12. Dovbysh, A. S. Foundation of intelligent systems designing [Text]: tutorial / A. S. Dovbysh. – Sumy: Publisher SSU, 2009. – 171 p.
13. Dovbysh, A. Optimization of the parameters of teaching the control system of the power unit of the heat power plant [Text] / A. Dovbysh, M. Bibyk, A. Rudyi // Compressor and Power Engineering. – 2015. – Issue 3 (41). – P. 37–40.
14. Kutsenko, O. Geoinformation system of personnel identification at re-construction of the area [Text] / O. Kutsenko, B. Kascheev, M. Mironenko // Bulletin of the NTU "KhPI". Series: System Analysis, Management and Information Technology. – 2017. – Issue 28 (1250). – P. 34–40.

UDC 629.735.083.02.004.58:534.647:004.032.26
DOI: 10.15587/1729-4061.2017.109968

MULTI-CLASS RECOGNITION OF OBJECTS TECHNICAL CONDITION BY CLASSIFIER BASED ON PROBABILISTIC NEURAL NETWORK

N. Bouraou
Doctor of Technical Sciences,
Professor, Head of Department*
E-mail: n.burau@kpi.ua
D. Pivtorak
PhD*
E-mail: p_diana@i.ua
S. Rupich
Postgraduate Student*
E-mail: xyqserg@ukr.net
*Department of Instrumentation and Orientation and Navigation Systems
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
Peremohy ave., 37, Kyiv, Ukraine, 03056

Роботу присвячено дослідженню ефективності класифікатора на основі імовірнісної нейронної мережі для багатокласової діагностики об'єкта за наявності багатоосередкового пошкодження. Використано багатовимірний вектор діагностичних ознак, що містить 5 елементів. Сформовано множини навчальних та тестових вхідних векторів, виконано навчання та тестування класифікатора. Проаналізовано ефективність багатокласового розпізнавання в залежності від характеристик класифікатора та множини навчальних векторів

Ключові слова: багатокласове розпізнавання, нейромережевий класифікатор, вектор діагностичних ознак, імовірність правильної класифікації

Работа посвящена исследованию эффективности классификатора на основе вероятностной нейронной сети для многоклассовой диагностики объекта при наличии многоочагового повреждения. Использован многомерный вектор диагностических признаков, содержащий 5 элементов. Сформированы множества учебных и тестовых входных векторов, выполнено обучение и тестирование классификатора. Проанализирована эффективность многоклассового распознавания в зависимости от характеристик классификатора и множества обучающих векторов

Ключевые слова: многоклассовое распознавание, нейросетевой классификатор, вектор диагностических признаков, вероятность правильной классификации

1. Introduction

Ensuring the reliability and efficiency of operation of complex spatial objects is a topical issue in the aviation, power, oil and gas indries, as well as for special-purpose engineering structures. In general, such objects are characterized by large dimensions, non-stationarity of processes, distribution of parameters, nonlinearity, incomparteness of control of external factors, conditions and modes of functioning. Design of structural elements of such objects is based on the principle of safe damage, which allows for a microdefect, but such that does not lead to efficiency loss and object destruction [1–3]. However, the presence of welded or rivet joints of structural elements of complex spatial objects poses a threat of the emergence and development of multi-site damages. This may lead to destruction characterized by a sudden and rapid propagation due to combining among themselves and absorbing small-size cracks. Such a nature of damage development, difficult operating conditions, limited information about the actual technical condition lead to the multi-classing of objects in both time and space. In order to ensure safe and effective operation of such objects, it is necessary to provide multi-class diagnostics for timely detection of damage, assessment of its extent, monitoring of its development and interaction on large-sized surfaces of complex spatial objects. This will contribute to ensuring the reliability and efficiency of operation, preventing the destruction of complex spatial objects and averting catastrophic consequences.

2. Literature review and problem statement

Continuous monitoring of the technical condition (TC) of structures in operation, development control of damage, operational loads can be implemented in monitoring systems based on the concept of Structural Health Monitoring

(SHM) [4, 5]. SHM systems are developed as extensive information networks that are similar to the human nervous system. The systems provide measurement, recording, conversion, transmission and complex analysis of data from a finite set of spatially distributed sensors of primary information. The sensors are constructed according to a variety of physical principles, permanently attached or built in a design and provide structural integrity. The synthesis of such systems is based on the optimum combination of modularity and multi-channeling principles, taking into account information aspects of diagnostic processes. The modular principle is implemented by a set of sensors for obtaining information sufficient for a comprehensive assessment of the operational load and current TC of one or more structural units of a controlled object. The principle of multi-channeling is realized both within one module (when implementing one physical principle and control method), and by combining several modules for solving the diagnostic problem in relation to one structural unit of an object. Methods of signal processing are selected separately for each module, depending on the information content of the physical quantities or characteristics used as diagnostic information. These can be determination of higher-order statistical and spectral characteristics, evaluation of distribution laws of informative parameters, time-scale analysis, fractal analysis.

Modern diagnostic and monitoring systems are characterized by the use of information technology based on artificial intelligence, which ensures the processing, comparison, image classification operations unavailable in traditional mathematics, the possibility of self-learning and self-organization. In particular, in [6], artificial neural networks (NN) are used in problems of acoustic emission signal classification, and in [7], the authors used a family of models of multidimensional classifiers based on the Bayesian network for multidimensional classification. The use of neural networks for the two-class diagnostics of rotor elements of aircraft gas turbine engines based on the analysis of vibration and acoustic signals in stationary and non-stationary modes is justified in [8]. Integrated approaches and classification methods based on artificial neural networks and genetic algorithms are proposed for the diagnostics of concrete structures [9]. In [10], the use of a multilayer perceptron in the SHM systems is proposed. The practical implementation of such a neural network for damage recognition in airframe components is presented in [11]. Classification of the structural component condition is performed according to the extent of damage, which throughout the study took discrete values (increased from a certain minimum value to the one that characterizes the maximum damage). In [12–14], the Probabilistic Neural Network (PNN) was used for the condition classification and damage identification. The PNN provides nonlinear division into classes, has high sensitivity to small changes in diagnostic features, is capable of distinguishing among conditions according to changes in the number of diagnostic features. In [12], the PNN is used to identify the damage in the aircraft wing structure according to changes in the natural frequency of the structural element. And in [13], the possibility of identification, localization and classification of two types of damage (crack and loss of rivets) is investigated. For classification, signals from eight built-in piezoceramic sensors were used, each signal being an input signal of a certain neural network. Thus, eight PNNs of the same architecture were built, each being designed for damage classification according to changes in the signal of the corresponding sensor.

In general, the above works deal with solving the problem of two-class diagnostics, when the fact of absence or presence of damage is established, or the type of damage among two possible ones is determined. However, the studies do not solve the problem of multi-class recognition of the condition of complex spatial objects in the event of emergence and development of multi-site damage.

In [14], the development of the PNN-based classifier for the multi-class recognition of the TC of the tank with environmentally hazardous substances was performed. The classifier is a part of a complex monitoring system based on the SHM concept.

Elements of such classifier (Fig. 1) are:

– a training set of images or diagnostic features (P vector);

– a set of target classes (T vector);

– a connectivity matrix T_c, which establishes the membership of the input vectors with the corresponding classes S_k;

– a neural network that performs classification and recognition of the object TC;

– a test set of images (P_{test} vector).

During functioning, the latter is replaced with a set of actual data coming from the array of sensitive elements.

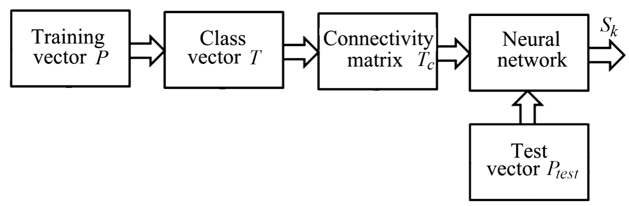

Fig. 1. General scheme of the condition classifier based on the neural network

The PNN is based on the architecture of a radial basis network, which consists of two layers. Neurons of the first layer have radial basis activation functions, and the second layer is called a competition layer. It estimates the probability of membership of the input vector with a particular class and compares the input vector with that class, the probability of membership with which is higher [8]. Each input vector of the NN corresponds to a certain initial or target value, and an "input/target" membership vector is formed for a set of input and output values. The training set contains Q pairs of "input/target" vectors. There are Z classes of possible membership of the input vector. As a result, the connectivity matrix T_c with the dimension ZxQ, which consists of zeros and units, can be formed. The rows of this matrix correspond to the membership classes, and the columns – to the input vectors. Thus, if the $T_c(i, j)$ element of the connectivity matrix is equal to 1, this means that the j-th input vector belongs to the i class. The number of neurons in the first layer is formed by the number of Q pairs of "input/target" vectors of the training set. The initial competition layer contains Z neurons, according to Z classes.

In general, the column diagnostic feature vector A_0, used for the condition recognition, may consist of any number of elements – diagnostic features

$$A_0 = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix}. \tag{1}$$

Diagnostic features a_n may include spectral, correlation, fractal, statistical characteristics of the measured signals. The number of features may vary depending on the number of measuring channels, the diagnostic value of features and the number of classes of the technical condition. However, it is desirable that the vector had at least 3 features for reliable classification.

We will consider the problem of multi-class diagnostics using the diagnostic feature vector A_0, which, for example, contains n=5 elements. Then we write down the vector (1) in the form:

$$A_0=\begin{pmatrix} a_1^0 \\ a_2^0 \\ a_3^0 \\ a_4^0 \\ a_5^0 \end{pmatrix}. \quad (2)$$

We denote the diagnostic features that characterize the defect-free condition through a_n^0, where $n=\overline{1,5}$, and let the nominal values of diagnostic features lie in the range of [1,0; 10.0], which is characteristic of dimensionless or normalized diagnostic features [8]. In addition, we will take into account, as in [10], the permissible deviation Δ_0=±5 % from the values of the parameters of a_n^0, with which the object technical condition will be considered defect-free. That is, the values of the elements a_n^0 with the deviation can be taken in the range of [0.95; 1.05] of their nominal values. Given the latter, the diagnostic feature vector (2) takes the form:

$$A_0^*=\begin{pmatrix} [0{,}95;1{,}05]\cdot a_1^0 \\ [0{,}95;1{,}05]\cdot a_2^0 \\ [0{,}95;1{,}05]\cdot a_3^0 \\ [0{,}95;1{,}05]\cdot a_4^0 \\ [0{,}95;1{,}05]\cdot a_5^0 \end{pmatrix}. \quad (3)$$

Let the proposed vector, containing 5 diagnostic features, in general, describe 6 classes of technical condition of the control object:

– the $S0$ class corresponds to the defect-free condition of the control object; this class includes all input sets, for which deviations of diagnostic feature values do not exceed the aforementioned permissible deviation Δ_0;

– the $S1$ class includes input vectors, in which deviations of values of any of the features exceed the permissible deviation Δ_0;

– the $S2$ class includes input vectors, in which deviations of values of simultaneously two any features exceed the permissible deviation Δ_0;

– the $S3$ class includes input vectors, in which deviations of values of simultaneously three any features exceed the permissible deviation Δ_0;

– the $S4$ class includes input vectors, in which deviations of values of simultaneously four any features exceed the permissible deviation Δ_0;

– the $S5$ class includes input vectors, in which deviations of values of simultaneously of all the features exceed the permissible deviation Δ_0.

Thus, the $S0$ class characterizes the defect-free condition of the control object, and the $S1$–$S5$ classes characterize the object condition after the appearance and development of damage. Among the identified TC classes, consideration of the latter two ($S4$ and $S5$) provides a certain theoretical synthesis of the research results. In practice, these classes can characterize rather serious operational irregularities and partial or complete loss of functionality of the control object.

Based on the above, we can formulate the problem of multi-class recognition of the object TC by the developed neural network classifier. In general, the problem lies in the error-free recognition of the $S0$–$S5$ classes of the technical condition according to the multidimensional diagnostic feature vector (3).

3. The aim and objectives of the study

The aim of the work is to analyze the efficiency and to ensure error-free multi-class recognition of the object technical condition by the developed neural network classifier.

To achieve this aim, the following objectives were accomplished:

– to form a set of training diagnostic feature vectors that characterize the $S0$–$S5$ classes of the technical condition, to perform training of the neural network classifier;

– to form sets of test diagnostic feature vectors for the $S0$–$S5$ classes and to check the classifier efficiency;

– to perform an efficiency study of multi-class recognition, depending on the value of the NN influence parameter and dimension of the set of training vectors.

4. Efficiency study of multi-class recognition by classifier based on probabilistic neural network

4. 1. Formation of a set of training vectors

An important stage in the development and efficiency study of the neural network classifier is the formation of sets of training and test images (multidimensional vectors). First, for each of the above classes of TC, we form a set of training vectors according to the above conditions for determining the $S0$–$S5$ classes.

Training vectors of the $S0$ class are:

– the diagnostic feature vector A_0 (2) without taking into account possible deviations of diagnostic feature values (Δ_0=0 %);

– two vectors with maximum permissible deviations ($+\Delta_0\cdot A_0$) and ($-\Delta_0\cdot A_0$);

– various combinations of deviations of elements A_0 in the permissible range of [0.95; 1.05].

For the set of training vectors of each defective condition corresponding to the $S1$–$S5$ classes, the following maximum deviations of diagnostic feature values Δ_M for training were selected and set: ±5.5 %; ±10 %; ±15 %; ±20 %; ±25 %; ±50 %. Taking into account features and values of possible deviations, each set of training vectors for the $S1$–$S5$ classes will consist of all possible combinations of deviations and diagnostic features by classes.

Then, to determine the TC, which is characterized by the $S1$ class, we train the NN on the following set of vectors:

$$P_{1M}^{S_1}=\begin{pmatrix} (1\pm\Delta_M)\cdot a_1^0 \\ a_2^0 \\ \dots \\ a_5^0 \end{pmatrix}; P_{2M}^{S_1}=\begin{pmatrix} a_1^0 \\ (1\pm\Delta_M)\cdot a_2^0 \\ \dots \\ a_5^0 \end{pmatrix}; \cdots P_{nM}^{S_1}=\begin{pmatrix} a_1^0 \\ a_2^0 \\ \dots \\ (1\pm\Delta_M)\cdot a_5^0 \end{pmatrix}. \quad (4)$$

From the expression (4) we form a single training input vector for the $S1$ class:

$$P_{S_1}=\left[P_{1M}^{S_1};\ P_{2M}^{S_1};\ \ldots;P_{nM}^{S_1}\right],$$

which can be written in the matrix form as:

$$P_{S_1}=\Delta_M^{S_1}\cdot A_0, \qquad (5)$$

where $\Delta_M^{S_1}$ is the matrix with the dimension $m_1 \times n$ (n – the number of elements of the column vector A_0; m_1 – the number of possible combinations of deviations and diagnostic features corresponding to the condition $S1$):

$$\Delta_M^{S_1}=\begin{pmatrix} 1\pm\Delta_M & 1 & \cdots & 1 \\ 1 & 1\pm\Delta_M & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1\pm\Delta_M \end{pmatrix}. \qquad (6)$$

For the $S1$ class in the matrix $\Delta_M^{S_1}$, the diagonal elements correspond to the deviations $1\pm\Delta_M$. All other elements of the matrix are equal to 1, indicating the invariance of the corresponding elements of the vector A_0. Taking into account possible deviations of diagnostic feature values for the defect-free condition [0.95; 1.05], individual elements of the matrix $\Delta_M^{S_1}$ will take any value in the range of [0.95; 1.05].

To determine the TC, characterized by the $S2$ class, we train the NN on the following set of vectors:

$$P_{1M}^{S_2}=\begin{pmatrix} (1\pm\Delta_M)\cdot a_1^0 \\ (1\pm\Delta_M)\cdot a_2^0 \\ a_3^0 \\ a_4^0 \\ a_5^0 \end{pmatrix};\ P_{2M}^{S_2}=\begin{pmatrix} (1\pm\Delta_M)\cdot a_1^0 \\ a_2^0 \\ (1\pm\Delta_M)\cdot a_3^0 \\ a_4^0 \\ a_5^0 \end{pmatrix};\cdots$$

$$\ldots P_{5M}^{S_2}=\begin{pmatrix} a_1^0 \\ (1\pm\Delta_M)\cdot a_2^0 \\ (1\pm\Delta_M)\cdot a_3^0 \\ a_4^0 \\ a_5^0 \end{pmatrix};\cdots P_{nM}^{S_2}=\begin{pmatrix} a_1^0 \\ a_2^0 \\ a_3^0 \\ (1\pm\Delta_M)\cdot a_4^0 \\ (1\pm\Delta_M)\cdot a_5^0 \end{pmatrix}. \qquad (7)$$

The set of vectors (7) shows that training takes place on all possible combinations with deviations $1\pm\Delta_M$. As in the previous case, from (7) a single input vector for the $S2$ class is formed:

$$P_{S_2}=\left[P_{1M}^{S_2};\ P_{2M}^{S_2};\ \ldots;P_{nM}^{S_2}\right],$$

which can be written in the matrix form as:

$$P_{S_2}=\Delta_M^{S_2}\cdot A_0, \qquad (8)$$

where $\Delta_M^{S_2}$ is the matrix with the dimension $m_2 \times n$ (n – the number of elements of the column vector A_0; m_2 – the number of possible combinations of deviations and diagnostic features corresponding to the $S2$ condition):

$$\Delta_M^{S_2}=\begin{pmatrix} 1\pm\Delta_M & 1\pm\Delta_M & \cdots & 1 & \cdots & 1 & 1 \\ 1\pm\Delta_M & 1 & \cdots & 1\pm\Delta_M & \cdots & 1 & 1 \\ 1 & 1\pm\Delta_M & \cdots & 1\pm\Delta_M & \cdots & 1\pm\Delta_M & 1 \\ 1 & 1 & \cdots & 1 & \cdots & 1 & 1\pm\Delta_M \\ 1 & 1 & \cdots & 1 & \cdots & 1\pm\Delta_M & 1\pm\Delta_M \end{pmatrix}. \qquad (9)$$

Individual elements of the matrix $\Delta_M^{S_2}$ for the $S2$ condition can take values in the range of [0.95; 1.05], as well as for the matrix $\Delta_M^{S_1}$.

Similarly, to the described method, we form training vectors for the $S3$–$S4$ classes taking into account the above conditions for the class definition.

For the $S5$ class, there is the only possible option of definition of the matrix $\Delta_M^{S_5}$:

$$\Delta_M^{S_5}=\begin{pmatrix} 1\pm\Delta_M \\ 1\pm\Delta_M \\ 1\pm\Delta_M \\ 1\pm\Delta_M \\ 1\pm\Delta_M \end{pmatrix}. \qquad (10)$$

Thus, in general, each diagnostic class $S1$–$S5$, for which simultaneously one or more any features in the training vectors exceed the permissible deviation Δ_0, corresponds to the combination matrix: $\Delta_M^{S_k}$, where k=1,..., 5.

Training vectors for the $S1$–$S5$ classes in the matrix form have a generalized view:

$$P_{S_k}=\Delta_M^{S_k}\cdot A_0,\quad k=\overline{1,5}.$$

The general set of training images for the six diagnostic classes $S0$–$S5$ can be written in the matrix form:

$$P=\left[P_{S_0};\ P_{S_1};\ P_{S_2};\ P_{S_3};\ P_{S_4};\ P_{S_5}\right]. \qquad (11)$$

The rows of the matrix P correspond to the number of diagnostic features, and the number of columns is equal to the number R of input training vectors. Based on the results of the given conditions, R=378 training vectors were formed according to the specified classes. The $S0$ class is trained on R_0=6 vectors, the $S1$ and $S4$ classes – on R_1=R_4=60, the $S2$ and $S3$ classes –on R_2=R_3=120, the $S5$ class – on R_5=12 vectors. On the formed set of training vectors for the six diagnostic classes $S0$–$S5$, the classifier training was conducted based on the probabilistic neural network according to the method described in [10].

4. 2. Formation of a set of test vectors

After the training, it is necessary to check the performance of the developed classifier, for which the following 3 sets of test vectors were formed:

– for the first set of test vectors, the deviation of diagnostic features does not exceed the permissible value (Δ_0=±5 %) of ±2.5 %; this means that testing is performed only for the defect-free condition of an object (class $S0$);

– for the second set of test vectors, deviations of the elements of all input vectors for the $S0$ class are within Δ_0, and diagnostic features of vectors for the $S1$–$S5$ classes have a deviation of ±9 % from the values of A_0. So, for $S1$, any element of the test vector has a deviation of ±9 %, while others do not differ from the diagnostic feature vector A_0 by more than ±5 %. For the $S2$–$S5$ classes, the number of the elements different from A_0 by ±9 % is two, three, etc., according to the class;

– the third set of test vectors for the $S0$–$S5$ classes is formed according to the algorithm of forming the second set of test vectors with an increased value of the deviation of the elements up to ±12 %.

For each of the sets, 84 test vectors that characterize the $S0$–$S5$ classes of the technical condition of an object were formed. The total number of test vectors is 252. The developed classifier trained on the general set of training images (11) has performed correct recognition of all test vectors from the above three sets.

4. 3. Analysis of multi-class recognition efficiency

For the developed neural network classifier and the formed sets of training and test vectors that characterize the multi-classing of the object TC, we will analyze the classification efficiency. Such a study is important for justifying the classifier characteristics, which provide error-free recognition of the condition.

The analysis of recognition efficiency of the object technical condition by the developed classifier will be carried out in 2 stages. The first step is classification, that is, the procedure of assigning the test vectors submitted to the classifier input, to the defined classes $S0$–$S5$, which characterize the object condition. In the second stage, we will evaluate the correct classification of test vectors, depending on the characteristics of the classifier and the set of training vectors.

The efficiency of multi-class recognition will be evaluated by the indicator K, which is determined in percentage as the ratio of the number of correctly classified vectors N_1 to the total number of input vectors N_0. The indicator K is a percentage of the probability of correct classification [8]:

$$K = \frac{N_1}{N_0} \cdot 100\ \%. \tag{12}$$

Let us study the influence of the factors associated with the NN characteristics and training process on the efficiency indicator K.

Study of the influence of the probabilistic neural network parameter spread. As noted in [8, 10], the probabilistic neural network parameter *spread* imposes functional conditions on classification accuracy. In the software implementation of NN, this parameter is related to the mean square deviation of the Gaussian function, which specifies the width of the activation functions of neurons and determines their influence on the estimation of the total probability density. Therefore, the *spread* parameter affects the result of classification, it can take any value in the range of [0; 1], during the network training this value is taken without additional justification. The optimum value of the *spread* parameter is determined experimentally during the network testing and directly in the process of classification of test vectors as such that provides error-free (K=100 %) recognition or with minimum possible errors.

In the previous testing of the NN, *spread*=0.05 was taken. Let us study the dependence of the efficiency indicator K on the value of the influence parameter. In this study, we will change the value of the *spread* parameter in the range of values from 0.01 to 0.1 with an increment of 0.01, and in the range of values from 0.1 to 1 – with an increment of 0.1. As the minimum value, we take *spread*=0.005. For the study, a new set of test vectors with the following deviations of diagnostic feature values δ was formed: ±2.5 %; ±6 %; ±9 %; ±10 %; ±12 %; ±15 %. Some of them (δ=±10 %; ±15 %) coincide with the previously taken deviations Δ_M of features for a set of training vectors, which is done to check the reproducibility of classification results by training vectors. According to such test vectors, recognition and determination of the indicator K are performed.

Study of the influence of the number of training vectors. The size of the radial-basis layer of the PNN depends on the number of images of the training set. On the one hand, the larger the NN size, the longer the network training, which negatively affects the classifier performance in real time. On the other hand, reduction of the number of training vectors can lead to a decrease in the recognition efficiency. Therefore, when developing neural network classifiers, it is important to analyze the effect of the dimension of the set of training diagnostic feature vectors on the classification accuracy in order to determine the possibility of error-free multi-class recognition at a certain minimum number of training vectors.

We will use the value of the influence parameter *spread*=0.05, for which error-free class recognition was provided in the previous test. As described above in paragraph 4.1, first, R=378 training vectors were formed that characterize the classes $S0$–$S5$ of the object TC with the following deviations of diagnostic feature values Δ_M: ±5.5 %; ±10 %; ±15 %; ±20 %; ±25 %; ±50 %.

We will reduce the number of training vectors for the $S1$–$S5$ classes relative to the above value R by removing the values of the set deviations Δ_M according to the following procedure:

– remove Δ_M=±50 %; the NN is trained on the set of training vectors with the following deviations of feature values Δ: ±5.5 %; ±10 %; ±15 %; ±20 %; ±25 %; the total number of training vectors R has decreased to 316;

– remove Δ_M=±50 % and Δ_M=±25 %; the NN is trained on the set of vectors with the deviations of feature values Δ: ±5.5 %; ±10 %; ±15 %; ±20 %; the total number of training vectors is R=254;

– remove Δ_M=±50 %, Δ_M=±25 % and Δ_M=±20 %; the NN is trained on the set of vectors with the following deviations of feature values Δ: ±5.5 %, ±10 %, ±15 %; the total number of training vectors is R=192;

– remove Δ_M=±50 %, Δ_M=±25 %, Δ_M=±20 % and Δ_M=±15 %; the NN is trained on the set of vectors with the deviations of feature values Δ: ±5.5 %, ±10 %; the total number of training vectors is R=130;

– for the NN training, we use training vectors only with one deviation value Δ=±5.5 %, the rest of the values of the set deviations Δ_M are removed; the total number of training vectors R is only 68.

We will conduct testing of the trained neuron network classifier on the set of test vectors with the modified and extended range of diagnostic feature deviations δ: ±2.5 %; ±10 %; ±15 %; ±17 %; ±20 %; ±25 %; ±30 %; ±35 %. Such changes allow examining the classifier efficiency for a wider range of possible deviations of diagnostic feature values for each class of the object technical condition.

5. Results of the study of multi-class recognition efficiency

Fig. 2 shows the graphs of the dependence of the recognition efficiency indicator K on the value of the NN parameter *spread*, obtained by the expression (12) with different deviations δ of test vectors.

As can be seen from the following results:

– with the deviation δ=±2.5 %, the neural network classifier provides error-free classification (K=100 %) in the range of *spread* values from 0.005 to 0.07. At the value of the influence parameter of 0.08, the coefficient K is 89.29 %, and further increase in *spread* leads to a decrease in the efficiency indicator;

– with the deviation $\delta=\pm6$ %, the classifier efficiency is 100 % in the range of values of the influence parameter *spread* from 0.005 to 0.1;

– with the deviation $\delta=\pm9$ %, the classifier provides error-free recognition at the *spread* values in the range from 0.01 to 0.1. Reduction of the influence parameter value negatively affects the classification quality and when *spread*=0.005, a decrease in the coefficient K to 92.86 % is observed;

– the deviation $\delta=\pm10$ % coincides with one of the training values of the deviation Δ_M; error-free classification is provided with the *spread* values in the range from 0.005 to 0.1;

– with the deviation $\delta=\pm12$ %, the coefficient K is equal to 100 % at the influence parameter values in the range from 0.02 to 0.1, and starting with the value of *spread*=0.01, the classifier efficiency significantly deteriorates;

– the deviation of elements of the test set $\delta=\pm15$ % also coincides with one of the training values Δ_M; error-free classification is achieved at the *spread* values in the range from 0.005 to 0.1.

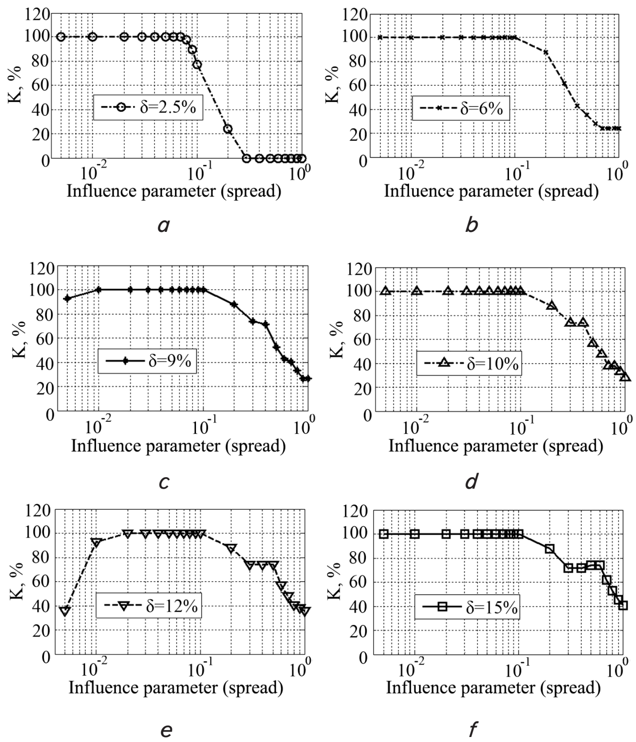

Fig. 2. Graphs of the dependence of the multi-class recognition efficiency indicator K on the probabilistic neural network parameter *spread* for the following values of deviations: *a* – δ=2.5 %; *b* – δ=6 %; *c* – δ=9 %; *d* – δ=2.5 %; *e* – δ=2.5 %; *f* – δ=2.5 %

The results of the study of the dependence of the efficiency indicator K on the dimension of the set of training vectors, conducted using the classifier testing results, have shown:

– for test vectors with deviations of diagnostic feature values $\delta\leq17$ %, the efficiency of recognition of the technical condition according to the $S0$–$S5$ classes is provided at the level of 100 % for all the considered values of R. The indicator K is not decreased even at close diagnostic feature values for defect-free and defective conditions (the difference between the diagnostic feature values of the $S0$ class and $S1$–$S5$ classes did not exceed 0.5 %);

– for test vectors with deviations of diagnostic feature values $\delta>17$ %, there is a decrease in the recognition efficiency with reduction of the number of vectors R in the training set.

The latter case is illustrated in Fig. 3 for the following values of diagnostic feature deviation δ: ±20 %; ±25 %; ±30 %; ±35 %.

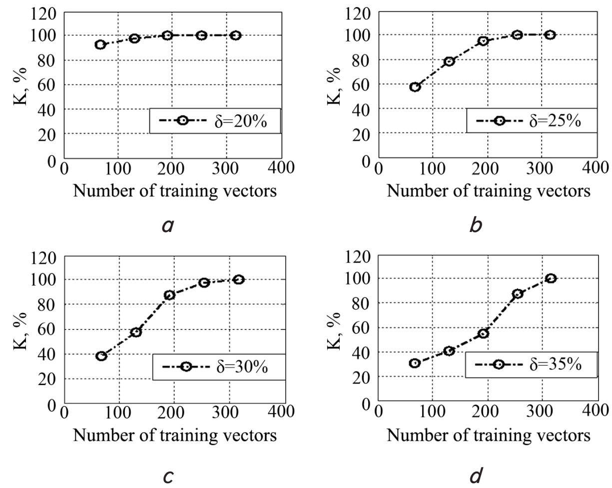

Fig. 3. Graphs of the dependence of the multi-class recognition efficiency indicator K on the number of training vectors for the following deviations: *a* – δ=20 %; *b* – δ=25 %; *c* – δ=30 %; *d* – δ=35 %

As can be seen from the results presented, the error-free recognition (K=100 %) for the entire considered range of deviations δ is provided only when the NN is trained on the set that includes R=316 training vectors. Reduction of the number of training vectors R leads to a decrease in the efficiency indicator (K<100 %) at the following values of diagnostic feature deviations δ:

– for the number of training vectors R=254 with deviations $\delta\geq30$ %;

– for the number of training vectors R=192 with deviations $\delta\geq25$ %;

– for the number of training vectors R=130 with deviations $\delta\geq20$ %;

– for the number of training vectors R=68 with deviations $\delta>17$ %.

6. Discussion of the results of the study of multi-class recognition efficiency

The results of the study of the influence of the probabilistic neural network parameter *spread* on the efficiency indicator K (Fig. 2) have shown the possibility of error-free multi-class recognition of the object condition by the developed classifier. This result is obtained for the entire set of input vectors with different values of diagnostic feature deviation provided that the value of the *spread* parameter lies in the range of [0.02; 0.07]. In general, when the *spread* parameter is increased, the probabilistic neural network will take into account the influence of adjacent neighboring conditions. At very small values of the *spread* parameter, the Gaussian distribution function covers a small number of influences, which affects the classification quality. The limitation of the range of the *spread* parameter values, obtained in the study is associated with the complex influence of the following factors:

– the number of the recognized classes (for the two-class diagnostics, the range of the *spread* parameter values is significantly expanded [8]);

– the dimension of diagnostic feature vectors;

– the differences in the diagnostic feature deviation values within the defined classes for training and test vectors.

The range of the *spread* parameter values that provides error-free recognition is significantly expanded if the deviations of diagnostic feature values in training and test vectors are the same. This is confirmed by the results of the studies, for example, for $\delta=\pm 10$ % and $\delta=\pm 15$ %. However, even in the case of the complex influence of these factors, the obtained results have shown the possibility of error-free multi-class recognition of the object TC by the developed classifier. This is important for monitoring the TC of complex spatial objects with multi-site damage.

The analysis of the obtained results of the influence of the dimension of the set of training vectors on the multi-class recognition efficiency has shown their dependence on the value of deviation δ. The greater the diagnostic feature deviation δ in test vectors, the greater the influence of the dimension of the set of training vectors on the multi-class recognition efficiency. Therefore, in order to ensure error-free (K=100 %) multi-class recognition by input vectors, in which deviations of diagnostic values exceed 17 %, it is necessary to expand the NN training. On the other hand, reduction of the number of training vectors does not lead to a decrease in the multi-class recognition efficiency in case of minimum differences in diagnostic feature values for defect-free ($S0$ class) and defective ($S1$–$S5$ classes) conditions of control objects. This indicates a high sensitivity of the probabilistic neural network when performing image comparison and classification. Such a classifier can provide high efficiency of recognition of changes in the TC of control objects at early stages of multi-site damage development in cases of incomplete information about the recognized images and a limited number of training images.

7. Conclusions

1. For recognition, multidimensional diagnostic feature vectors that characterize the defect-free ($S0$ class) and defective ($S1$–$S5$ classes) object conditions are used. Division into classes is carried out depending on the diagnostic feature deviation value and the number of features in the vector, the deviation of which exceeds the permissible value $\Delta_0=\pm 5$ %. Formation of a set of training diagnostic feature vectors with a wide-range feature value deviations is carried out. The formed training vectors correspond to the six classes $S0$–$S5$, followed by the neural network “training” for setting the network parameters.

2. Three sets of test vectors, in which the maximum diagnostic feature deviation value does not exceed ±5 %, ±9 % and ±12 %, respectively, are formed. The developed classifier, trained on the general set of training images has correctly performed recognition of all test vectors.

3. The study of the multi-class recognition efficiency, depending on the characteristics of the neural network and the set of training vectors, is conducted. It is found that the developed classifier provides error-free multi-class recognition of test vectors, if the value of the network influence parameter *spread* is in the range of [0.02; 0.07]. The minimum sizes of the set of training vectors and limit values of diagnostic feature deviations in test vectors, which provide error-free multi-class recognition, are determined. It is revealed that reduction of the number of training vectors does not lead to a decrease in the multi-class recognition efficiency in case of small (less than 0.5 %) differences in diagnostic feature values for defect-free and defective conditions of control objects.

References

1. Shanyavskiy, A. A. Bezopasnoe ustalostnoe razrushenie elementov aviakonstrukciy [Text]: monografiya / A. A. Shanyavskiy. – Ufa, 2003. – 802 p.
2. Goranson, U. G. Damage tolerance. Facts and fiction [Text] / U. G. Goranson // Keynote Presentation in Int. Conf. on Damage Tolerance of Aircraft Structure. – Delft Technical University, 2007. – 58 p. – Available at: http://dtas2007.fyper.com/userfiles/file/Paper%2011_Goranson.pdf
3. Ignatovich, S. R. Ocenka dostovernosti kontrolya treshchin v aviacionnyh konstrukciyah [Text] / S. R. Ignatovich, L. A. Humaryan // Visnyk Ternopilskoho natsionalnoho tekhnichnoho universytetu. – 2013. – Vol. 72, Issue 4. – P. 117–124.
4. Nagarajaiah, S. Structural monitoring and identification of civil infrastructure in the United States [Text] / S. Nagarajaiah, K. Erazo // Structural Monitoring and Maintenance. – 2016. – Vol. 3, Issue 1. – P. 51–69. doi: 10.12989/smm.2016.3.1.051
5. Bouraou, N. I. Investigation of the model of the vibration measuring channel of the complex monitoring system of steel tanks [Text] / N. I. Bouraou, S. A. Cybul’nik, D. V. Shevchuk // Eastern-European Journal of Enterprise Technologies. – 2015. – Vol. 5, Issue 9 (77). – P. 45–52. doi: 10.15587/1729-4061.2015.50980
6. Ali, Y. H. Acoustic Emission Signal Analysis and Artificial Intelligence Techniques in Machine Condition Monitoring and Fault Diagnosis: A Review [Text] / Y. H. Ali, R. Abd Rahman, R. I. Raja Hamzah // Jurnal Teknologi. – 2014. – Vol. 69, Issue 2. doi: 10.11113/jt.v69.3121
7. Bielza, C. Multi-dimensional classification with Bayesian networks [Text] / C. Bielza, G. Li, P. Larrañaga // International Journal of Approximate Reasoning. – 2011. – Vol. 52, Issue 6. – P. 705–727. doi: 10.1016/j.ijar.2011.01.007
8. Bouraou, N. I. Rozpiznavannia tekhnichnoho stanu ob‘yektiv na osnovi shtuchnykh neironnykh merezh [Text]: monohrafiya / N. I. Bouraou, O. V. Zazhytskyi. – Kyiv: NAU, 2014. – 120 p.
9. Hoła, J. Non-destructive and semi-destructive diagnostics of concrete structures in assessment of their durability [Text] / J. Hoła, J. Bień, Ł. Sadowski, K. Schabowicz // Bulletin of the Polish Academy of Sciences Technical Sciences. – 2015. – Vol. 63, Issue 1. doi: 10.1515/bpasts-2015-0010

10. Balageas, D. Structural Health Monitoring [Text] / D. Balageas, C.-P. Fritzen, A. Gemes. – John Wiley & Sons, 2006. – 495 p. doi: 10.1002/9780470612071
11. Dworakowski, Z. Application of Artificial Neural Networks for Damage Indices Classification with the Use of Lamb Waves for the Aerospace Structures [Text] / Z. Dworakowski, L. Ambrozinski, P. Packo, K. Dragan, T. Stepinski, T. Uhl // Key Engineering Materials. – 2013. – Vol. 588. – P. 12–21. doi: 10.4028/www.scientific.net/kem.588.12
12. Shen, T. Damage location and identification of the wing structure with Probabilistic Neural Networks [Text] / T. Shen, F. Wan, B. Song, Y. Wu // 2011 Prognostics and System Health Managment Confernece. – 2011. doi: 10.1109/phm.2011.5939524
13. Palomino, L. V. Probabilistic Neural Network and Fuzzy Cluster Analysis Methods Applied to Impedance-Based SHM for Damage Classification [Text] / L. V. Palomino, V. Steffen, R. M. Finzi Neto // Shock and Vibration. – 2014. – Vol. 2014. – P. 1–12. doi: 10.1155/2014/401942
14. Bouraou, N. I. Syntez neironnoi merezhi dlia bahatoklasovoi diahnostyky elementiv konstruktsiy v ekspluatatsiy [Text] / N. I. Bouraou, A. H. Protasov, P. S. Myronenko, S. S. Rupich // Metody ta prylady kontroliu yakosti. – 2015. – Issue 2 (35). – P. 83–93.

UDC 004.62:004.023
DOI: 10.15587/1729-4061.2017.113046

IMPROVING EFFICIENCY OF PROVIDING DATA GROUP ANONYMITY BY AUTOMATING DATA MODIFICATION QUALITY EVALUATION

O. Chertov
Doctor of Technical Sciences, Associate Professor*
E-mail: chertov@i.ua
D. Tavrov
PhD*
E-mail: dan.tavrov@i.ua
*Department of Applied Mathematics
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
Peremohy ave., 37, Kyiv, Ukraine, 03056

Запропоновано модифікацію методу розв'язання задачі забезпечення групової анонімності на основі міметичного алгоритму, яка не передбачає участі експерта на етапі оцінювання розв'язків задачі. Автоматизація оцінювання розв'язків підвищує ефективність процесу групової анонімізації даних. Модифікацію методу проілюстровано шляхом розв'язання задачі анонімізації на основі реальних даних

Ключові слова: міметичний алгоритм, групова анонімність, мікрофайл, викид, модифікований метод тау Томпсона

Предложена модификация метода решения задачи обеспечения групповой анонимности на основе меметического алгоритма, которая не предусматривает участия эксперта на этапе оценивания решений задачи. Автоматизация оценивания решений повышает эффективность процесса групповой анонимизации данных. Модификация метода проиллюстрирована путем решения задачи анонимизации на основе реальных данных

Ключевые слова: меметический алгоритм, групповая анонимность, микрофайл, выброс, модифицированный метод тау Томпсона

1. Introduction

Around the world, amounts of digital data keep increasing with each year. A great share of these data are published in open access in their primary, non-aggregated form. Such data sets are called *microdata*. Microdata can be used for numerous purposes, including:

– dissemination of clinical data to facilitate medical research. E. g., in the U.S., this is regulated by the corresponding bills [1, 2];

– enforcing transparency of public policy. E. g., in the EU, protection of personal data is subject to the corresponding law [3];

– sharing census and other statistical research data to enable conducting economic, demographic, and other kinds of research.

At the same time, there is a certain risk that providing public access to the data in their unchanged form will not only achieve its primary goal but also lead to disclosing confidential information about an individual or a group thereof. E.g., open access to clinical data facilitates medical research. At the same time, publishing medical records can enable unique identification of a patient. Moreover, outliers in a regional distribution of patients might point to areas with exceeded sickness rate threshold.

Therefore, it is important to provide data anonymity at the stage of creating the content for open information resources. Anonymity of a subject can be seen as its property of being not identifiable (uniquely characterized) within a set of subjects [4]. Anonymity comes in two variants:

– *individual anonymity* concerns information about single respondents (persons, households, enterprises);

– *group anonymity* concerns distribution of information about a group of respondents.

Methods for providing individual anonymity have been a subject of research for more than 20 years and are developed