, Issue 1. P. 16–28. doi: https://doi.org/10.1109/tfuzz.2015.2418000
9. Lozynskyy A., Demkiv L. Synthesis of fuzzy logic controller of nonlinear dynamic system with variable parameters // Computational problems of electrical engineering. 2016. Vol. 6, Issue 2. P. 91–98.
10. Orlowska-Kowalska T., Kaminski M., Szabat K. Implementation of a Sliding-Mode Controller With an Integral Function and Fuzzy Gain Value for the Electrical Drive With an Elastic Joint // IEEE Transactions on Industrial Electronics. 2010. Vol. 57, Issue 4. P. 1309–1317. doi: https://doi.org/10.1109/tie.2009.2030823
11. Adaptive Fuzzy Control of a Class of Nonlinear Systems by Fuzzy Approximation Approach / Chen B., Liu X. P., Ge S. S., Lin C. // IEEE Transactions on Fuzzy Systems. 2012. Vol. 20, Issue 6. P. 1012–1021. doi: https://doi.org/10.1109/tfuzz.2012.2190048
12. Anastassiou G. A. Higher Order Multivariate Fuzzy Approximation by basic Neural Network Operators // Cubo (Temuco). 2014. Vol. 16, Issue 3. P. 21–35. doi: https://doi.org/10.4067/s0719-06462014000300003
13. Piegat A. Fuzzy Modeling and Control. Springer, 2001. 728 p. doi: https://doi.org/10.1007/978-3-7908-1824-6
14. Interactive collaborative Robotics / A. Ronzhin, G. Rigoll, R. Meshcheryakov (Eds.) // Proceedings of the Third International Conference on Interactive Collaborative Robotics, ICR. Springer, 2018. 302 p. doi: https://doi.org/10.1007/978-3-319-99582-3
15. Function approximation approach to the inference of neural network models of genetic networks / Kimura S., Sonoda K., Yamane S., Matsumura K., Hatakeyama M. // IPSJ Transactions on bioinformatics. 2007. Vol. 48. P. 9–19.

UDC 004.032.26
DOI: 10.15587/1729-4061.2019.160670

INCREASING THE SHARE OF CORRECT CLUSTERING OF CHARACTERISTIC SIGNAL WITH RANDOM LOSSES IN SELF-ORGANIZING MAPS

S. Shapovalova
PhD, Associate Professor*
E-mail: lanashape@gmail.com
Yu. Moskalenko
Postgraduate student
E-mail: yuramuv@gmail.com
*Department of Automation of Designing of Energy Processes and Systems
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
Peremohy ave., 37, Kyiv, Ukraine, 03056

Проведено аналіз способів оптимізації алгоритмів функціонування нейронних мереж Кохонена – карт самоорганізації (Self-organizing map – SOM) за швидкістю навчання та часткою коректної кластеризації. Визначено ефективну оптимізацію карт самоорганізації за другим критерієм – Enhanced Self-Organizing Incremental Neural Network (ESOINN). Визначено, що у випадку неповного вхідного сигналу, тобто сигналу з втратами в невідомі моменти часу, частка коректної кластеризації неприпустимо низька на будь-яких алгоритмах SOM, як базових, так і оптимізованих.

Неповний сигнал представлено як вхідний вектор нейронної мережі, значення якого подані єдиним масивом тобто без урахування відповідності моментів втрат поточним значенням і без можливості визначення цих моментів. Запропоновано та програмно реалізовано спосіб визначення відповідності неповного вхідного вектора до вхідного шару нейронів для підвищення частки коректного розпізнавання. Спосіб засновано на пошуку мінімальної відстані між поточним вхідним вектором та вектором-ваг кожного з нейронів. Для зменшення часу роботи алгоритму запропоновано оперувати не окремими значеннями вхідного сигналу, а їх неподільними частинами та відповідними групами вхідних нейронів. Запропонований спосіб реалізовано для SOM та ESOINN. Для доведення ефективності реалізації базового алгоритму SOM проведено його верифікацію з існуючими аналогами інших розробників.

Розроблено математичну модель для формування прикладів повних сигналів навчальної вибірки на основі еталонних кривих другого порядку та сформовано навчальну вибірку. За цією навчальною вибіркою було проведено навчання всіх нейронних мереж, реалізованих з використанням запропонованого способу та без нього. Розроблено схему імітації втрат та згенеровано тестові вибірки для обчислювальних експериментів на неповних сигналах.

На основі експериментів доведено ефективність запропонованого способу для класифікації за неповним вхідним сигналом на основі карт самоорганізації як для реалізацій базового алгоритму SOM, так і для ESOINN

Ключові слова: карта самоорганізації, SOM, ESOINN, нейронні мережі Кохонена (Kohonen), сигнал з втратами, втрати в часовому ряді, класифікація за характеристичним сигналом

1. Introduction

One of the basic problems in systems of object control in terms of input signal of their characteristics is classification problem. If the signal is represented as a time series, it is expedient to use neural networks to ensure high level of recognition accuracy. When a part of a signal or some of its values are lost at unknown time points, it is impossible to apply varieties of networks in which size of the input layer corresponds to size of the input vector. However, a Kohonen network can be used, namely self-organizing maps (SOM). Compared with other neural networks, conceptual advantage of SOM is ability to teach them by a small number of examples of training samples which is essential for automatic control systems, especially for pre-fault modes of equipment operation.

The SOM neural networks have come into wide use in solving present-day applied problems of classification in terms of the characteristic signal in a variety of industries, for example:

– diagnosis of quality of apples and corn seeds based on dynamic speckle [1]. Results of seed state descriptors serve as the SOM input information;

– classification of astronomic objects. The SOM input is data of the Gaia space telescope [2];

– creation of samples of local galaxies after astration [3]. Database of Sloan Digital Sky Survey (SDSS) – the project for training images and spectra of red shift of stars and galaxies – is the basic source of information;

– comparison of protein ensembles [4]. Results of computer modeling of protein ensembles are the input information.

Result in these examples is represented by the SOM visualization in the form of a colored topographic map.

However, when the SOM signal with losses is clustered in the network, quotient of correct relation to the corresponding class is also significantly reduced which makes it impossible to use this network when solving practical problems, for example, anomaly detection. Therefore, improvement of the SOM functioning algorithms for recognition of signals with losses is an urgent problem of practical importance.

2. Literature review and problem statement

Kohonen's neural networks, in particular SOM, are based on the concept of self-organization [5]. Such networks in their specialization are adapted for data visualization and cluster analysis. Advantages of the SOM include resistance to noisy data, fast training and a small number of examples of the training sample.

Supersensitivity to the initial distribution of weight values is the big disadvantage of this type of neural networks. Solution to this problem is proposed in [6–11]. An effective SOM initialization scheme is proposed in [6]. It consists of the follows: the training sample examples in which vectors differ substantially are specially initialized in different parts of the SOM. Modeling results have proved that primary initialization significantly accelerates training. An example of primary initialization of vertices of the SOM clusters before basic training by examples for which membership in a particular class is known in advance is given in [7]. This has made it possible to reduce training time and subject the SOM clusters to forced disconnection.

In addition to improvement of primary initialization, the SOM algorithms are optimized according to the following criteria: acceleration of training and raising the share of correct clustering in the network working mode. According to these criteria, it is possible to conditionally distinguish two approaches to optimization of the SOM algorithms.

The first approach involves methods for training acceleration.

Criteria for termination of conventional SOM training include absence of significant changes in the neural network weights at the current step or reaching the specified maximum number of steps. A new criterion for stopping training is proposed in [8]: the degree of maintaining the topology. An optimized network is more likely to form clusters that are not topologically ordered.

Modification of the algorithm of training by large sets of dissimilar data is proposed in [9] for cutting training time. The algorithm of training by sparse data with a cut of processing demand has been improved in [10].

Application of R-trees and k–d trees for indexing data in order to accelerate the training process by multidimensional data sets is proposed in [11].

General disadvantage of the varieties of neural networks presented in [6–11] is relatively low share of correct recognition in the operating mode. Therefore, the second approach to optimizing SOM algorithms is aimed at increasing the share of correct recognition due to a clear definition of cluster boundaries.

This approach consists in creation of a SOM with a variable structure, for example, by providing an additional space for new clusters. This is realized by adding new fragments of the SOM grid which are called nodes, units or clusters in [12–16]. Node is a neuron or a group of neurons that create a new whole cluster or its «share» added in the training process.

Two SOM models are presented in [12]. The first corresponds to Kohonen's basic concept of training without a teacher. However, problems are solved using the process of controlled growth of the SOM dimension and removal of nodes. The second model is the result of combining the previous one with application of radial basic functions.

A self-organizing map model is presented in [13]. It adds a new node if weights differ significantly from those available in the network. The study [14] describes modification of the Kohonen's network, Growing Self-Organizing Map (GSOM), which solves the problem of choosing optimal dimension of a self-organizing map. The GSOM network is used for problems of nonlinear diminution of dimension, approximation, and clustering.

Modification of the self-organizing map, Self-Organizing Incremental Neural Network (SOINN), is presented in [15]. The network has two layers that are trained one after another. The first layer is used to determine topology of clusters and the second is to determine number of clusters and identify nodes for them. The first layer is the input for the second one. Similar neurons of the first layer which are united by the bond determine the same class. When the first layer completes learning, the second layer starts learning by a similar algorithm. The map «grows» in the training process and is less sensitive to noise of the input signal.

A self-organizing map called Enhanced Self-Organizing Incremental Neural Network (ESOINN) is proposed in [16]. The following SOINN drawbacks are eliminated in it:

1) uncertainty of the moment of stopping training the first layer and start of training the second;

2) the problem of merging neighboring clusters. The ESOINN network better divides close and partially overlapping clusters.

Thus, ESOINN is optimal concerning accuracy of SOM recognition as evidenced, in particular, by the results of trials on the MNIST data sets in [17].

However, both SOM and ESOINN do not provide a sufficient share of correct clustering of signals with losses. The problem arises because the winning neuron is chosen in the SOM by the minimum distance between the current input vector and the vector of weights of the given neuron. If a complete signal is sent, then dimension of the current input signal is equal to dimension of the input layer of the self-organizing map. Thus, each i-th neuron of the input layer corresponds to the value of the i-th element of the input vector. If the input signal is incomplete (a part of the signal is lost), that is, dimension of the input vector is not equal to dimension of the network input layer, values of the corresponding indices of the input signal and the input neuron layer do not coincide, error grows and the signal is clustered incorrectly. The problem has to be solved by additional preliminary determination of conformity of the input signal to the neurons of the input layer. Therefore, it is necessary to conduct study and improve algorithms of the SOM functioning in order to achieve correct classification in a case of signals with losses.

3. The aim and objectives of the study

The study objective is raising accuracy of recognition of the input signal with losses based on the SOM neural network.

To achieve this objective, the following tasks have to be solved:

– to develop a method of classification in terms of the characteristic signal with random losses on the basis of self-organizing maps;

– to develop a mathematical model for formation of examples of training and testing samples;

– to select existing and develop own program implementations of the SOM according to the basic algorithm and the algorithm supplemented by the proposed method;

– to verify proposed program implementations with existing counterparts based on computational experiments;

– to carry out computational experiments with recognition of incomplete input signals.

4. Classification in terms of the characteristic signal with random losses on the of SOM basis

4. 1. Problem statement for a neural network complex

The problem of recognizing the input signal consists in the general case in determining conformity of the input signal.

$$\overline{X}_{SOM} = [x_1, x_2 \ldots, x_n], \quad (1)$$

where x_i is the value of the input signal at the i-th time point to the class a_k which belongs to the set:

$$A = \{a_1, a_2, ..., a_d\}. \quad (2)$$

Complexity of classification lies in the fact that it is not known in advance at what moments of time t_j the corresponding dimensions of the signal x_j were lost. That is, a single data array which has no spaces is fed to the SOM input.

The SOM solves only the clustering problem, that is, it defines the SOM cluster that corresponds to the current input signal. After that, it is necessary to solve the problem of classification, that is, determine correspondence of this cluster to class a_k. To solve the classification problem, another neural network, for example, a multilayer perceptron (MLP) is most often used. Fig. 1 shows a schematic representation of the SOM and MLP complex.

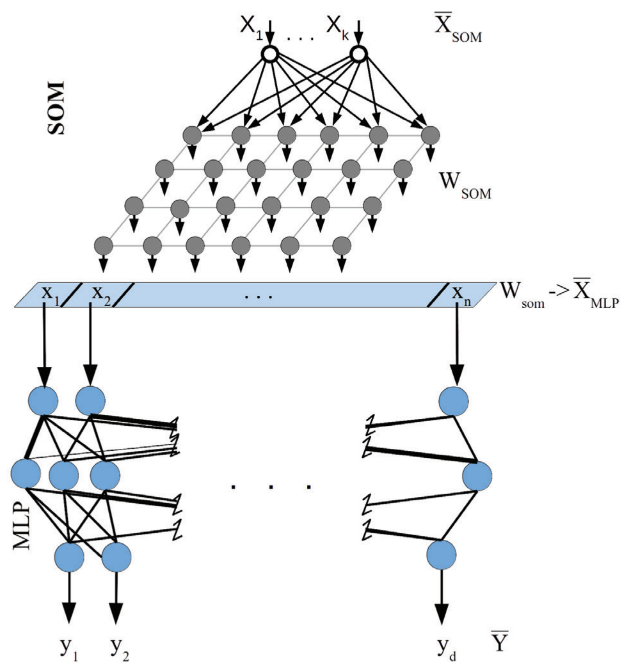

Fig. 1. Solution to the classification problem with SOM and MLP neural networks

The trained SOM receives an input signal in a working mode. The clustering problem solution consists in determining the most appropriate SOM cluster. This is reflected in the matrix of all outputs of neurons of the W_{som} grid. To solve the classification problem in the MLP, all rows of the W_{som} matrix are written sequentially and create an input of the MLP vector. Thus, dimension of the input layer of the MLP depends on dimension of the SOM matrix. Dimension of the output layer of the MLP equals to the number of classes of recognition of the application problem. Each neuron corresponds to a certain a_k class from the set A (2). The output vector of the MLP Y displays a set of probabilities of the input vector conformity to each of the classes. The result of the classification problem is the number of the neuron that defines the recognized class.

However, if a part of the characteristic signal is lost, the input vector does not conform to the input layer of the SOM network and the clustering problem is solved incorrectly. In order to avoid this, it is necessary to preliminarily determine conformity of the input signal to neurons of the input layer.

4. 2. Determination of conformity of components of the input vector to neurons of the input layer of the SOM

The problem of correct recognition of incomplete signals can be solved on the basis of preliminary determination of conformity of neurons of the input layer of the neural network to elements of the input signal. Conformities between values of the input signal x_i and the neuron n_j of the input layer of the neural network should be established.

A signal is represented as an input vector of the neural network the values of which are given by a single array, that is, without taking into consideration conformities of the moments of losses to the current values even in a case of impossibility of determining these moments. Fig. 2 shows schematic diagram of such conformity for complete (Fig. 2, *a*) and incomplete (Fig. 2, *b*) input vectors. If a signal with losses (that is losses of neurons of the input layer are greater than values of $\overline{X}_{som}$) is sent to the neural network input, then only the first k neurons of the input layer will be involved resulting in incorrect result of recognition.

This problem is solved by determining comparison of neurons of the input layer with the input vector of the SOM. This comparison is optimal for the current signal. The following algorithm is used:

1. Determining initial combination C_1 of comparison of x_i with n_j for all k neurons of the input signal according to the rule:

$$x_i = n_i. \quad (4)$$

2. Feeding $\overline{X}_{som}$ to the SOM input.

3. Determining the winning neuron by the SOM algorithm.

4. Checking completion of the algorithm:

Since the number of neurons involved in the SOM is equal to the number of neurons in the input layer, current error of the winning neuron (with no account of missed values) is calculated as the Euclidean distance between the vector of its weights and the input signal:

$$D = |\overline{X}_{som} - W_j|^2, \quad (5)$$

where $\overline{X}_{som}$ is the current input vector of the SOM; W_j is the vector of weights of the j-th winning neuron.

The algorithm completes in the following cases:

– if the current error is less than the specified minimum error value δ:

$$D < \delta, \quad (6)$$

number of the current winning neuron is returned;

– if all combinations of comparison of components of the input vector and neurons of the input layer were enumerated without meeting condition (6), the result is returned with the smallest error D.

5. Changing the current combination of comparison C_p for C_{p+1} occurs as one step of transposition of connection x_i to n_j. For the first iteration, shift occurs for $i=k$. With each subsequent iteration, i is decreased by 1.

6. Returning to p. 2.

As a result of execution of the algorithm of comparison of components of the input signal with neurons of the input layer, an optimal combination of comparison is determined. After that, clustering is performed in the SOM.

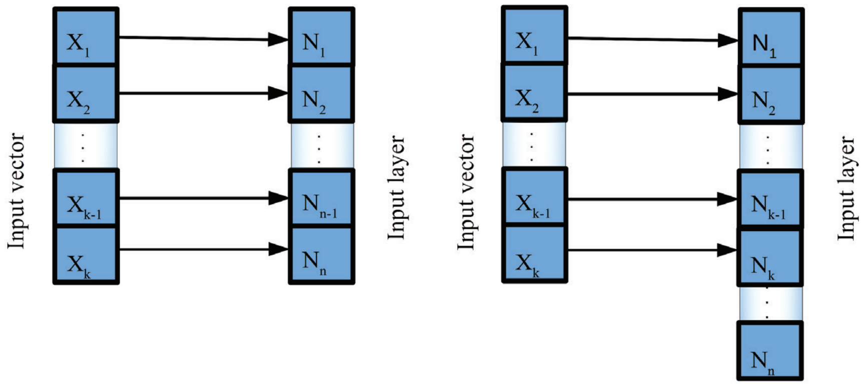

Fig. 2. Conformity of the input vector to the input layer: a case of a complete signal ($k=n$) (*a*); a case of a signal with losses ($k<n$) (*b*)

This method disadvantage consists in the fact that its computational complexity is factorial and depends on the amount of lost values of the input vector. If condition (6) was not met at early stages in the cycle, then a complete enumeration of conformity of connections of components of the input vector and neurons of the input layer is performed to find a combination with a minimum error.

In order to solve the problem of rapid growth of computational complexity, we propose to operate not with individual values of the input signal but with their groups. Conformity between the input layer and the input vector is sought not for each value of the input vector but for a group of values. A group of neurons is an indivisible fragment of the input signal with g values of x^g. This fragment is indivisible in terms of the fact that all connections and elements of the group correspond to the neurons of the input layer on the principle:

$$X_i^g = n_j + \text{const}, \quad (7)$$

that is, all connections within the group in each iteration are equally shifted by one value.

Optimal size of the group g is determined by means of available computing resources and can be determined experimentally as the time spent for classification by the test examples.

The smaller dimension of the group of values of the input vector the greater computational complexity. On the other hand, the larger the number of components of the input vector the higher accuracy can be achieved.

The computational experiments performed with the test task have determined that correct recognition is retained at a loss of up to 20 % of the input signal. Thus, with a complete input vector of 100 values, the order of complexity of the algorithm is 10^{20} and when using groups of neurons with initial dimension 7, it decreases by 10^{18} times.

4. 3. Stages in solving the problem of classification with incomplete signal

1. Implementation of the SOM and MLP software complex according to the architecture shown in Fig. 1.

2. Training an autonomous SOM by a training sample with examples of a complete signal.

3. Training in the MLP neural network complex using the SOM output with the same training sample.

4. Determination of conformity of components of the current input signal to the neurons of the input layer of the SOM by the proposed method.

5. Classification on the SOM-MLP neural network complex where the SOM is brought to conformity with the current incomplete signal.

The first three stages are typical for classification and are performed once. After that, the neural network complex is ready to classify in terms of a complete signal. The fourth stage provides opportunity to classify incomplete signals at the fifth stage according to the SOM and MLP operation algorithms.

Thus, classification of each current signal with losses at unknown time points consists in performing the last two stages.

5. Mathematical model of formation of examples of complete signals in the training sample

The test example is the problem of recognizing curves of the second order in terms of similar fragments: upper parts of the circle, ellipse and parabola.

The input signal $\overline{X}_{som}$ is a list of discrete values of a function that defines the corresponding curve.

In order to collect the necessary number of examples of the training sample, the ε value is added to the x_i value of the corresponding curve. This value belongs to the range [–1.5·δ; 1.5·δ], where δ is the maximum distance between the x_i points belonging to different reference curves. The ε value is calculated as the Gaussian distribution:

$$\varepsilon = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \quad (8)$$

where σ=0.4.

If the random value goes beyond the range, then assume that ε=0.

Examples of the training sample are calculated by formulas (9)–(11):

– for a circle:

$$x_i = \sqrt{(R^2 - \tau^2)} + \varepsilon, \quad (9)$$

where x_i is ordinate of the point at the i-th moment of time; R is the circle radius;

– for an ellipse:

$$x_i = \sqrt{(1 - \tau^2 / a^2_{elips}) \cdot b^2_{elips}} + \varepsilon, \quad (10)$$

where a_{elips}, b_{elips} are parameters of the ellipse;

– for a parabola:

$$x_i = a_{parabola}\tau^2 + b_{parabola} + c_{parabola} + \varepsilon, \quad (11)$$

where $a_{parabola}$, $b_{parabola}$, $c_{parabola}$ are parameters of the parabola.

To complicate the problem, similar fragments of curves were selected (Fig. 4). The standard curves were given by the following equations (2 curves for each class):

– circle:

$$x_i = \sqrt{(9 - (\tau - 3)^2)} + \varepsilon, \quad (12)$$

$$x_i = \sqrt{(9 - (\tau - 2.9)^2)} + \varepsilon; \quad (13)$$

– ellipse:

$$x = -1/3\tau^2 + 2\tau + \varepsilon, \quad (14)$$

$$x = -1/3\tau^2 + 2.05\tau + 0.1 + \varepsilon; \quad (15)$$

– parabola:

$$x_i = \sqrt{(3 \cdot 3 - \frac{(\tau - 3)^2 \cdot 3^2}{3.2^2}} + \varepsilon, \quad (16)$$

$$x_i = \sqrt{(3 \cdot 3 - \frac{(\tau - 2.95)^2 \cdot 3^2}{3.2^2}} + \varepsilon. \quad (17)$$

Fig. 3 shows reference curves. It is shown that the range of values x_i taking into consideration the random ε variable contains all possible values of all examples built on the basis of these curves.

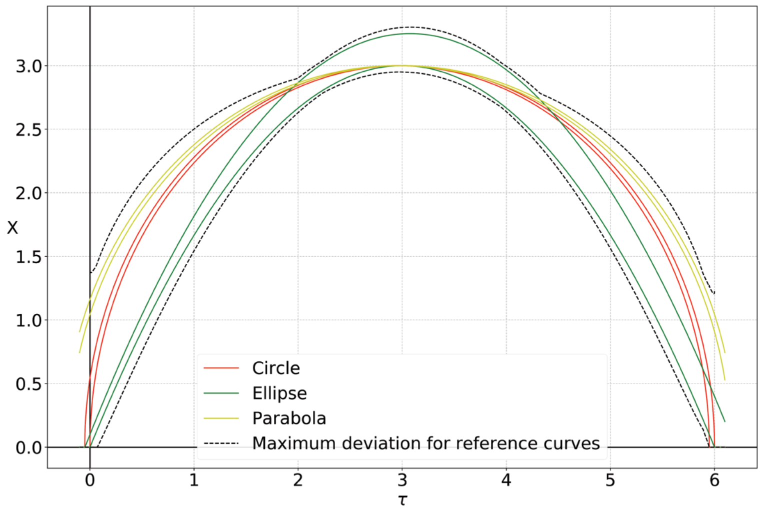

Fig. 3. Reference curves of the test problem

All SOM neural networks involved in further computational experiments were trained on the same training sample.

6. Realizations of neural networks

To prove effectiveness of the proposed method for determining conformity of the input vector to neurons of the input layer, the following versions of the neural networks were developed:

1) SOMbase self-organizing map implemented on the basic SOM algorithm;

2) SOMmod self-organizing map implemented on the basic SOM algorithm supplemented with the proposed method;

3) ESOINNmod self-organizing map implemented on ESOINN algorithm supplemented with the proposed method;

4) MLPbase multilayer perceptron implemented on the multilayer perceptron algorithm.

The first two versions of the SOM were implemented on the same data structures and basic algorithms for correct comparison of the SOM algorithm with the algorithm supplemented with the proposed method.

To compare the recognition results in optimized SOMs, ESOINN was selected from the resource [18] on the basis of which the proposed method was implemented in the ESOINNmod network.

To test the investigated Kohonen networks, the developed MLPbase was used with the exception of ENCOG [19] and NeuraPH [20] networks for which MLP was used from their libraries. This relates to the fact that the SOM only solves the clustering problem and then it is necessary to solve the classification problem according to the defined SOM clusters, that is, determine which cluster corresponds to which class of the applied problem.

In order to verify the developed software implementation of the SOM by the basic algorithm, computational experiments of recognition of the complete signal for comparison with the existing counterparts of other SOM developers: Encog, NeuroPH. Thus, correctness of the own implementation of the basic SOMbase algorithm was proved to be further developed by the proposed method.

Computational experiments with recognition of incomplete input signals were carried out in:

1) the SOM networks without the proposed method: SOMbase, NeuroPH, Encog, SOINN, ESOINN, GSOM;

2) the SOM networks with the proposed method: SOMMod, ESOINNmod.

After performing computational experiments, shares of correct recognition were compared by all methods.

The neural networks were trained according to the training sample which contained 100 examples (34 examples of a circle, 33 examples of an ellipse, 33 examples of a parabola). Each example reproduced a complete signal consisting of 100 values calculated by formulas (8), (12) to (17). The following self-organizing maps have been studied: NeuraPH, ENCOG, GSOM, SOINN, SOMbase, ESOINN, SOMmod, ESOINNmod and multilayer perceptron (MLP). Initial training speed for the SOM was 0.1. Speed of training all SOMs in the first 200 steps was forcibly reduced by 5 times and subsequently decreased by 5 times every 50 steps. It took 400 epochs to train the SOM.

The initial training speed of the multilayered perceptron was 0.01 and forcedly decreased by 5 times if during 5 epochs there were no changes in accuracy of classification of the test sample. Further, training was terminated if there were no changes in error during 15 steps.

For a final solution of the classification problem, matrix of outputs of all neurons of each SOM was sent to the MLP as a sequence of rows which determined the corresponding class membership. Encog, NeuroPH networks used MLPs from their libraries and the MLP of own development in other cases.

7. Verification of implementation of basic algorithms of the SOM and MLP

The first series of computational experiments was conducted to verify the SOM network developed by the basic algorithm with existing counterparts divided into networks without optimization: Encog [19], NeuraPH [20] and with optimization: SOINN [21], ESOINN [18], GSOM [22].

The test sample with which comparison was made contained complete signals of the input vector without losses individually generated by formulas (8), (12)–(17).

Since the data sample is balanced, accuracy of classification can be determined by the metric of accuracy:

$$Accuracy = \frac{P}{N}, \qquad (18)$$

where P is the number of classes correctly recognized by the classifier; N is the size of the training sample.

The purpose of the experiment was to determine share of correct complete signal recognition by each neural network under identical conditions. In all subsequent experiments, share of correct recognition was determined as follows:

$$Cor = Accuracy \cdot 100\,\%, \qquad (19)$$

where Accuracy is the accuracy metric.

Results of recognition of all SOM implementations participated in current and further calculations are given in Table 1. Presence or absence of optimization of the current SOM was also indicated for correct comparative analysis.

Table 1

Results of comparison of SOM implementations with a complete signal

No.	SOM implementation name	Availability of SOM optimization	Share of correct recognition, %
1	SOMbase	–	80
2	NeuroPH	–	79
3	Encog	–	81
4	SOINN	+	87
5	ESOINN	+	90
6	GSOM	+	86

Thus, effectiveness of own implementation of the basic algorithm (Table 1, rows 1–3) was proved and it was confirmed that optimized varieties of the SOM with a complete signal give better result in comparison with the neural networks implemented by the basic algorithm (Table 1, rows 4–6).

8. Computational experiments of classification in terms of a signal with random losses

The second series of computational experiments was designed to determine effectiveness of the SOM functioning in a condition of signal losses.

Since recognition efficiency is influenced by the extent of losses and the time intervals of their occurrence, several experiments that reproduced different cases of loss moments were conducted. From this point of view, recognition problems were divided into the following classes:

1. Loss of a single part of the signal.
2. Signal losses at unknown time points.

According to these classes, 4 test samples were generated (the first 3 for the first class of problems and the rest for the second class):

1. Loss of successive final values.
2. Loss of initial values.
3. Loss of successive values at a given interval from t_{21} to t_{79} with a random start.
4. Loss of signal values over the entire range from t_1 to t_{100} at random moments of time.

For each task, 100 examples of incomplete signal were generated. All of the neural networks used in experiments were trained by examples of a complete signal. When tested in the operating mode, a signal with losses was sent to the SOM input.

Results of the test for an incomplete signal are given in Table 2. Names of the neural networks whose developers implemented not basic but optimized SOM algorithm are shown in italics. The last two columns contain the recognition results using the proposed method of basic implementation of the algorithm (SOMmod) and one of the best optimizations in terms of the criterion of share of correct recognition (ESOINNmod).

Table 2

Recognition results for an incomplete vector

No.	The problem specification		Share of correct recognition in implementations of neural networks, %							
			Without the proposed method						With the proposed method	
	Range of losses at time points τ_i	Amount of lost values	NeuraPH	ENCOG	GSOM	SOINN	SOMbase	ESOINN	SOMmod	ESOINNmod
Loss of a single fragment of the input signal										
1	80–100	5	84	84	85	85	84	93	85	93
		10	82	82	86	86	82	91	83	91
		20	79	81	85	87	82	90	82	90
2	1–20	5	32	32	38	38	32	21	73	77
		10	19	20	35	35	20	17	70	73
		20	16	16	17	14	16	14	66	69
3	1–100	5	48	48	56	58	48	62	77	81
		10	41	41	54	49	41	42	74	77
		20	32	33	35	35	33	38	68	69
Losses of the input signal at unknown time points										
4	1–100	5	68	69	69	71	68	74	74	79
		10	54	54	58	60	54	56	61	66
		20	37	37	39	39	37	43	69	71

From the point of view of the signal loss moments, signals that have lost their initial values are recognized the worst.

The results of recognition of such signals in all optimized and non-optimized SOMs without implementation of the proposed method were unacceptably low and measured:

– at a 5 % loss of the signal: from 21 % to 38 %;

– at a 10 % loss of the signal: from 19 % to 38 %;

– at a 20 % loss of the signal: from 14 % to 16 %.

All networks are most resistant to the loss of finite values. Approximately the same results were shown by networks with the proposed method and without it. Accuracy ranged from 79 % to 95 %.

In the case of a signal loss in the range from the 1st to the 100th value, the result was worse. However, the experiments have shown that the proposed method significantly increased share of correct recognition, especially when a large number of values were lost.

Results of the incomplete signal classification (Table 2) indicate that the proposed method significantly increases share of correct clustering at incomplete input data. This applies to both the network implemented by the basic SOM algorithm and for the optimized network whose algorithm is supplemented with the proposed method.

9. Discussion of classification results in terms of incomplete input signal based on self-organizing maps

A method for determining conformity of incomplete input vector to the input layer of neurons based on the difference between the input signal vectors and weights of the input neuron layer was proposed. According to this method, it is necessary to make a preliminary determination of combination of links between components of the input vector and the neurons of the input layer.

This preliminary preparation of signal sending to the neural network reduces the classification error. This is because the desired cluster is forcedly activated in the SOM. To do this, a combination of conformity of components of the incomplete input vector to neurons of the SOM input layer is determined. In this way, distance between the current input vector and the current weight vector is minimized by formula (5) and, accordingly, the classification error.

The proposed method is presented in the form of an algorithm, each stage of which has necessary formulas and a design diagram. The method enables solution of the problem of diagnostic signal recognition with losses at unknown time points.

To prove this method, an appropriate test problem was proposed. It was necessary to create a classification problem in which it is possible to model examples of signals in the form of time series and change the moments of loss of their values. To complicate statement of the classification problem, a condition was set that the signals should be similar in description so that when imitating noisy values, they partly overlap one another. Mathematical model of generation of training examples based on the analytical equations of similar fragments of curves of the second order (graphical representation and overlap area are shown in Fig. 4) has become a solution to this problem The training sample is based on formulas (8), (12) to (17). Thus, a mathematical model of simulation of noisy diagnostic signals was created. With its help, training and test samples can be generated for solving the complex classification problem.

For further experiments, own SOM implementation by the basic algorithm, the SOMbase, was developed. In order to prove its effectiveness, comparison with counterparts of other developers was made (Table 1).

For the computational experiments conducted to prove effectiveness of the proposed method, existing SOM software implementations were selected based on the basic algorithm (NeuraPH, Encog) and with three optimization variants of other developers (GSOM, SOINN, ESOINN). Two implementations of the SOM (own SOMbase and one of the most effective in terms of the share of correct recognition, ESOINN) were supplemented with the proposed method. Other implementations remained unchanged.

The test results (Table 2) proved that the proposed method significantly improves accuracy of clustering, however, the time points of losses significantly affect its efficiency. With the least effective use, indicators remain the same or are slightly better (by 1–3 %). In the best cases, the share of correct recognition increases from 14 % to 69 % in the ESOINN neural network supplemented with the proposed method.

Thus, efficiency of the proposed method of incomplete signal classification was experimentally proved.

Practical significance of the study consists in the possibility of using the proposed method for diagnosing state of technical objects in time series of their characteristics. The feature is that there is no need for conversion of the original signal and extraction of additional characteristics, such as frequency. In many application problems of control, this has demanded current characteristics in real time with possible signal losses caused by external factors.

This method disadvantage is growth of computational complexity, that is, the classification time. To eliminate this disadvantage, application of groups was proposed. This has made it possible to reduce 10^{18} times time of classification of the test sample.

Likelihood of incorrect classification of signals with surges is one of limitations of this method. It is recognition of such signals and raise of SOM stability that is the prospect of further studies. This is necessary, in particular, to solve problems of finding anomalies.

10. Conclusions

1. A method for classification in terms of the characteristic signal with losses at random moments of time based on self-organizing maps was proposed and the program implemented. The existing classification procedure was supplemented with definition of conformity of the incomplete input vector to the input neuron layer to increase the share of correct recognition.

2. A mathematical model of generation of examples of complete signals of a training sample on the basis of reference curves of the second order was developed. In this way, special test problems were added to determine efficiency of neural networks in terms of speed and accuracy of operation. Unlike the existing data sets that represent application problems, the proposed model enables simulation of signal parameters and size of the areas of class intersection. A simulation design of losses was developed and test samples were generated for experiments with incomplete signals. This has allowed us to model losses of discrete values or entire fragments of the signal at different time points.

3. Analysis of methods for optimization of the SOM algorithm was made. Existing SOM implementations and their optimizations for computational experiments were selected: ENCOG, NeuroPH, SOINN, GSOM, ESOINN. Effective optimization of SOM, ESOINN was determined. Software implementation of the SOM was developed for computational experiments using the basic algorithm and the algorithm supplemented with the proposed method and the ESOINN program implementation was supplemented with the proposed method.

4. To prove effectiveness of implementation of the SOM basic algorithm, it was compared with existing counterparts: NeuroPH, Encog. It was proved that the share of correct classification of own development (80 %) corresponds to the results of the same algorithm of other developers (78 % and 81 %, respectively).

5. Based on computational experiments, efficiency of the proposed method for classification in terms of incomplete input signals based on self-organizing maps was proved for both implementation of the basic SOM algorithm and the most effective optimization of the ESOINN. Depending on the loss time points, the share of correct recognition in the ESOINN neural network supplemented with the proposed method slightly improves by 1–3 % in the worst case and from 14 % to 69 % in the best case.

References

1. Unsupervised learning segmentation for dynamic speckle activity images / Passoni L. I., Dai Pra A. I., Meschino G. J., Guzman M., Weber C., Rabal H., Trivi M. // arXiv. 2014. URL: https://arxiv.org/abs/1408.3818
2. An approach to the analysis of SDSS spectroscopic outliers based on Self-Organizing Maps / Fustes D., Manteiga M., Dafonte C., Arcay B., Ulla A., Smith K., Borrachero R., Sordo R. // arXiv. 2013. URL: https://arxiv.org/abs/1309.2418
3. A large sample of Kohonen selected E+A (post-starburst) galaxies from the Sloan Digital Sky Survey / Meusinger H., Brünecke J., Schalldach P., in der Au A. // Astronomy & Astrophysics. 2017. Vol. 597. P. A134. doi: https://doi.org/10.1051/0004-6361/201629139
4. Fraccalvieri D., Bonati L., Stella F. Self Organizing Maps to efficiently cluster and functionally interpret protein conformational ensembles // Electronic Proceedings in Theoretical Computer Science. 2013. Vol. 130. P. 83–86. doi: https://doi.org/10.4204/eptcs.130.13
5. Kohonen T. Self-organized formation of topologically correct feature maps // Biological Cybernetics. 1982. Vol. 43, Issue 1. P. 59–69. doi: https://doi.org/10.1007/bf00337288
6. Su M.-C., Liu T.-K., Chang H.-T. Improving the self-organizing feature map algorithm using an efficient initialization scheme // Tamkang Journal of Science and Engineering. 2002. Vol. 5, Issue 1. P. 35–48.
7. Shapovalova S. I., Sharaievskyi H. I. Kompiuterne modeliuvannia karty samoorhanizatsiyi dlia rozviazannia zadachi rozpiznavannia syhnaliv // Visnyk natsionalnoho universytetu «Lvivska politekhnika». 2007. Issue 574. P. 75–80.

8. Su M.-C., Chang H.-T. Genetic-algorithms-based approach to self-organizing feature map and its application in cluster analysis // 1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98CH36227). 1998. doi: https://doi.org/10.1109/ijcnn.1998.682372
9. El Golli A. Speeding up the self organizing map for dissimilarity data // Proceedings of International Symposium on Applied Stochastic Models and Data Analysis. Brest, 2005. P. 709–713.
10. Conan-Guez B., Rossi F., El Golli A. Fast algorithm and implementation of dissimilarity self-organizing maps // Neural Networks. 2006. Vol. 19, Issue 6-7. P. 855–863. doi: https://doi.org/10.1016/j.neunet.2006.05.002
11. Cuadros-Vargas E., Romero R. F., Obermayer K. Speeding up algorithms of SOM family for large and high dimensional databases // In Workshop on Self Organizing Maps. Kitakyushu, 2003.
12. Fritzke B. Growing cell structures – A self-organizing network for unsupervised and supervised learning // Neural Networks. 1994. Vol. 7, Issue 9. P. 1441–1460. doi: https://doi.org/10.1016/0893-6080(94)90091-4
13. Interconnected growing self-organizing maps for auditory and semantic acquisition modeling / Cao M., Li A., Fang Q., Kaufmann E., Kröger B. J. // Frontiers in Psychology. 2014. Vol. 5. doi: https://doi.org/10.3389/fpsyg.2014.00236
14. Growing self-organizing map approach for semantic acquisition modeling / Cao M., Li A., Fang Q., Kroger B. J. // 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom). 2013. doi: https://doi.org/10.1109/coginfocom.2013.6719269
15. Furao S., Hasegawa O. An incremental network for on-line unsupervised classification and topology learning // Neural Networks. 2006. Vol. 19, Issue 1. P. 90–106. doi: https://doi.org/10.1016/j.neunet.2005.04.006
16. Furao S., Ogura T., Hasegawa O. An enhanced self-organizing incremental neural network for online unsupervised learning // Neural Networks. 2007. Vol. 20, Issue 8. P. 893–903. doi: https://doi.org/10.1016/j.neunet.2007.07.008
17. Algoritm Uluchshennoy Samoorganizuyushcheysya Rastushchey Neyronnoy Seti (ESOINN). URL: https://habr.com/post/206116
18. An enhanced self-organizing incremental neural network for online unsupervised learning. URL: https://github.com/BelBES/ESOINN
19. Encog Machine Learning Framework. URL: https://github.com/encog/encog-java-core
20. Neuroph – Java Neural Network Platform Neuroph. URL: https://github.com/neuroph/neuroph
21. Self-Organizing Incremental Neural Network. URL: https://github.com/fukatani/soinn
22. Growing Self-Organizing Map. URL: https://github.com/philippludwig/pygsom