The Second International Conference "Problems of Cybernetics and Informatics"
September 10-12, 2008, Baku, Azerbaijan. Section #2 "Intellectual Systems"
www.pci2008.science.az/2/24.pdf

PROGRAM COMPLEX OF THE PATTERN RECOGNITION WITH THE HELP NEURAL NETWORKS – MIRANDA MP

Dmitry Bagayev[1], Pavel Khrustalev[2]

Kovrov State Technological Academy named after V.A. Degtyareva, Kovrov, Russia
[1]*dmitry_bag@mail.ru*, [2]*pav405@yandex.ru*

Now the computer industry explicates with fast enough rates. Huge significance is received by systems of a pattern recognition based on the theory of neural networks. Neural networks (NN) exist in many, excellent from each other variants. In the given paper the most popular NN for the various reasons – multilayer perceptron, trained on algorithm of back propagation. This algorithm for tutoring multilayer perceptron for the first time became officially known owing to operation Rumenharts, Hilton and Williams in 1986.

The purpose of carried out research consist in study, some correction and possible finishing of the optimized methods of tutoring of multilayer neural networks and application of the received results for construction of a program complex of a pattern recognition. As plants of discernment 10 digits (from 0 up to 9 inclusively) have been selected.

According to a task in view were 2 multilayer NN with various parameters are designed. The first for discernment of invariant digits to turn of several various fonts. The second for discernment of the hand-written digits entered in set area.

The NN executes nonlinear multi-dimensional transformation of an input vector x on an output vector y. Dimension of the last in this case is equal 10. Considering the aforesaid, one of webs has been selected with following parameters: the output layer contains 10 neurons - images meeting amount which the web should distinguish. After testing working capacity and productivity of a web it has appeared, that the optimum amount of neurons in the latent layer makes 75. Dimension of an input vector reflected height and width of the treated image in pixels and was equaled 5029. The above described NN intended for discernment of invariant digits to turn of several various fonts, is displayed on figure 1.

Other NN was constructed for the decision of more simple problem and as investigation had rather more simple structure: $x_M = 3500$, $N_L = 50$, $N_K = 10$.

The lead analysis has displayed, that for the decision of a task in view use of a web with the big amount of layers not expediently as thus there is a big interest of an error, the multilayer neural networks with one latent layer which has displayed the best in this case result of convergence to minimum value of an error therefore has been applied. As algorithm of tutoring of NN has been selected, as it was spoken above, the algorithm of inverse distribution of an error, tutoring of NN on which is carried out in exactitude as shown in figure 2.

Below we shall illustrate some elements the block - schemes of algorithm.

At the first stage at initialization synoptical links it is necessary to consider some significant factors. To weighting coefficients differing from each other, small values differing from each other in the certain range, for example, from an interval from 0 up to 1 should be appropriated. If synoptical weights accept big initial values neurons, as a rule, quickly enough will reach a condition of saturation. After the given situation local gradients of algorithm of inverse distribution will accept small values that will cause significant lowering speed of tutoring of a web. If to weighting coefficients to appropriate small values the algorithm will languidly work in an environ of an origin of coordinates of a surface of errors. It is linked by that, unfortunately, the origin of coordinates is a saddle point, i.e. a stationary point where generating surfaces of errors along one axis have a positive gradient, and along another - negative. For these reasons it is not meaningful to use as too big, and too small initial values of weighting coefficients. As always the golden mean is somewhere in between these extreme measures.

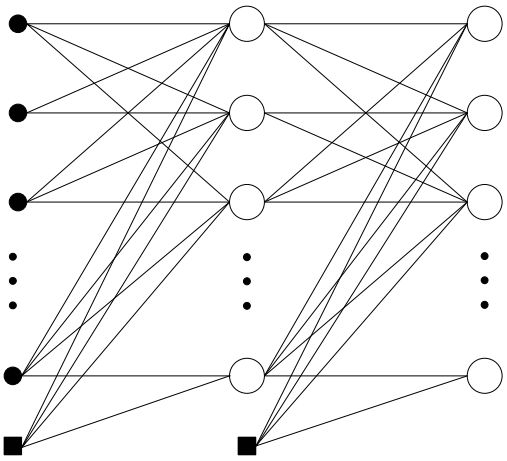

Figure 1. The two-layer neural networks used for pattern recognition

At the second stage there is a feeding of the next input vector from training Samples on an input of a web. An advisable condition for more productive tutoring a web is normalization of inputs. All input a vector and their values should be preliminary treated, so that average value on all training set was close to zero. For the decision of this problem it is possible to use one of the spreaded ways - process of scaling [1]:

$$x_м = (x_u - n)\, k, \qquad (1)$$

where x_u – an initial vector, $x_м$ – scaled. A vector n – the averaged value of set of input data's. k – factor of process of scaling.

At presentation on an input of a web of not normalized vector following problems are possible:

- Neurons of an input layer or will appear in fixed saturation, or all time will be inactive;
- Weighting coefficients during tutoring will accept very small or very much great values (depending on a variance) and as consequence, process of tutoring will be stretched for an indefinite period and exactitude of classification will decrease.

At the third stage output signals of neurons in each of layers are evaluated. Evaluations occur sequentially from a layer to a layer. After calculation of signals of all neurons of last layer the error of output value of each neuron in this layer is evaluated.

On the most important – the fourth stage of operation of algorithm there is a set up of weighting coefficients of a neural networks. In algorithm of inverse distribution of an error adjusting of weighting coefficients Δw_{ij} is made aside, opposite to a gradient of a surface of an error. Such approach ensures the optimum decision of a problem of minimization of functional of the error $\varepsilon = \varepsilon(\Delta w_{ij})$ which have been thought up Windrow and Hoff [2]. Though besides it sometimes use casual search of weighting coefficients Δw_{ij} with the subsequent evaluations and comparisons of functions of errors among themselves corresponding these factors. Last approach to the decision of an optimization problem is simple enough in comparison with gradient, but considerably yields to it by efficiency in complex variations of the main task of NN - maps of an input vector x to an output vector y.

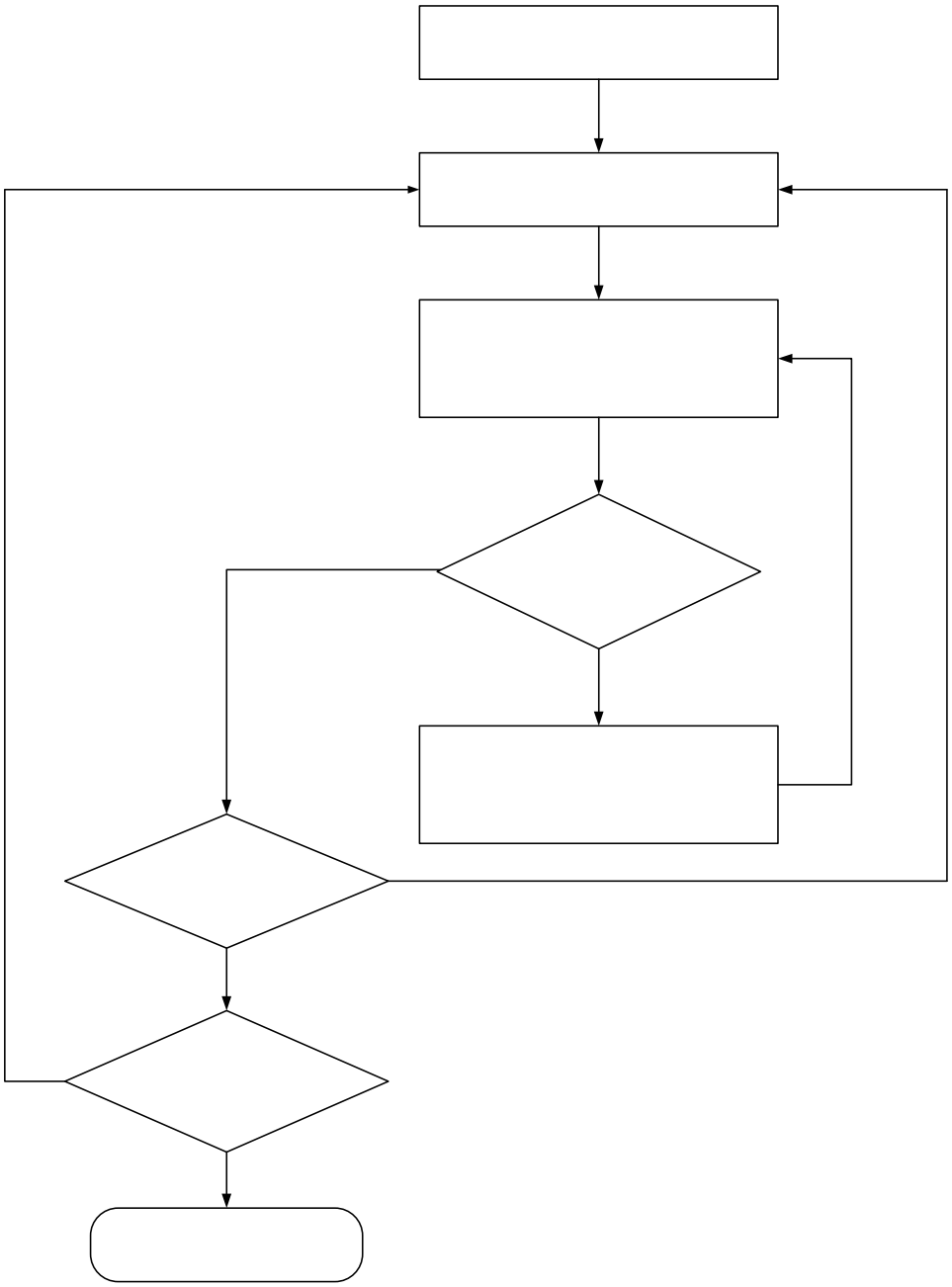

Figure 2. Algorithm of tutoring multilayer perceptron by method of Back Propagation

After the fourth stage of algorithm return on the third is necessary for an objective choice of a place of a stopping of tutoring by means of the account of an error for each input vector from training sample.

The epoch of tutoring is considered the web of last input vector finished only after presentation from training sample. After that the total error of discernment is evaluated, and there is its comparison with what or in advance certain magnitude. Tutoring of a neural networks appears finished after the total error on exits of a web will appear less than this magnitude.

Tutoring in multilayer NN which carries, in a greater degree, creative character was spent according to the various methods offered in the literature, and their fractional modification, following formulas therefore have been applied:

- For adjusting values of weighting coefficients on each step

$$\Delta w_{ij} = \eta(d_j - y_j) f_{sigm}'(S_j) x_i, \quad (2)$$

where $(d_j - y_j)$ – a difference between a demanded output signal and real, $f_{sigm}(S_j)$ – activation function, x_i – value signal's i, η – factor of speed of tutoring.

For acceleration of tutoring and heightening of productivity of a web series of receptions was used. Introduction in the formula of an impulse (2) factors k:

$$\Delta w_{(t+1)ij} = \eta (d_j - y_j) f_{sigm}'(S_j) x_i k(\Delta w_{(t)ij}) \quad (3)$$

The great value of factor k exhausts a web in a local minimum, and very far from global, i.e. high value k instead of rendering stabilizing effect on a web – destabilized it, and from small value practically there was no improvement of productivity, therefore k it is recommended to select equal 0,5.

For the offered set of plants, signals with which move in a web on discernment, the sequential algorithm of tutoring was used. As definition of probability of correct separation on classes it was used pertaining to bayes criterion of optimum classification [3].

As activation functions $f_{sigm}(S_j)$ has been selected a hyperbolic's tangent, ensuring wide range of values and an antisymmetry in comparison with usual logistical function.

On the basis of the reduced formulas, with use of language Object Pascal©, program Miranda MP has been developed (figure 3). The program operates in two conditions: discernment of digits of printed fonts, discernment of hand-written digits. Advantages of the given program complex are: discernment of invariant symbols to turn; visualization of operation of a web; flexibility of construction of various two-layer NN and their tutoring.

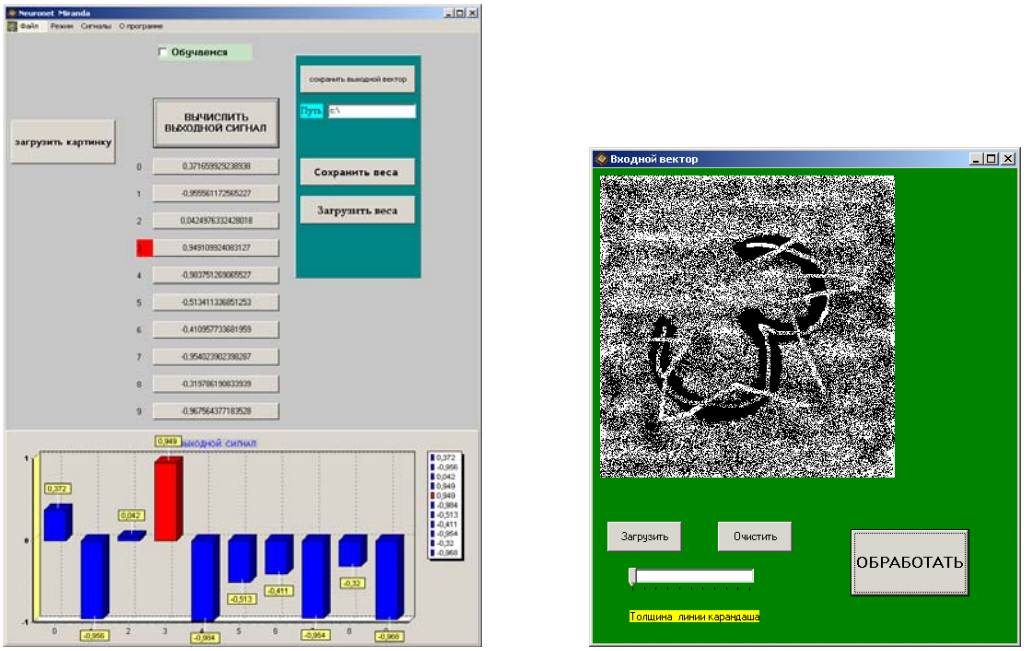

Figure 3. Presentation to a web, distorted, noise and exposed to turn, digit 3

Now development of the program complex containing various aspects of neural networks for even more qualitative discernment of complex plants.

Literature

1. Zaentcev I.V. Neural networks: basic models. (In Russian), Voronezh (1999) 76 p.
2. Yasnitsky L.N. Introduction to artificial intelligence. (In Russian) "Academia" Publishing House, Moscow (2005) 176 p.
3. Haykin S. Neural Networks: A Comprehensive Foundation // MacMillan College Publishing Co. New York, 1994.