Computer modelling of statistical, structural and neural network methods of pattern recognition

N M Novikova
Department of Applied Mathematics, Informatics and Mechanics, Voronezh State University, 1, University Square, Voronezh, 394018, Russia
E-mail: nov.nelly@gmail.com

Abstract. Pattern recognition is one of the most important tasks of both intelligent control systems and artificial intelligence. The paper substantiates the relevance of the study of computer modelling of statistical, structural and neural network methods of pattern recognition. The article presents a comparative analysis of the quality of pattern recognition using the Hamming neural network and the statistical algorithm. The analysis shows that the use of the Hamming neural network is preferable in most cases. Computer modelling of the structural method using the Freeman code gives a description that allows us to unambiguously assign an object to its class. Based on the analysis of the results of computer modelling, the positive and negative aspects of each method are revealed. As a result, the structural method is the most optimal.

1. Introduction
Pattern recognition is one of the most interesting and relevant problems of cybernetics. This problem is closely connected with research in the field of artificial intelligence, with construction of diagnostic systems in industry and medicine, with solution of many problems of administration, radiolocation and communication. A special place in these tasks is occupied by pattern recognition because when we use images, we can compactly represent a large amount of information. Currently, much attention is paid to pattern recognition using neural networks.

Cognitron-based face recognition is considered in [1]. The recognition process took a lot of time and turned out to be complicated and time-consuming.

The work [2] is devoted to the study of pattern recognition using a convolutional neural network and a fuzzy hybrid classifier. With such a neural network architecture, a convolutional neural network is responsible for image preprocessing. The fuzzy logic network does a good job of making the final classification decision. It should be noted that an essential condition for the successful operation of convolutional neural networks is the use of large data samples at the learning stage. The obvious consequence of working with large amounts of data is significant computational cost.

In [3], the basic principles of constructing multilayer perceptrons using the example of digit recognition are described. The results of architecture operation depending on the number of learned parameters are analyzed.

In [4], multilayer neural-like structures and algorithms of their dynamic synthesis for pattern recognition tasks are studied. The results in the field of learning and before training...
convolutional neural networks using growth algorithms of their structures are presented. It is shown that such networks have stable noise immunity to pattern recognition. In [5], the comparative analysis of pattern recognition quality indicators using statistically optimal and neural network algorithms is considered. The results of a computational experiment on the recognition of complex images with large similarity measure value are presented. In [6], the computer model of statistical pattern recognition is presented. The image similarity measure is introduced, and its values are computed for pairs of images. The method for removing the singularity of covariance matrices is shown. The algorithm of the recognition device is given. In a statistical experiment on a computer, the estimates of the probabilities of errors of the first and second kind are obtained. Investigations of the behaviour of these estimates for various sizes of control and learning samples, dimensions of the feature space, and similarity measures of recognized images are made.

In these papers, the discriminant approach is used when, during recognizing images, each object is associated with a point in a certain space. The class of similar objects then seems to be a compact set of such points, and the problem is to construct discriminant or probabilistic decision rules relating the newly appearing points to a particular image (class). When we use the structural approach, another situation arises — the construction of a short description of one image or class of images — a pattern. The problem arises when it is necessary to compactly encode images for subsequent reproduction of the basic structural characteristics of the image, and not its details. This problem is closely related to the problem of pattern recognition, i.e. if an object can be replaced by its description, then the description elements can be used as features of the object, and the description itself can be used as the basis for constructing a recognition space. The objects considered in this class of problems are often complex, and the number of required features is very high. Therefore, the idea of describing a complex object in the form of a hierarchical structure of simpler subimages is used. Therefore, the basis for describing the image is information about the structure of the objects, and the recognition procedure additionally describes the differences between objects, excluding recognition errors.

It should be noted that despite the large number of works on pattern recognition, there are no works on structural recognition and analysis of various recognition methods.

The purpose of the paper is the construction of computer models based on known algorithms and the analysis of these models in order to identify the most optimal pattern recognition methods.

2. Materials and techniques
The problem statement was the same for all methods. The statement of the problem is as follows: there are images of objects that make up the set A and a number of images of objects related to the set B. The objects of the set B have a certain number of features identical to those of the set A. Let us call the images of the objects of the set A the main targets, and the images of the objects of the set B — the traps. In the process of recognition, one object from the set A and one object from the set B are presented.

When constructing computer models for recognition, images of ships shown in figures 1–4 are presented, one of which is the target, and three images are the false targets or traps.

![Figure 1. Target](image1.png) ![Figure 2. Trap 1](image2.png)
When modelling both statistical and neural network recognition methods, it should be noted that all traps have different measures of similarity with the target. The similarity measure refers to the number of identical features and the degree of coincidence of similar features. In metric space, such measure can be any non-increasing distance function. To calculate the similarity measure, it is necessary to represent images with vectors in n-dimensional space.

The reciprocal of the Euclidean distance between a pair of \( n \)-dimensional vectors is chosen as similarity measure:

\[
L(X_k, X_i) = \left( \sum_{j=1}^{n} (x_{jk} - x_{ji})^2 \right)^{-1}
\]

Based on the statistical theory of pattern recognition, the following recognition algorithm can be constructed.

1. Feature space formation.
2. Learning.
3. Decision making.
4. Computation of recognition probability.

**Feature space formation**

Images are presented for recognition in the form of a certain set of observations written in the form of the matrix \( X \).

Each column of the matrix \( x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T \) is a \( p \)-dimensional vector of the observed values of the features \( X_1, X_2, \ldots, X_p \).

The totality of features should most reflect those properties of images that are important for recognition. The computational complexity of the recognition procedures and the validity of the results depend on the dimension of the feature space. The procedures for selecting features and forming feature space are highly dependent on a specific task.

Since the recognition images are two-level ones and presented in the "bitmap" format, we can work with them as with matrices of 0 and 1, where 1 is a black dot, 0 is a white one. This matrix will be a matrix of features. Recognition of images in grayscale and especially in color complicates the task, significantly increasing the dimension of the matrix of features. This complicates the programming and does not affect the purpose of the study.

**Learning**

The source of information about recognizable images is the totality of results of independent observations (sampled values) that make up the learning and control samples. Under the conditions of the problem under consideration, the samples are \( p \)-dimensional random variables.

Since we are dealing with a normal multidimensional distribution, the prior uncertainty will relate only to the parameters of this distribution, so the goal of learning in this case is to obtain estimates of these parameters. It should be noted that there is a distinction in case of supervised learning, as a result of which a classified learning sample is obtained. Regarding the elements of the sample, it is known to which distribution law they belong. As a result of unsupervised
learning, estimates are formed with respect to unclassified samples. Thus, given the normal multidimensional distribution, it is enough to compute the maximum likelihood estimation of the vector of the mean and covariance matrices based on the classified learning samples. These estimates are computed according to the formulas (1) and (2), respectively [7].

\[ m_j = \frac{1}{N_j} \sum_{x \in \omega_j} x, \quad (1) \]

\[ C_j = \frac{1}{N_j} \sum_{x \in \omega_j} xx^T - m_j m_j^T. \quad (2) \]

**Decision making**

The selection of a decision rule that allows us to assign the control sample of observations to one of the mutually exclusive classes is made in accordance with the theory of statistical decisions, considering a priori information and learning dataset. According to this theory, all decision rules for \( k > 2 \) classes are based on a comparison of likelihood relations between themselves or with a certain threshold \( R \), the value of which is determined by the selected quality criterion.

The decision rule is as follows: the sample \((x_1, x_2, \ldots, x_n)\) refers to the \(p\)-dimensional normal law with the parameters \((m_2, C_2)\) if the inequality (3) holds [7].

\[ E_n = \frac{1}{2} \sum_{i=1}^n \left[ (x_i - m_1)^T C_1^{-1} (x_i - m_1) - (x_i - m_2)^T C_2^{-1} (x_i - m_2) \right] + \frac{n}{2} \ln \left( \frac{\det C_1}{\det C_2} \right) \geq \ln R, \quad (3) \]

where \( E \) is the estimate of the likelihood ratio logarithm;

\( \ln R \) is the threshold; in recognition, we shall use \( R = 1 \).

The matrices \( C_1 \) and \( C_1 \) can be degenerate. Note that in this case there are no inverse matrices for them. Therefore, in the computations, we shall use pseudoinverse (generalized inverse) matrices instead of simple inverse matrices. A pseudoinverse matrix coincides with the inverse one if the inverse matrix exists.

The matrix \( A^+ \) is called a pseudoinverse matrix for the matrix \( A \) if it satisfies the following criteria:

1. \( AA^+ A = A \).
2. \( A^+ AA^+ = A^+ \).
3. \( (AA^+)^* = AA^+ \).
4. \( (A^+ A)^* = A^+ A \).

Under these conditions * denotes the Hermitian conjugation, the Hermitian conjugate matrix for a matrix whose elements are real numbers coincides with the transpose of a matrix.

Consider one of the methods for finding a pseudoinverse matrix – the Ben-Israel method. This method is iterative, and it allows us to compute the pseudoinverse of a given matrix with some degree of accuracy. The number of iterations necessary to obtain the result is not known in advance and the iterative process continues until the stop criterion is met [8].

The method is based on the following recursive formula:

\[ X^{(k+1)} = 2X^{(k)} - X^{(k)}AX^{(k)}. \]

In this case, the recursion basis is defined as follows:

\[ X(0) = \alpha A, \]

where \( \alpha \) is the number for which

\[ 0 \leq \alpha \leq \frac{2}{\lambda_{\text{max}}}, \]
where $\lambda_{\text{max}}$ is the maximum eigenvalue of the matrix $A$.

As a result of iterations, the matrix $X$ converges in the norm to the pseudoinverse of $A$, and we can assume that $X^{(n)} = A^+$. To search for the matrix determinant, we use the Gauss method, in which the matrix is reduced to the upper-triangular form, and then the determinant is calculated as the product of diagonal elements.

### Hamming neural network

Neural networks, like statistical recognition algorithms, use the object feature space for recognition.

Among the various configurations of artificial neural networks (NN), there are those whose classification according to the principle of learning, strictly speaking, does not suit supervised learning or unsupervised learning. In such networks, synapse weighting factors are computed based on information about the processed data only once before the start of network operation, and all network learning comes down to this calculation [9]. Of the networks with similar behaviour, the most famous are the Hopfield network and the Hamming network, which are usually used to organize associative memory.

The idea of the network operation is to find the Hamming distance from the tested image to all samples. The Hamming distance is the number of distinct bits in two binary vectors. The network must select a sample with a minimum Hamming distance to an unknown input signal, as a result of which only one network output corresponding to this sample will be activated.

At the initialization stage, the weighting factors of the first layer and the threshold of the activation function are assigned the following values:

$$w_{ik} = \frac{x^k_i}{2}, \quad i = 0 \ldots n - 1, \quad k = 0 \ldots m - 1,$$

$$T_k = \frac{n}{2}, \quad k = 0 \ldots m - 1.$$

Here $x^k_i$ is the $i$-th element of the $k$-th sample.

The weighting factors of inhibitory synapses in the second layer are taken equal to the certain value $0 < \varepsilon < 1/m$. The synapse of a neuron linked with its own axon has a weight of $+1$.

The Hamming network operation algorithm is as follows:

1. The unknown vector $X = \{x_i : i = 0 \ldots n - 1\}$ is fed to the network inputs, based on which the states of the neurons of the first layer are calculated (the upper index in parentheses indicates the layer number):

$$y_j^{(1)} = s_j^{(1)} = \sum_{i=0}^{n-1} w_{ij} x_i + T_j, \quad j = 0 \ldots m - 1.$$

After that, the values of the axons of the second layer are initialized with the obtained values:

$$y_j^{(2)} = y_j^{(1)}, \quad j = 0 \ldots m - 1.$$

2. Compute new states of neurons of the second layer:

$$s_j^{(2)}(p + 1) = y_j(p) - \varepsilon \sum_{k=0}^{m-1} y_k^{(2)}(p), \quad k \neq j, \quad j = 0 \ldots m - 1$$

and the values of their axons:

$$y_j^{(2)}(p + 1) = f\left[s_j^{(2)}(p + 1)\right], \quad j = 0 \ldots m - 1.$$
The activation function $f$ has the form of a threshold, and the value $F$ should be large enough so that any possible values of the argument do not lead to saturation.

3. Check whether the outputs of the neurons of the second layer have changed during the last iteration. If yes, go to step 2. Otherwise — the end.

It can be seen from the evaluation of the algorithm that the role of the first layer is very conditional: using the values of its weighting factors once in step 1, the network no longer addresses it, therefore the first layer can be completely excluded from the network (replaced by the matrix of weighting factors) [9]. At the learning stage, the calculation of weighting factors represents finding the mean value of a given element of the feature vector over the entire learning sample.

**Structural recognition method**

The structural approach is applied to pattern recognition problems in which information describing the structure of each object is important. And the recognition procedure is required to make it possible not only to assign the object to a certain class (to classify it), but also to describe those aspects of the object that exclude its assignment to another class.

In order to represent hierarchical (tree-like) structural information contained in each image, i.e. describe the image using simple subimages, and again describe each subimage with even simpler subimages, a syntactic or structural approach was proposed. This approach is based on the analogy between the structure of images (hierarchical or tree-like) and on the syntax of languages. Within the framework of this approach, it is assumed that images are constructed from subimages connected in various ways. Obviously, this approach is only useful if it is easier to recognize the selected simplest subimages, called non-derived elements, than the images themselves. A "language" that provides a structural description in terms of a set of non-derived elements and operations of compositions of these elements is sometimes called an "image description language." The rules for the composition of non-derived elements are usually set using the so-called grammar of the image description language. The recognition process is carried out after identifying non-derived elements in the object and providing a description of the object. Recognition consists in syntactic analysis or parsing of the "sentence" describing a given object. This procedure determines whether this sentence is syntactically correct as related to a given grammar. In parallel, parsing provides some structural description of the sentence (usually in the form of a tree-like structure) [10, 11].

The main stages of structural recognition:

1. Preprocessing.
2. Segmentation.
3. Selection of non-derived elements.
4. Representation of the image.
5. Decision making.

### 3. Results and discussion

Processing the results of computer modelling was carried out for all the considered methods of pattern recognition. The probability of correct recognition and errors was determined by the formula

$$P = \frac{n}{N},$$

where $n$ is the number of experiments, the result of which was the correct assignment of the object to the class, and $N$ is the total number of experiments.

In the statistical method, it is currently not possible to find error probabilities of recognition of $\alpha$ and $\beta$ in a final analytical or quadrature form, therefore, to obtain estimates $\hat{\alpha}$ and $\hat{\beta}$ of probabilities of errors $\alpha$ and $\beta$ for arbitrary values of the learning sample sizes, it is advisable to
resort to a statistical experiment using the Monte Carlo method [12]. In a statistical experiment on a computer, estimates of the probabilities of errors of the first and second kind are obtained. Investigations of the behaviour of these estimates for various sizes of control and learning samples, dimensions of the feature space, and the similarity measures of recognized images were made.

The similarity measure between Trap 1 and Target is 0.02016, between Trap 2 and Target — 0.01247, between Trap 3 and Target — 0.010401. It is worth noting that the greater the similarity measure, the more the feature spaces of the two images intersect, and this increases the probability of error of the first kind, that is, an image from the ”target” class can be mistakenly assigned to the ”trap” class.

When constructing a computer model of the Hamming neural network, a learning set consisting of vectors defined on the feature space was used. The sample size varies from 10 to 60. Each element of the learning sample feature vector consists of 4800 elements. These components are fed to the input of 4 neurons of the Hamming network, so the network learning requires the calculation of 19,200 coefficients.

The results of computer modelling of statistical and neural network pattern recognition methods are presented in tables 1–3.

Table 1. Dependence of the probability of correct recognition of the dimension of the feature vector.

| Feature vector dimension | The probability of correct recognition |
|--------------------------|---------------------------------------|
|                          | Statistical   | Neural network |
| 3                        | 0.53          | 0.725          |
| 5                        | 0.57          | 0.81           |
| 7                        | 0.75          | 0.94           |
| 9                        | 0.96          | 0.98           |
| 11                       | 0.99          | 0.99           |
| 13                       | 0.99          | 0.99           |

| 3            | 0.53          | 0.883          |
| 5            | 0.56          | 0.89           |
| 7            | 0.73          | 0.95           |
| 9            | 0.86          | 0.96           |
| 11           | 0.98          | 0.99           |
| 13           | 0.99          | 0.99           |

| 3            | 0.52          | 0.70           |
| 5            | 0.55          | 0.86           |
| 7            | 0.68          | 0.99           |
| 9            | 0.89          | 0.99           |
| 11           | 0.97          | 0.99           |
| 13           | 0.99          | 0.99           |
Table 2. Dependence of the probability of error of the first kind on the size of the learning sample.

| Learning sample size | Target – Trap 1          | Target – Trap 2          | Target – Trap 3          |
|----------------------|--------------------------|--------------------------|--------------------------|
|                      | Statistical: 0.2          | Statistical: 0.43        | Statistical: 0.4          |
|                      | Neural network: 0.04      | Neural network: 0.01     | Neural network: 0.01      |
| 10                   |                          |                          |                          |
| 20                   | 0.035                    | 0.285                    | 0.14                     |
| 30                   | 0.0001                   | 0.18                     | 0.03                     |
| 40                   | 0.0001                   | 0.1                       | 0.01                     |
| 50                   | 0.0001                   | 0.05                     | 0.01                     |
| 60                   | 0.0001                   | 0.02                     | 0.001                    |

From the analysis of the results of computer modelling presented in table 1, it can be seen that the Hamming neural network algorithm provides a greater probability of correct recognition than the statistical one for all dimensions of feature vectors. An analysis of the data given in table 2 shows that the Hamming neural network algorithm provides a lower probability of error of the first kind for all learning sample sizes for three pairs of images.

Table 3 shows the dependences of the probability of an error of the second kind on the size of learning samples for three pairs of images. It can be seen from the above data that the Hamming neural network algorithm provides a lower probability of error of the second kind for all learning sample sizes for three pairs of images.

When processing the results of computer modelling, it was found that the data for structural pattern recognition have similar structures, namely, the ships have aft and bow parts, as well as a deck superstructure. In a detailed study of the images, it was revealed that significant differences between them exist mainly in deck superstructures, so in the future we shall consider only them. To do this, we divide the ship into its three component parts and select the central one.

The Freeman chain code was chosen as non-derived elements. Chain codes are used to represent a boundary as a sequence of straight-line segments of a certain length and direction. This representation is based on a 4- or 8-connected grid. The length of each segment is determined by the resolution of the grid, and the directions are set by the selected code. In this
Table 3. Dependence of the probability of error of the second kind on the size of the learning sample.

| Learning sample size | Statistical | Neural network |
|----------------------|-------------|----------------|
| 10                   | 0,32        | 0,01           |
| 20                   | 0,035       | 0,01           |
| 30                   | 0,025       | 0,001          |
| 40                   | 0,005       | 0,0001         |
| 50                   | 0,0001      | 0,0001         |
| 60                   | 0,0001      | 0,0001         |

Target – Trap 2

| Learning sample size | Statistical | Neural network |
|----------------------|-------------|----------------|
| 10                   | 0,385       | 0,01           |
| 20                   | 0,25        | 0,001          |
| 30                   | 0,013       | 0,001          |
| 40                   | 0,08        | 0,001          |
| 50                   | 0,006       | 0,0001         |
| 60                   | 0,0001      | 0,0001         |

Target – Trap 3

| Learning sample size | Statistical | Neural network |
|----------------------|-------------|----------------|
| 10                   | 0,28        | 0,01           |
| 20                   | 0,06        | 0,001          |
| 30                   | 0,01        | 0,0001         |
| 40                   | 0,001       | 0,0001         |
| 50                   | 0,001       | 0,0001         |
| 60                   | 0,0001      | 0,0001         |

case, an 8-connected grid is used (figure 5), and the length of each segment is 1. The length 1 is chosen so that objects of different sizes can be recognized.

Figure 5. Freeman chain code
After combining successive equal segments into one, we obtained the following chain codes.

1. Target:
   7-1-7-1-3-1-7-1-7-1-7-5-7-1-3-1-3-5-3-1-7-1-3-1-3-1-7-1-3-1-3-1-3-1-3-1.

2. Trap 1:
   7-1-7-1-7-1-7-1-7-1-7-1-3-5-3-5-3-1-7-1-3-1-3-1-7-1-3-1-3-1-3.

3. Trap 2:
   1-7-5-7-1-7-1-7-1-3-5-3-1-7-1-3-1-3-1.

4. Trap 3:
   7-1-7-1-7-1-7-1-7-1-7-1-7-1-7-3-5-1-3-5-3-1-3-1-3-1-7-1-3-1-3-1-3-1.

We see that in this case it was possible to use chain code that uses 4 directions, but the programme was written for the general case.

Having analyzed the obtained chains, we can describe the presented objects. The target has a wide deck superstructure with many ”steps”. Trap 2 has the narrowest deck superstructure. Trap 1 has fewer ”steps” in the aft superstructure, and Trap 3 has fewer steps in the bow part. This description allows us to unambiguously assign an object to its class.

Let us consider the difficulties that arise in computer modelling of statistical and neural network methods.

Firstly, in computer modelling at the learning stage, it is necessary to use learning samples of large sizes, which leads to large computational costs.

Secondly, it should be noted that these methods are based on the identification of features of an object for recognition, which is a very difficult task. In addition, the features of the object are represented in the form of matrices.

Thirdly, in computer modelling of the statistical method, difficulties arise in working with matrices, which must be converted into pseudoinverse matrices using iterative methods. This complicates the programming, which also increases the computational cost.

Fourthly, computer models of statistical and neural network methods use similarity functions, which also complicates the programme operation.

There are no such difficulties in computer modelling of the structural method. The structural method not only allows you to assign the object to a certain class of images, but also allows you to obtain information that describes the structure of the object. The structural approach is based on the analogy between the structure of images (hierarchical or tree-like) and the syntax of languages. The advantages of this method include the fact that with the successful selection of non-derived elements, recognition does not depend on the size of the objects.

Based on the analysis of computer models of statistical, neural network and structural methods, we can conclude that the structural method is easily programmed, it has no computational costs, it is reliable. This is especially important for computer modelling of artificial intelligence.

4. Conclusion

Computer models of statistical, neural network and structural methods for pattern recognition are presented. Two main groups of recognition methods are distinguished: methods based on the selection of features of an object using discriminant or probabilistic decision rules and structural methods. The structural method not only allows you to assign the object to a certain class of images, but also allows you to obtain information that describes the structure of the object. The algorithms of statistical and neural network methods and the results of computer modelling are presented. The dependences of the probability of correct recognition on the dimension of the feature vector, the dependences of the probability of an error of the first and second kind on the size of the learning sample are obtained. The analysis shows that the use of the Hamming neural network in most cases is preferable. The description, which makes it possible to unambiguously assign an object to its class is obtained for the structural method. A comparative analysis of the
Results showed that it is most preferable to use the structural method for recognizing complex images, since it is distinguished by the simplicity of implementation and the reliability of the data obtained.

References
[1] Kozlov P V and Yuzhakov A A 2014 Neurocomputers: development, application 12 57–61
[2] Dudenkov V M and Novikova N M 2015 Neurocomputers: development, application 2 43–47
[3] Fedorenko Y S and Gananyuk Y E 2017 Neurocomputers: development, application 2 24–30
[4] Stepanyan N V and Znep N N 2018 Neurocomputers: development, application 5 4–11
[5] Novikova N M and Prasolova A E 2004 Neurocomputers: development application 4 28–33
[6] Novikova N M and Noaman S A 2012 Scientific statements of BelSU 13 (132) 135–141
[7] Gonzalez R and Woods R 2005 Digital image processing (Moscow: Technosphere)
[8] The main methods for finding pseudo-inverse matrices URL http://hub.hexlet.org/blog/algebra/55.html (accessed 03.05.2013)
[9] Khai kin S 2006 Neural networks: full course (Moscow: Williams)
[10] Fu K 1977 Structural methods in pattern recognition (Moscow: Mir)
[11] Novikova N M and Budko V N 2011 Intelligent interfaces (Voronezh: Publishing and Polygraphic Center of VSU)
[12] Fomin A Y and Tarlovsky G R 1986 Statistical theory of pattern recognition (Moscow: Radio and communications)