On neural network structure selection to solve problem of iron ore preparation process identification

Yu I Eremenko, D A Poleshchenko and Yu A Tsygankov
Department of Automation and Information Technology (AIT), STI NUST MISIS, 42, microdistrict Makarenko, Stary Oskol, 309516, Russia
E-mail: TsY-18@yandex.ru

Abstract. The possibility of creation the neural network model for predicting the iron content in the output product of grinding process of the ore-dressing industry is considered in this paper. The use of this range of tools is determined by the multifactoriness and non-linearity of the process, and by the need of the model to adapt to changing parameters. The study consisted in modeling of various structures to achieve the desired quality of the model’s output signal. The network with a structure of 4 neurons in the input layer, 60 neurons in a hidden layer and 1 neuron in the output layer had the best result in the generalization and accuracy of test set reproduction. The result allows to assume the possibility of using a neural network tool for the development of aggregates of technological grinding process, and the creation of a single control system with predictive models.

At present, the problem of production efficiency is rather topical at most industrial enterprises [1]. In recent years, domestic enterprises take considerable interest in methods of operational efficiency improvement, including because of the prevailing political and economic conjuncture in Russia. The above is also true not only for human resources and industrial units, but also for technological production in general, trying to optimize the indicators of quality, depending on changes at the input.

However, most of the objects of large technological enterprises today are quite highly automated complexes, and this complicates the task of increasing the efficiency of their operation. One of the aspects, available for modernization, is the improvement of algorithmic support of control systems [2].

At the moment, the vast majority of control systems are the set of local circuits based on the principles of PID control. This algorithm is the most universal and the most common in production. However, the drawback of this approach is that each particular circuit is aimed at performing the task for certain variable and completely does not take into account the requirements of other circuits. The overwhelming majority of modern technological processes are multiparameter control systems, between which the interrelation is observed. In these conditions, the control by changing a job in only one circuit can not be beneficial for the operation of the system as a whole. As a rule, this means the maintaining a specific regime in some technically justified limits. Most regulators are set during the adjustment to provide transient processes under the so-called “moderate” conditions. This does not allow the algorithm to provide the best control in the possible range of input characteristics change. As a result, with the maximization of the output of any variable in this section, negative consequences can occur in the management at the next stages of production [3].
In addition, the PID regulators used in the production of PI are linear in nature. However, the vast majority of technological systems and aggregates are nonlinear, and, accordingly, characterize their work and try to implement management within the optimal technical and economic boundaries with the use of this functional extremely difficult.

In such conditions, it is only the experienced operator who is able to effectively manage the process, with the maintenance of the product of the required quality. For this reason, there is a kind of "dependence" on the experience of this employee, which is not always possible to transfer to other employees. But even an experienced operator does not always manage the process efficiently in terms of economic indicators, i.e. he sometimes prevents the situation of the issue of marriage, is managing with the overexpenditure of raw materials, energy and other, creating the so-called "margin for quality" [4].

Therefore, in recent years, the so-called APC-systems (APC — AdvancedProcessControl) are being actively developed and implemented. This type of system is based on the well-known principle of predictive control. When implementing this type of system, the models of certain units of production processes are developed, and a control algorithm is accomplished, allowing to take into account the multifactority of the process. This is achieved by uniting the models of various facilities of production process into a single system, aimed at maximizing a certain output parameter, responsible for the quality of work. As a result, such setting actions on local regulators are calculated, which allow to keep the output parameters of particular system circuits within the limits, allowing to achieve the set goal not only at a specific circuit, but also at the output of the system as a whole [4–6].

It is worth noting that works in this direction began quite a long time ago. The first works dated back to the 1980s, they were made by Russian scientists [7]. However, then this principle of construction of complex object management systems was brought to the finished product by foreign leaders in the field of automation — Honeywell, Schneider Electric [8–10]. These systems are of particular value for such technological processes, where several complex aggregates are used, whose operation is divided into several subprocesses [11, 12]. Apparently, therefore, APC-systems are widely implemented in oil refining and chemical industries, in pulp and paper industry and non-ferrous metallurgy [10, 13]. However, it should be said that the technological process of grinding in the mining industry also fits the above characteristics [14]. Therefore, it seems possible to solve a similar problem for this production.

The disadvantage of the above systems is the use of standard identification methods in the description of nonlinear objects, the models of which are implemented on the basis of linear transfer members. Sometimes, in case of wide range of possible changes in the input characteristics of the object, a set of such members is implemented for each object; however the members are adjusted for some average characteristics.

To cope with this shortcoming, the decision was made to create models of production process, using the neural network tool.

Artificial neural networks are built on the principles of organization and functioning of their biological analogs. They are able to solve a wide range of problems of image recognition, identification, prediction, optimization, management of complex objects.

The choice of the neural network structure is carried out in accordance with the features and complexity of the task. To solve certain types of problems, there are already optimal configurations. If the problem can not be reduced to any of the known types, it is necessary to solve the complex problem of synthesizing a new configuration. It is necessary to be guided by the following basic rules: network possibilities increase with increase in number of network neurons, density of connections between them and number of layers; the introduction of feedbacks, together with the increase in the capacity of the network, raises the question of the dynamic stability of the network; the complexity of the algorithms of network operation, the introduction of several types of synapses helps to increase the power of the neural network.
Since the problem of the synthesis of a neural network strongly depends on the problem being solved, it is difficult to give general detailed recommendations. In most cases, the optimal variant is obtained on the basis of intuitive selection, although the literature provides evidence that for any algorithm there is a neural network that can implement it.

Many problems of pattern recognition (visual, speech), performance of functional transformations in signal processing, control, prediction, identification of complex systems, boil down to the following mathematical formulation. It is necessary to construct a mapping $X-Y$ such that the correct output signal $Y$ is formed for each possible input signal $X$. The mapping is given by a finite set of training examples, the number of which is substantially smaller than the total number of possible combinations of the values of the input and output signals.

In pattern recognition problems, $X$ is a representation of the image (image, vector), and $Y$ is the number of the class to which the input image belongs. In control tasks, $X$ is a set of controlled parameters of a managed object, and $Y$ is a code that determines the control action corresponding to the current values of the monitored parameters. In prediction tasks, time series representing the values of controlled variables over a certain time interval are used as input signals. The output signal is a set of variables that is a subset of the variables of the input signal. When identifying $X$ and $Y$, the input and output signals of the system are respectively.

Most of the applied problems can be reduced to the implementation of some complex functional multidimensional transformation. As a result of the mapping $X-Y$, it is necessary to ensure the formation of the correct output signals in accordance with: with all examples of the training sample; with all possible input signals that are not included in the training sample.

The second requirement greatly complicates the task of forming a training sample. In general, this problem has not yet been solved, but in all known cases a particular solution can be found [15].

Possessing the ability to approximate complex dependencies due to nonlinear activation functions in neurons, and the possibility of revealing hidden dependencies, neural networks (NN) will allow to obtain models of technological objects of the grinding process with high quality indicators [3].

From a large set of initial variables, it was necessary to select those, which have a greater connection with the main indicator of grinding process — the content of ferrum in the output product (Fe, %). Correlation analysis was performed for these purposes. It showed that three indicators had the high correlation level with the output parameter — the density of hydrocyclone overflow, operating in the process after sump at the second stage of grinding of each half-section of the grinding section ($p_1(t)$, $p_2(t)$), as well as the hydrocyclones, operating after one of the output sumps of the section ($p_3(t)$). As a result, 3 input and 1 output signals were selected, that made it possible to describe the process using the functional dependence, shown in figure 1.

To conduct training and testing procedures, data from a real industrial system was used in the work. The data sampling corresponded to 1 month of technological process operation, with 1-minute recording increment by input parameters, and 2 hours by iron content, that is due to the absence of automated control system. This means a wide range of changes in the input variables and state of objects. It will allow to take them into account in a model, in case of achievement of high modeling quality.

The analysis of the data showed that in the initial set there were data sections, which related to the periods of stoppages during operation of the objects or malfunctions in the measuring system. They had dips to zero level or signal levels, physically unattainable on the site. Such sets, in view of their absolute inapplicability for processing and lack of informativeness, were removed from further experiments.
During the experiments, the obtained data sets were filtered and normalized. To smooth the interference of the measurement system, a moving average filter was used:

$$x_{fj} = \frac{1}{N} \sum_{i=1}^{N} x_i,$$

where $N$ — is the number of data points, processed by the filter, $x_i$ — are the values of the input signals, $x_{fj}$ — are the values of the filter output.

In order to bring the values of real signals into compliance with the values from the tolerance range of activation functions, their normalization was performed:

$$x_{nj} = \frac{x_{fj} - \min}{\max - \min},$$

where $i = 1, N$ — is the number of data points of the vector, min, max — are the minimum and maximum values of the vector, $x_{fj}$ — are the values of the input signals, and $x_{nj}$ — are the normalized values of the vector.

Then, the entire available data set was divided into 2 parts: training set with a size of 15,000 points, and test set with a size of 3,500 points.

After carrying out the preparatory procedures, a series of experiments was carried out to identify the architecture of neural network, allowing to obtain a model with the best quality indicators. In the course of these experiments, the structures of multilayer neural network of direct signal propagation with a different number of hidden layers, neurons in them, activation functions by layers and the number of input signals were tested.

Figure 2 shows the overall architecture of the multilayer neural network with direct signal propagation, adapted to solve the problem, and used in experiments.

In the above figure, $X1, X2, X3$ — are the neurons of the input layer of the network, $Y$ — is the neuron of the output layer of the network, $s_1 = \sum_{i=1}^{3} w_{ni}^1 x_i + b_n^1$, $s_2 = \sum_{n=1}^{N} w_{mn}^2 x_n + b_m^2$, $s_3 = \sum_{m=1}^{M} w_{km}^3 x_m + b_k^3$, $s_4 = \sum_{k=1}^{K} w_{k}^4 x_k + b^4$ — are the formulas for calculation the values of the weighted sums in the layers, $f_1(s_1), f_2(s_2), f_3(s_3), f_4(s_4)$ — are the functions of activation in the layers, $w_{ni}^1, w_{mn}^2, w_{km}^3, w_k^4$ — are the values of weight coefficients in the layers, respectively.

In the work [16] the investigations were carried out, and 3 NN architectures were selected, capable of approximating such signals with the achievement of high quality.
Figure 2. Typical architecture of the multilayer neural network, used during experiments.

The first of them is a structure, consisting of one hidden layer and the number of neurons in it is 60; 1 neuron in the output layer; the activation functions are sigmoidal and linear, respectively (NN\textsubscript{1}). The functioning of this neural network is carried out according to the model:

$$Y_1(t) = f_2\left(b_2^0 + \sum_{n=1}^{N} w_{mn}^2 f_1\left(\sum_{i=1}^{3} w_{ni}^1 x_{nj}(t) + b_1^0\right)\right), \quad (3)$$

where $N = 60$ — is the number of neurons in the hidden layer, $b_1^0$, $b_2^0$ — are the displacements of neurons, $w_{ni}^1$, $w_{mn}^2$ — are the weight coefficients in the layers.

The model, based on the NN architecture with 2 hidden layers, 60 and 30 neurons, 1 neuron in the output layer and activation functions for the layers: tangential, sigmoidal and linear (NN\textsubscript{2}) was used as a more complex structure. Its functioning is described by the following equation:

$$Y_2(t) = f_3\left(b_3^0 + \sum_{k=1}^{M} w_{km}^3 f_2\left(b_2^0 + \sum_{n=1}^{N} w_{mn}^2 f_1\left(\sum_{i=1}^{3} w_{ni}^1 x_{nj}(t) + b_1^0\right)\right)\right), \quad (4)$$

where $N = 60$, $M = 30$ — are the numbers of neurons in the corresponding layers, $b_1^0$, $b_2^0$, $b_3^0$ — are the displacements of neurons, $w_{ni}^1$, $w_{mn}^2$, $w_{km}^3$ — are the weight coefficients in the corresponding layers.

In order to improve the quality of modeling, it was made an attempt to complicate the NN by introducing one more hidden layer. As a result, the structure was obtained, with the number of neurons 55, 50 and 45 in the hidden layers, 1 neuron in the output layer and activation functions: tangential, sigmoidal in the second and the third hidden layer, linear in the output layer (NN\textsubscript{3}). The model of this structure is presented below:

$$Y_3(t) = f_4\left(b_4^0 + \sum_{k=1}^{K} w_{km}^4 f_3\left(b_3^0 + \sum_{k=1}^{M} w_{km}^3 f_2\left(b_2^0 + \sum_{n=1}^{N} w_{mn}^2 f_1\left(\sum_{i=1}^{3} w_{ni}^1 x_{i}(t) + b_1^0\right)\right)\right)\right), \quad (5)$$

where $N = 55$, $M = 50$, $K = 45$ — are the numbers of neurons in the corresponding layers, $b_1^0$, $b_2^0$, $b_3^0$, $b_4^0$ — are the displacements of neurons, $w_{ni}^1$, $w_{mn}^2$, $w_{km}^3$, $w_{km}^4$ — are the weight coefficients in the corresponding layers.

To evaluate the functional efficiency of models, based on these neural network structures, such numerical quality indicators as the quadratic criterion (6) and the correlation coefficient of the real and model values of the output signal of the ferrumcontent were used (7).

$$F = \sum_{i=1}^{n} (Y_{\text{mod}_i} - Y_{\text{exp}_i})^2, \quad (6)$$
where $Y_{\text{mod}}$ — are the values of the real graph of current, $Y_{\text{exp}}$ — are the simulated values of the model output.

$$
\bar{r} = \frac{\sum_{i=1}^{n}(Y_{\text{mod}} - \bar{Y}_{\text{mod}})(Y_{\exp} - \bar{Y}_{\exp})}{\sqrt{\sum_{i=1}^{n}(Y_{\text{mod}} - \bar{Y}_{\text{mod}})^2 \sum_{i=1}^{n}(Y_{\exp} - \bar{Y}_{\exp})^2}},
$$

where $\bar{Y}_{\text{mod}} = \frac{1}{n} \sum_{i=1}^{n} Y_{\text{mod}}$, $\bar{Y}_{\exp} = \frac{1}{n} \sum_{i=1}^{n} Y_{\exp}$.

The numerical values of quality indicators, obtained during the training of the above models, are given below in table 1.

**Table 1.** Numerical indicators of modeling quality.

| The tuple of input signals of neural network | Neural network architecture | $r_{\text{NN}_1}$ | $r_{\text{NN}_2}$ | $r_{\text{NN}_3}$ |
|---------------------------------------------|-----------------------------|-------------------|-------------------|-------------------|
| $[\rho_1(t), \rho_2(t), \rho_3(t)]$         | $\text{NN}_1$               | 338.87            | 0.7454            | 294.2             |
|                                           | $\text{NN}_2$               |                   |                   | 0.7837            |
|                                           | $\text{NN}_3$               |                   |                   | 207.65            |
|                                           | $r$                         |                   |                   | 0.8531            |

In order to improve the training quality, it was made an attempt to change the architecture of neural network.

To do this, in threes neurons — density values with the first-order lag were added to the input layer of all neural networks: $[\rho_1(t-1), \rho_2(t-1), \rho_3(t-1)]$. The overall architecture of neural networks, constructed according to this principle, is shown in figure 3.

**Figure 3.** Typical architecture of the multilayer neural network with additional neurons of processing of input signals with the first-order lag.
Figure 4. Typical architecture of the multilayer neural network with additional neurons of processing of input signals with the first-, the second-, and the third-order lags.

During the experiments with the modified structure of the input layer on the tested architectures, the numerical values of quality indicators were obtained. They are presented in table 2.

**Table 2.** Numerical indicators of modeling quality.

| The tuple of input signals of neural network | Neural network architecture |
|--------------------------------------------|-----------------------------|
|                                            | NN_1 | NN_2 | NN_3 |
| $[\rho_1(t), \rho_2(t), \rho_3(t), \rho_1(t-1), \rho_2(t-1), \rho_3(t-1)]$ | 324.96 | 0.7575 | 234.5 | 0.8321 | 110.69 | 0.9246 |

It can be seen that the quality of modeling has increased, relative to the base models. In order to further increase the quality, it was decided to try to increase the number of neurons in the input layer, with the purpose of introduction the additional lags in the input channels. As a result, the structure was obtained. Its general form is shown in figure 4.

The results of modeling, using this modification of the basic structures of neural networks, are presented in table 3.

**Table 3.** Numerical indicators of modeling quality.

| The tuple of input signals of neural network | Neural network architecture |
|--------------------------------------------|-----------------------------|
|                                            | NN_1 | NN_2 | NN_3 |
| $[\rho_1(t), \rho_2(t), \rho_3(t), \rho_1(t-1), (t-2), (t-3), \rho_2(t-1), (t-2), (t-3), \rho_3(t-1), (t-2), (t-3)]$ | 295.83 | 0.7823 | 184.26 | 0.8708 | 74.47 | 0.949 |

The next step was an attempt to change the original structure, shown in figure 2, by introducing an additional neuron into the input layer of the basic network architecture, designed
to processing the output signal with the first-order lag. It is achievable under real production conditions. As a result, the following general structure was obtained (figure 5).

\[ s_i = \sum_{n=1}^{N} w_{in}^{i} x_n + b_{i}^{1}, s_2 = \sum_{n=1}^{N} w_{m}^{i} x_n + b_{i}^{2}, s_3 = \sum_{n=1}^{M} w_{kn}^{i} x_n + b_{i}^{3}, s_i = \sum_{k=1}^{K} w_{k}^{i} x_n + b_{i}^{4}. \]

**Figure 5.** Typical architecture of the multilayer neural network with additional neurons of processing of output signal with the first-order lag.

Experiments with neural networks, having this kind of structure of input layer, delivered the best result in training. This is also evidenced by the numerical quality indicators, listed below in table 4.

**Table 4.** Numerical indicators of modeling quality.

| The tuple of input signals of neural network | Neural network architecture | NN_1 | NN_2 | NN_3 |
|--------------------------------------------|----------------------------|------|------|------|
| \([\rho_1(t), \rho_2(t), \rho_3(t), Fe(t-1)]\) | F | F | F | F |
|                                           | r | r | r | r |
|                                           | 3.524 | 0.9977 | 2.11 | 0.9986 | 1.197 | 0.9992 |

Moreover, it should be noted that when supplementing additional neurons to the input layer, which process the output signal of ferrum content with the second-, and the third-order lags, there was no significant improvement in the quality of model functioning. Therefore, the results of these experiments are not reflected in the work.

The next step in testing the quality of models’ functioning was the experiments, related with the supply to the network of a test set with a size of 3,500 points. It corresponded to a sufficiently long period, during which there were changes in the characteristics of input raw materials and external conditions. Finally, the results were obtained, and presented in table 5.

It can be seen from the table that when the NN structure becomes more complicated, first there is the improvement of the model functioning on the training set, and then the deterioration of operation on the test set, with respect to the network with a simpler structure. This is a good example of such a phenomenon as retraining of the neural network.

Therefore, further attempt to improve the quality of the model and its adaptation to the possibility of functioning under production conditions was performed using the modified network architecture NN_1.

For the possible reduction of the number of training epochs and prevention of the phenomenon of neural network retraining, the algorithm of selection of validation set in the process of training was used. This mechanism allowed to reduce the number of necessary epochs to 17, without loss.
Table 5. Numerical indicators of modeling quality.

| The tuple of input signals of neural network | NN_1 | NN_2 | NN_3 |
|---------------------------------------------|------|------|------|
| Test set                                    | 2.342| 0.9912| 16.808|
|                                             | 0.9444| 0.8681|

of model quality, that is reflected in the numerical quality indicators: $F = 2.1436; r = 0.992$. Training in such short terms allows to talk about the possibility of using this model in real production conditions, where the frequency of laboratory analysis is 2 hours. That is, under the given conditions it can be said about the possibility of additional training of the neural network in case of changing the input characteristics to previously unknown ones, and continuation the further work of the model. Figure 6 presents the result of the work of this neural network, using the algorithm of validation during training, where the heavy line shows the real graph of ferrum content in the output concentrate, obtained on the basis of laboratory analysis, and the thin dot-dash line shows the output signal of the model.

![Figure 6](image-url)

**Figure 6.** The result of operation of the ball mill model, based on the modified structure NN_1.

Conclusion
In the process of work, the training and test sets were analyzed; the most informative signals were selected from the total number of variables; the data was pre-processed. In the research, a number of experiments were carried out to model the ore grinding process, using various structures of neural networks and combination of signals in the input layer. On the basis of numerical quality indicators, the structures were selected, which allowed to obtain the most qualitative result of modeling for a given set of input data. Neural network architectures, having high accuracy in the process of operation on the training set, were evaluated on the test set. This was the main criterion for the quality of the procedure for identifying the process of grinding by a neural network.
In the course of the work, the model of grinding stage, based on the neural network, was obtained. It allows to predict the yield of iron in the finished concentrate, according to the selected technological parameters of the grinding process. This presupposes the possibility of using given model in the management system of real production cycle, and developing the models of other aggregates for creation a single system of process control.

Acknowledgements

The study was carried out with the financial support of applied scientific research of the Ministry of Education and Science of the Russian Federation, contract 14.575.21.0133 (RFMEFI57517X0133).

References

[1] Monov V, Sokolov B and Stoenechev S 2012 Cybernetics and Information Technologies 12 51–68
[2] Poleshchenko D A and Tsygankov Y A 2016 International Siberian Conference on Control and Communications (SIBCON-2016), National Research University "Higher School of Economics" (Moscow, Russian Federation) pp 227–232
[3] Eremenko Y, Poleshchenko D and Glushchenko A 2014 Applied Mechanics and Materials 682 80–86
[4] Dozortsev V M and Kneller D V 2005 Transducers and systems 10 56–62
[5] Dozortsev V M, Itskovich E L and Kneller D V 2013 Automation in industry 1 12–19
[6] Dozortsev V M and Kreidlin E Y 2009 Automation in industry 6 11–16
[7] Dozortsev V M, Itskovich E L, Nikiforov I V and Perel'manp I I 1983 Proc. IFAC / IFIP Symp. Real-Time Digital Control Appl. Guadalajara (Mexico) 1 205–210
[8] Anosov A A, Borodin P E, Dozortsev V M, Elitov G L and Kneller D V 2011 Automation in industry 8 29–37
[9] Kneller D V and Zakharkin M A 2010 Transducers and systems 10 57–71
[10] Jonas R K and Kneller D V 2009 Automation in industry 8 32–38
[11] Comacho E F and Bordons C 1998 Model Predictive Control (Springer Verlag)
[12] Tatjevskyr P 2010 Advanced Control of Industrial Processes: Structures and Algorithms (L.: Springer)
[13] Zozulya Y I and Sletnyov M S 2013 Automation, telemachainization and communication in the oil industrys 1 7–14
[14] Umucu Y, Çağlar M F, Gündüz L, Bozkurt V and Deniz V 2011 International Symposium on Innovations in Intelligent Systems and Applications (Istanbul) pp 344–348
[15] Kruglov V V and Borisov V V 2002 Artificial neural networks. Theory and practice (Moscow: Hotline-Telecom)
[16] Eremenko Y I, Poleshchenko D A and Tsygankov Y A 2017 J. Fundam. Appl. Sci. 9(7S) 706–721