Neural network application for predictive modeling

K E Tokarev¹,², Yu A Orlova², A F Rogachev¹, A Yu Rudenko¹ and M P Protsyuk¹

¹ Volgograd State Agricultural University, 26, University ave, Volgograd, 400002, Russia
² Volgograd State Technical University, 28, Lenin ave., Volgograd, 400005, Russia

E-mail: tke.vgsha@mail.ru, yulia.orlova@gmail.com, rafr@mail.ru, rudenkovgsxa@mail.ru, maria_protsuk@mail.ru

Abstract. The article analyzes the approaches to the construction of predictive models based on the apparatus of artificial neural networks, in particular, the method of back propagation of errors by iterative adjustment of weight coefficients.

1. Introduction

Artificial neural networks (ANN) are representatives of the cybernetic direction in science, based on the belief that natural systems in the form in which we see them today, are a community of winners in the centuries-old competition for survival.

Human, as the creator of artificial worlds, only need to understand the key mechanisms of their sustainable functioning and self-development in order to use this understanding in their own practice.

Artificial neural networks are computer systems that simulate the work of biological neurons with the possibility of parallel processing of information, the ability to learn and summarize the accumulated knowledge.

The evolution of ANN has led to the fact that currently some of them to some extent really simulate biological neural networks, while some have forgotten about the nature of its creation. Artificial neural networks were originally conceived as some abstractions, mathematical models that mimic the activity of the human brain. Distinguish the following properties harmoniously connecting with each other biological and artificial NS: the ability to learn; the ability to find solutions based on noisy, distorted and even contradictory data; fault tolerance of hardware-implemented neural networks).

In this regard, the development of methods for constructing neural network models of productivity, substantiation of their architecture and technological parameters is an important theoretical and applied task, the solution of which allows to increase the efficiency of planning and management of agricultural production by making more informed management decisions based on the use of tools for multivariate analysis of alternatives.

2. Methodology

Artificial neural networks are biological structures consisting of interconnected simple elements – formal neurons. The core of the concepts used is the idea that neurons can be modeled with fairly simple automata, and all the complexity of the brain, the flexibility of its functioning and other important qualities are determined by the connections between neurons. Each link is represented as a very simple element that serves to transmit the signal. It is assumed that the system of connections is rich enough in its capabilities and redundant enough to compensate for the poverty of the choice of elements, their unreliability, the possible destruction of the links[1, 2, 7].
The most well-deserved and probably the most important element of neurosystems is an adaptive adder. The adaptive adder computes the scalar product of the input vector \( x \) by the parameter vector. We call it adaptive because of the presence of a vector of adjustable parameters. For many problems it is useful to have a linear inhomogeneous function of the output signals. Its calculation can also be represented by an adaptive adder having \( n+1 \) input and receiving a constant single signal at the 0-th input (Figure 1).

The nonlinear signal converter is shown in Figure 2. He receives a scalar input \( x \) and transforms it into \( \phi(x) \). The most commonly used sigmoid function:

\[
\phi(x) = \frac{1}{1 + e^{-x}}
\]

An interesting property has a derivative of the sigmoid-it exists on the entire real axis and \( \phi' = \phi(1-\phi) \).

![Figure 1. Inhomogeneous adaptive adder](image1)

![Figure 2. Nonlinear signal converter](image2)

The branch point is used to send one signal to several addresses (Figure 3). It receives the scalar input \( x \) and transmits it to all its outputs.

![Figure 3. Branch point](image3)

The branch point is used to send one signal to several addresses (Figure 3). It receives the scalar input \( x \) and transmits it to all its outputs. The standard formal neuron is composed of an input adder, a nonlinear transducer, and a branch point at the output (Figure 4).

![Figure 4. Formal neuron](image4)

Among NS, the most common and most studied class is layered networks (as it is otherwise called, direct distribution networks).
In such networks neurons are located in several layers (Figure 5). The neurons of the first layer receive input signals, convert them and transmit them through the branching points to the neurons of the second layer. Next, the second layer is triggered, etc. to the k-th layer, which outputs signals to the interpreter and the user. If the opposite is not specified, then each output signal of the i-th layer is fed to the input of all neurons i+1. The number of neurons in each layer can be any and has nothing to do with the number of neurons in other layers. Standard input method: all neurons in the first layer receive each input signal. The first layer is called input, the last-output, the rest-hidden.

![Artificial neural networks with layers](image)

**Figure 5.** Artificial neural networks with layers

Note that in the literature in different ways can be called the layered network. Most often, the first (input) layer is not taken into account when naming NS of direct distribution, because its neurons are simply network inputs, but they are not fully functional neurons. Thus, a two-layer network consists of input, hidden and output layers. It can also be called a direct distribution network with one hidden layer.

The activation function of neurons (characteristic, transfer function) – a nonlinear converter that converts the output signal of the adder (Figure 4) – can be the same for all neurons in the network. In this case, the network is called homogeneous (homogeneous). Drawing up a network of neurons of the standard form is not mandatory.

Learning NS is the process of finding the synaptic weights of all neurons that are optimal for some functional quality, called learning error. It describes an integrated measure of the closeness of the outputs of the network $y(M)(k)$ and the instructions of the teacher $y^*(k)$, $k=1,...,K$ training examples:

$$J(w) = \sum_{k=1}^{K} Q(\varepsilon(w,k))$$

where $Q(\varepsilon(w,k))$ is called the instantaneous quality criterion.

Here $W=(w(M)t,w(M-1)t,...,w(1)t)$ is the vector of weights of all networks, $\varepsilon(w,k)=y(M)(k) - y^*(k)$. Often taken $Q(\varepsilon(w,k))=\varepsilon(w,k)^{tr} \varepsilon(w,k)$, $R$ – a positive definite matrix.

D.E. Rumelhart [3, 4] developed an algorithm for iterative adjustment of the weight coefficients of the NS, called the method of back propagation of the error. This algorithm is not without its drawbacks, but its role is very large for training NS direct distribution.

Let the network consist of $M$ layers with $n_\mu$ neurons in the $\mu$-th layer ($\mu=1,...,M$). A pair $(\mu, i)$ will denote the $i$-th neuron of the $\mu$-th layer. Inhomogeneous adaptive summation of this neuron:

$$\text{net}^{(\mu,i)} = w_0^{(\mu,i)} + \sum_{j=1}^{n_{\mu-1}} w_j^{(\mu,i)} y_j^{(\mu-1,i)} = w^T^{(\mu,i)} u^{(\mu,i)},$$

$$w(\mu,i)=(w0(\mu,i),w1(\mu,i),...,wn(\mu,i)), \mu=1,...,M, \text{ be the vector of weight coefficients } (n=N_{\mu}-1 \text{ – number of neurons in the previous layer}), y(\mu-1,i)=(1,y1(\mu-1,i),...,yn(\mu-1,i)), \mu=1,...,M, \text{ – advanced}$$
input vector of the neuron \( \mu \)-th layer (output neuron of the previous layer), here \( y(0,i) = x(i) \) is inputs of network. The output of the neuron is written as: \( y(\mu,i) = F(\text{net}, I) \), \( I = 1, \ldots, N_\mu \); \( \mu = 1, \ldots, m \).

Let us specify a set of \( K \) training pairs \( \{x(k), y^*(k)\}, k = 1, \ldots, K \). the Training (finding the weight coefficients) as follows.

1. Given \( \eta > 0 \) (usually taken as 0.2) is the learning constant (step length), and \( E_{\text{max}} > 0 \) is the maximum allowable learning error (optimization accuracy).

2. The weights of the network are initialized by small random numbers (if equal weights are set, and different weights are required for the proper functioning of the network, the network will not be able to learn). \( E := 0 \) – accumulated error, \( k := 1 \) – training pair number.

3. The input vector \( x(k) \); \( x := x(k) \), \( u \) – extended input vector is supplied to the network input, the vector \( y := y(M) \) – output of this network (the last layer) is calculated at a given input.

4. The error \( e = y - y^* = s(M) \) of the network operation is calculated for this example.

5. The weights of the last layer are corrected: \( w(M) := w(M) - \eta e F'(\text{net}(M)) u(M) \).

6. \( \mu := M - 1 \) is the number of the last hidden layer.

7. \( i := 1 \) – the number of the first neuron in the layer.

8. \( w(\mu,i) := w(\mu,i) - \eta S(\mu,i) F'(\text{net}(\mu,i)) u(\mu,i)^T \), where \( s(\mu,i) = s(\mu + 1) Td(\mu + 1,i) \), and \( d(\mu + 1,i) = F'(\text{net}(\mu + 1)) w(\mu + 1,i) \) – vector.

9. If \( i < N_\mu \), then \( i := i + 1 \) and return to step 8.

10. If \( \mu > 1 \), \( \mu := \mu - 1 \) and return to step 7.

11. The generalized error of the network functioning is calculated: \( E := E + 0.5 \cdot e^2 \).

12. If \( E < E_{\text{max}} \), then the training is over, otherwise \( E := 0 \), \( k := 1 \) and return to step 4.

This algorithm is called the standard error back propagation algorithm. It should be noted that adjusting the weights is a minimization of the error function. The standard error back propagation method is actually a gradient method for finding the optimum of the error functional. The weights change according to the rule: \( w(t) = w(t-1) - \eta Q(e(w(t-1), k)) \). Here \( Q(e(w(t-1), k)) \) is the gradient of the instantaneous functional – the descent direction, and \( \eta \) is the step length in the descent direction. It is worth noting the very slow convergence of this method.

Currently, more efficient methods of adjusting the network weights are used. For this purpose, subject to such well-known optimization techniques to search for the direction of decreasing functions as conjugate direction methods and the well-known quasi-BFGS and DFP methods [5,6].

The use of ANN in forecasting is quite transparent. They can be used in several ways, the most famous of which is similar to the methods of statistics, called the method of Windows. The Windows method assumes the use of two windows \( W_i \) and \( W_o \) with fixed sizes \( n \) and \( m \) respectively. These Windows are able to move with some step in the time sequence of historical data, starting with the first element, and are designed to access the data of the time series, and the first \( W_i \) window, receiving such data, transmits them to the input of the neural network, and the second \( W_o \)– to the output. The resulting \( W_i \rightarrow W_o \) pair at each step is used as a training sample element (a recognizable image, or observation).

One-step forecasting is used for short – term forecasts, usually absolute sequence values. The forecast is only one step ahead, but the actual value is used instead of the predicted value to make the forecast for the next step. Multi-step forecasting is used to make a long-term forecast and is designed to determine the main trend and the main points of trend change for a certain period of time in the future. In this case, the predictive system uses the obtained (output) data for the time moments \( k+1 \), \( k+2 \), etc. as input data for forecasting at time points \( k+2 \), \( k+3 \) and etc.

3. Results

As an experiment the problem of approximation of a number of dynamics of power generation of electric energy, i.e. construction of model on a final set of points on concrete examples about generation of the power released by the companies entering into Uniform power system of Russia taking into account two influencing factors was considered: losses in networks at transfer and average
daily temperature. In this case, the following tasks were set: search for influencing factors on the
dynamics of power generation; search for the optimal representation of statistical data on the dynamics
of generation and the results of the approximation of the behavior of the series under study. The initial
data for solving the problem are shown graphically in Figure 6, 7.

To solve the problem of constructing a model of the dynamics of the power generation process,
taking into account the selected influencing factors using the neural network apparatus, the RBF
network structure was chosen (Figure 8).

Figure 6. Power generation, GWh

Figure 7. Network losses, GWh

Figure 8. The structure of RBF network
The RBF network has an intermediate layer of radial elements, each of which reproduces the Gaussian response surface. Since these functions are nonlinear, it was enough to take one intermediate layer, determining only the desired number of radial elements. In order to get out of the network, it was necessary to combine the outputs of hidden radial elements. It is enough to create a linear combination of radial elements [8, 9].

The great advantage of ANN is their ability to work in multidimensional spaces, including spaces of mixed type, in which some variables are continuous and some are discrete. But the implementation of this ability implies the presence of the developer of the relevant data, as well as the ability to properly structure them in a form that is easy to remember. In the presence of powerful computer technology and excellent optimization programs, this remains the most difficult stage in the development of neuromodels, poorly amenable to formal description due to the specificity of the application problem to be solved in each case.

Figure 9. Structure of RBF network components for forecasting economic indicators in the energy sector

Thus, the RBF network has an output layer consisting of elements with linear activation functions (Figure 9, 10).

Figure 10. The neural interface system
In this sense, modeling with the help of ANN continues to be something of an art. Not meeting the expectations that they can be used to create models with the usual level of versatility and transparency, neuromodels, however, are much better prepared for use for the management of specific objects [10, 11]. Comparing the advantages and disadvantages of ins with the features of other modeling technologies, one can unwittingly come to the conclusion that their purpose is not in the total displacement of the classical models of the functional type, but in solving those problems in which traditional technologies have shown weakness. ANN provide such an opportunity, and with such a powerful tool in hand, it is unwise to ignore its capabilities.

4. Conclusions

Thus, the experiments show that artificial neural networks can be successfully used to predict the time series of power generation, released by the companies of the energy complex, as well as other economic indicators of energy. Moreover, in some cases, for example, when solving problems with unknown regularities, they give more accurate results than traditional forecasting models, and are more resistant to "noise" and adapted to environmental changes.

In addition, the range of computational problems of energy, in which the use of neural network modeling is quite extensive:

- short-term forecasting (forecasting of demand for electric energy, forecasting of demand for thermal energy, forecasting of demand for natural gas, forecasting the marginal price of the system in the energy market, forecasting the volume of electricity generation by wind farms, forecasting the temperature of the outside air);
- diagnostics and fault localization of station and network equipment;
- assessment of power system stability;
- modeling of the intensity of environmental pollution by power units.

Another disadvantage of neural models is the significant time and other resources required to build a satisfactory model. This problem is not very important if you are investigating a small number of time sequences. However, a typical predictive system in production management can include from several hundred to several thousand time sequences.

However, despite these shortcomings, the model has a number of advantages. There is a convenient way to modify the model as new observations appear. The model works well with time sequences in which the observation interval is small, i.e. a relatively long time sequence can be obtained. For this reason, the model can be used in areas where we are interested in hourly, daily or weekly observations. These models are also used in situations where you need to analyze a small number of time sequences. In addition, neural network forecasting is particularly effective when using additional information, i.e. when taking into account various factors affecting the subject of study.

Artificial neural networks are a universal tool for solving multidimensional problems of modeling and control in power engineering. The relatively slow pace of development of transients in energy systems, combined with the dramatically increased computing power of personal computers allows neural networks to make their way in new and new applications. Their application opens up real opportunities for the introduction of new, more sophisticated approaches to the methodology of modeling and management, promising to rethink the practice of exit the operation of energy facilities and to make energy consumers more quality and cost-effective.

In order for ANN to begin to bring real benefit to the domestic energy sector, it is necessary to bring them closer to the energy sector itself, namely, to give into the hands of those who work at energy enterprises, determines the situation on the ground. This requirement is in some contradiction with the way of allocation of responsibility between scientific, design, Supervisory and operational units of the power complex, in accordance with which the scientific service of energy is carried out. To be more precise-carried out in soviet times.

Today the system of scientific service of energy, left without proper financial support of the state, almost lost its former influence on scientific and technical progress in the industry. Therefore, energy companies have no other choice but to Finance scientific research on their own in order to strengthen
their own competitiveness by improving the practice of technological, organizational and economic management.

ANN, in this respect, have excellent chances to attract attention from the new owners of energy companies, representing a market-like tool for the study of "particulars", rather than protruding "communities".

And the last circumstance which can promote wide introduction of neural network models in power is closer coordination of practice of technological management of objects with economic indicators and criteria, and also active use in the course of planning of work of the enterprise of prognostic models. So, the wholesale market of the electric power already works today. There is no doubt that local thermal energy markets will also work. Short-term focus shift forecasting the risks and costs of the energy enterprise will make ANN a more popular tool, create incentives for their active development by the management personnel of the energy sector.

Acknowledgment
The reported study was funded by RFBR and Volgograd region according to the research project № 19-416-343006.

References
[1] Alekseev A A, Rozaliev V L, Orlova Y A and Zaboleeva-Zotova A V 2016 Context-sensitive image analysis for coloring nature images Advan. in Intellig. Syst. and Comput. 451 133–41
[2] Gagarin A G and Rogachev A F 2012 Prognozirovane urozhajnosti na osnove analiza kross-regional’nyh dannyh Proc. of Lower Volga Agrouniver. Complex: Sci. and Higher Vocat. Ed. 4 339–46
[3] Lisetti C L and Rumelhart D E 1997 Recognizing facial expressions with a neural network Proc. of the nineteenth annual conf. of the cognit. Sci. society p 1100
[4] Widrow B, Rumelhart D E and Lehr M A 1994 Neural networks – applications in industry, business and science Communicat. of the ACM 37(3) 93–105
[5] Ovchinnikov A S and Bocharnikov V S 2012 New technical solutions improve the efficiency of resource-saving irrigation methods Proc. of Lower Volga Agrouniver. Complex: Sci. and Higher Vocat. Ed. 1(25) 119–24
[6] Timoshenko M A, Rogachev A F, Medvedeva L N and Tokarev K E 2019 Analysis and support of decisionmaking with the use of latest information and communication technologies for the development of social and economic area with regard to capabilities of different categories of population Advan. in Intellig. Syst. and Comput. 726 989–1004
[7] Nikitin N A, Rozaliev V L, Orlova Y A and Alekseev A A 2018 Automated sound generation based on image colour spectrum with using the recurrent neural network CEUR Workshop Proc. 2212 399–408
[8] Kayvan T and Alireza S 2010 Hopfield Neural Networks as Pseudo Random Number Generators, in North America Fuzzy Information Processing Society (NAFIPS) pp 1–6
[9] Wen S, Zhigang Z, Tingwen H and Yide Z 2014 Exponential adaptive lag synchronization of memristive neural networks via fuzzy method and applications in pseudorandom number generators IEEE Transact. on Fuzzy Syst. 22(6) 1704–13
[10] Kashem M A., Jlasmon G B., Mohamed A and Moghavvemi M 1998 Artificial neural network approach to network reconfiguration for loss minimization in distribution networks Electr. Power & Energy Syst. 20(4) 247–58
[11] Anikin I and Khaled A 2016 Pseudo-random number generator based on fuzzy logic In Control and Communications (IEEE (SIBCON) pp 1–4