Analysis of Neural Networks for Predicting Time Series When Assessing Industrial Safety Risks

Sergey Ya. Nagibin, Dmitry I. Loskutov

Moscow Aviation Institute (MAI), 125993, Volokolamskoe highway, 4, Moscow, Russia

Abstract – **The paper describes the choice of an artificial neural network (ANN), the most effective for use in problems of modeling the behavior of complex dynamic systems with the subsequent solution of the forecast problem. The choice is made to implement a risk-based approach in the domestic trusted innovation umbrella system «Zodiac» when monitoring the industrial safety of the enterprises of the Fuel and Energy Complex (FEC).**

Keywords – **Coefficient of determination, industrial safety, Kullback-Leibler divergence, monitoring, neural network, risk-based approach, «Zodiac» software.**

1. Introduction

Ensuring industrial safety is one of the most important tasks solved during the operation of hazardous production facilities (HPF). From the industrial safety's state depend both industrial enterprises' continuity and effectiveness, and citizens' life and health as well as environment preservation. Currently, the paradigm of state supervision in the industrial safety field is changing and a new model of state regulation is being implemented taking into account the accidents' risk degree and the scale of their possible consequences, which will reduce the control measures number.

The changes' guiding principle is the risk-based approach ensuring the safe operation of HPF of the FEC, which involves the identification, analysis and prediction of industrial accidents' dangers, the risk assessment and possible scale of the accidents' consequences at the HPF in real time to organize the necessary organizational and technical measures for the accidents prevention, to prevent the threats emergence to large industrial accidents and increasing the ensuring industrial safety efficiency at a separate HPF and (or) in the supervised facilities system as a whole [1], [2], [3], [4].

Industrial accidents have their own characteristics; the main ones are the comparative rarity of accidents compared to the production life cycle and a significant spread in the consequences magnitude. The danger of industrial accidents is usually assessed by the severe damage possibility, or the threat of their causing.

A modern quantitative assessment of the accident risk as a tool for analyzing industrial hazards was mainly formed in the West in the 1970-1980s, as a reaction to the major industrial accidents phenomenon that swept through these years in industrialized countries.

Attempts to directly apply the well-known and well-developed methods of the reliability theory to estimating rare unique events frequency, and probability theory to determining random values of accident damage in complex technical and social systems, did not bring satisfactory results, due to these theories limitations. For example, the reliability theory operates with a random time amount between successive failures. For "unique" accidents, this value tends to infinity. In addition, the causes of accidents are not only technical failures, but also poorly formalized human errors and poorly predictable external influences.

Thus, the technical risks of devices failure that make up the technological chain and the risks of the technological process failure as a whole pose a threat to the HPF industrial safety and significantly affect the generalized assessment of HPF industrial safety state.

Taking into account the tasks and features of developing a HPF industrial safety monitoring system, this article proposes to use an umbrella operational monitoring system to assess the emergency situations risks to collect parameters that can be used to assess the most critical systems' and elements' state of a monitored facility, collect statistics and use it for calculating the emergencies risk integral indicator. As such an umbrella system to create a system for remote industrial safety monitoring in Russian FEC, the Russian "Zodiac" software is used.

The implemented approach's peculiarity is that the risk assessment is based on the information analytical processing on the prerequisites for accidents and incidents identified at the early dangerous events' development stages. In this case, reliable, real-time information on the current control object's state, its components and subsystems are used.

The proposed approach allows implementing the proactive industrial safety risk management, the implementation of operational preventive measures to prevent accidents, pre-emergency situations and incidents.

2. Grapho-Analytical Method for Determination of Geometric Parameters of the Shock Wave Zone

When assessing industrial safety risks, the equipment's reliability characteristics are not always the determining factors. In these cases, as a rule, mathematical modeling methods are used, which make it possible to reduce studying the various processes' problems to mathematical model's properties studying task.

To simulate production technological processes taking into account their influence on the production facility industrial safety level, along with the reliability theory and mathematical statistics, the "Zodiac" software uses approaches and methods of catastrophe theory, chaos theory, theory of nonlinear dynamic systems and neural networks [5], [6].

To study complex nonlinear systems with feedback, it is possible to use methods based on the deterministic chaos theory. Chaos theory considers complex systems that are highly dependent on initial conditions in which small changes in parameters can lead to unpredictable consequences. In particular, deterministic chaos explains irregular behavior and anomalies in systems that are not stochastic. Chaos theory (chaotic dynamics) offers completely new concepts and algorithms for time series analysis, which provide a more complete understanding of the signals' nature. This theory provides a wide variety of methods, including attractor recovery in the lag phase space and calculation of Lyapunov exponents.

The ideas underlying the ANN theory are based on a number of theorems, according to which any multidimensional continuous function can be represented in the form of sums and products of one-dimensional functions [7]. Then the basic structure of the complex system's behavior, containing all the information about it, namely the attractor of a dynamic system (that is, a subset of the phase space that attracts trajectories in the infinite time limit), can be restored by measuring the observed characteristics of this dynamical system, fixed as a time series. When analyzing time series, the main task is the dynamic system's reconstruction that generated this series. In this case, fewer variables may be required to describe the system than in the general case. Therefore, the system can be characterized using its small dimension projection.

Multilayer neural networks perform operations that are very similar to projection: they calculate the weighted sum of a time series elements. Therefore, if the required small dimension projection exists, then the neural network can find it. Therefore, using properly designed neural networks, it is possible to make predictions in situations that are complex in nonlinear dynamics terms [8].

The time series prediction, solved by neuroanalysis methods, reduces to approximating the function of many variables from a given set of examples using the procedure of immersing a series in an m-dimensional lag space. Choosing a sufficiently large m, we can guarantee an unambiguous dependence of the series' future value on its m previous values, i.e. time series prediction is reduced to the problem of interpolating the function of many variables. The neural network can then be used to restore this unknown function from a set of examples given by the history of a given time series [9].

To analyze and synthesize the nonlinear systems behavior, the stated theoretical prerequisites should allow neural networks to solve the following problems:

- predict complex non-linear system behavior with a built-in mechanism for calculating the "forecast horizon" based on a modified method for calculating Lyapunov indices;
- automatically calculate the confidence prediction interval, i.e. "dynamic thresholds";
- carry out an automatic search for anomalies in the nonlinear dynamic systems behavior.

The selected neural network should allow to efficiently solving the risk assessment issues using the accumulated statistics of accidents and failures at the HPF facilities.

Particularly dangerous industrial enterprises are complex heterogeneous systems, the behavior of which is manifested in the form of signals' chaotic

sequences from sensors that control production processes, presented as temporary radar. Nonlinear dynamics theorems allow relating these sequences to the dynamical system main characteristics, such as a strange attractor and system's fractal dimension.

To solve these systems, it is advisable to use neural networks, with their training methods, and use non-linear multidimensional optimization methods.

Currently, various methods for developing ANN are known: perceptrons, networks with backpropagation of errors, recurrence networks, Jordan networks, Elman networks, deep learning networks, etc. Each of them has its own characteristics of work and training [10].

When developing intelligent monitoring systems in the innovative "Zodiac" software, designed for the Russian Federation FEC HPF industrial safety control systems, it became necessary to conduct research and comparative analysis to select neural networks based on various operating principles, the most effective for use in problems of modeling the complex dynamic systems behavior with the forecast problem subsequent solution.

To model data series behavior, for example, describing the signals' behavior of the "safety industrial complex integral estimation", is possible only using the "black box" approach, i.e. systems, the internal structure and mechanism of which are very complex, unknown or unimportant in the framework of this task. While studying such systems, instead of the system components' properties and interconnections, the system reaction, as a whole, to changing conditions is studied. This is mainly because the complex industrial objects' studied signals are chaotic fractals with behavior defined as "deterministic (or dynamic) chaos", and therefore are described by multidimensional models of differential equations that are non-linear in parameters [11], [12].

3. Methodology

To evaluate and compare the results of neural network operation, ANN based on a multilayer architecture and recurrent neural network based on the Long-Short-Term-Memory (LSTM) approach were selected. The main research goal in this article is to forecast time series (any) future values, as well as to compare the results of neural network operation based on a multilayer architecture and recurrent neural network operation based on the LSTM approach.

The procedure for calculating ANN parameters is the task of complex multidimensional nonlinear optimization, for the solution of which a number of methods have been developed: an algorithm for backpropagating errors, learning with a schedule,

learning with noise, stochastic optimization (genetic methods, simulated annealing method), etc. [13].

ANN during its formation may contain superfluous elements, and when optimizing the neural network architecture, its neurons, weights and thresholds that change little during training, are excluded from the network.

Since, as a rule, in the learning algorithms, a local gradient is calculated for each multilayer perceptron neuron, knowledge of derivative activation function associated with this neuron is required.

For such a derivative function to exist, activation must be continuous. In other words, differentiability is the only requirement that the activation function have to satisfy.

To configure the multilayer perceptron, the error backpropagation algorithm is used, which assumes the absence of a priori information, therefore, synoptic weights and threshold values are formed using a uniformly distributed number sensor.

Images from the training set are put into the neural network. For each image, a direct and reverse pass is performed sequentially.

Let the training example be represented by a pair $(x(n), d(n))$, where $x(n)$ is the input vector presented to the input nodes layer; $d(n)$ is the desired response provided by the output neurons layer to form the error signal. Let us calculate the induced local fields and functional signals of the neural network, passing through it in layers in the forward direction. The induced local field of the neuron j of layer l is calculated by the formula

$$v_j^{(l)}(n) = \sum_{i=0}^{m_0} w_{ji}^{(l)}(n) y_i^{l-1}(n),$$

where $y_i^{l-1}(n)$ is an output signal of neuron i located in the previous layer l-1 at iteration n; $w_{ji}^{(l)}(n)$ is a synoptic weight of the connection between the neuron j of the layer l and the neuron i of the layer l-1. For $i = 0$ $y_0^{l-1}(n) = +1$, $w_{j0}^{(l)}(n) = b_j^l(n)$ is a threshold applied to the neuron j of layer l is expressed as follows:

$$y_j^{(l)}(n) = f_j(v_j^{(l)}(n)).$$

If neuron j is in the first hidden layer, then

$$y_j^{(0)}(n) = x_j(n),$$

where $x_j(n)$ is j-th element of the input vector $x(n)$. If neuron j is in the output layer (L is the network depth), then

$$y_j^{(L)}(n) = o_j(n).$$

Let us calculate the error signal

$$e_j(n) = d_j(n) - o_j(n),$$

where $d_j(n)$ means desired neuron response.

During the return pass, let us calculate the network nodes local gradients according to the following formula:

$$\delta_j^{(l)}(n) = \begin{cases} e_j^{(L)}(n)f_j'\left(v_j^{(l)}(n)\right) \\ f_j'\left(v_j^{(l)}(n)\right)\sum_k \delta^{(l+1)}(n)w\delta_{kj}^{l+1}(n) \end{cases}.$$

For neuron j of the output layer L and for neuron j of the hidden layer l.

The change in the network layer l synoptic weights is performed in accordance with the generalized delta rule

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha\left[w_{ji}^{(l)}(n-1)\right] + \eta\delta_j^l(n)y_j^{(l-1)}(n),$$

where η is a learning speed parameter.

Forward and backward passages are iteratively performed, presenting to the neural network all era learning examples until the stop criterion is reached. Under the era is understood a single passage of data through a neural network in the forward and reverse direction.

To train the neural network, let us use the second-order quasi-Newton method, namely the Bruden-Fletcher-Goldfarb-Channo (BFGS) algorithm, the advantage of which, in contrast to the Newtonian methods, is that the Hessian of the function $\nabla^2 f(x_k)$, i.e. there is no need to find second-order partial derivatives. Instead, the Hessian is calculated approximately based on the steps taken before.

$$B_k d_k = -\nabla f(x_k) \leftrightarrow d_k = -H_k \nabla f(x_k),$$

where $H_k = B_k^{-1}$,

B_k is some matrix, d_k is a direction of descent.

Further from the point x_k, a step made towards d_k:

$$x_{k+1} = x_k + \alpha_k d_k,$$

where $\alpha_k > 0$ is a step length.

Starting from $B_0 = I$, the algorithm recalculates the Hessian approximation according to the rule $B_{k+1} = B_k + U_k$, where there is some low-ranking update. A small rank U_k is necessary for developing an effective procedure for calculating the inverse matrix $H_{k+1} = B_{k+1}^{-1} = (B_k + U_k)^{-1}$.

The specific type of U_k update follows from several requirements fulfillment. The main one is the secant equations: $B_{k+1}(x_{k+1} - x_k) = \nabla f(x_{k+1}) - \nabla f(x_k)$, which is valid for all quasi-Newton methods.

This equation alone is not enough to uniquely determine B_{k+1}. A specific quasi-Newtonian scheme is obtained by imposing additional requirements on the Hessian approximation.

Let us introduce the notation:

$$s_k = x_{k+1} - x_k$$
$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

Then, the recalculation scheme BFGS:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k},$$

$$H_{k+1} = \left(I - p_k s_k y_k^T\right)H_k\left(I - p_k y_k s_k^T\right) + p_k s_k s_k^T,$$

where $p_k = \frac{1}{y^T s_k}$.

Each iteration can be accomplished with a cost of $O(n^2)$ (plus the cost of computing the function and estimating the gradient). There are no $O(n^3)$ operations such as solving linear systems or complex mathematical operations. Even if Newton's methods converge much faster (quadratically), the cost of each iteration is higher, since it is necessary to solve linear systems.

LSTM-network is a kind of recurrent neural networks architecture, and is well adapted for training on the classification problems, processing and forecasting time series in cases where important events are separated by time lags with an indefinite duration and boundaries.

Ordinary neural networks have a fixed number of inputs and perceive each of them as independent. In recurrent networks, communications between neurons can go not only from the lower layer to the upper, but also from the neuron to itself, more precisely, to the previous value of this neuron itself or other neurons of the same layer. This allows to reflect the variable dependence on its values at different points in time: the neuron learns to use not only the current input and what the previous levels neurons did with it, but also what happened to it by itself and, possibly, other neurons at the previous inputs.

The recurrent networks disadvantage is that they do not cope effectively with situations where it is needed to remember something for a long time: the influence of a hidden state or input from step t on the subsequent states of a recurrent network decays exponentially.

4. Results

Let us evaluate the obtained results' quality and the neural networks operation forecast using classical criteria, namely, the Kullback-Leibler divergence (KL) and the determination coefficient R^2 [14].

The initial data for solving this problem are time series (industrial enterprise's measuring sensors readings), where neighboring points can depend on each other, and this dependence cannot be ignored.

The data are divided in the ratio of 90% to 10% in order to train the neural network in 90% of the initial data and test on the remaining 10% [14].

After obtaining the time series, let us reconstruct the phase space. Calculation of *delay* using the mutual information method shows the value 12. System dimension (embed = 264), let us calculate using the false neighbors method [15], [16].

The result is a matrix consisting of 264 columns and 12 rows. Based on the obtained matrix, neural network training was carried out. The last column of the matrix is taken separately as a "teacher" when teaching ANN. Data is normalized in the range from -1 to 1 [16], [17], [18].

In the first case, let us build a neural network in the multilayer perceptron architecture.

After training the network according to the algorithm described above, let us test for 10% of the data that was cut off at the very beginning and calculate the coefficients R^2 and KL.

Empirically, it was found that the neural network optimal architecture for the source data is the presence of three hidden layers with the number of neurons 12 on the first two layers and 2 on the third layer with a logistic activation function and a forecast on the last 20% of the initial data. The form of sigmoidal nonlinearity in general is defined as follows:

$$f_j\left(v_j(n)\right) = \frac{1}{1+\exp\left(-av_j(n)\right)}, a > 0,$$

where $v_j(n)$ is a local neuron j field. The neuron output signal amplitude with such an activation function lies in the range $0 \leq y_j \leq 1$.

R^2 training=-0.9034,KL training=0.2631,R^2 control=-0.8218,KL control=0.3223, number of neurons=[12,12,2], hidden layers=3, Training data=90%,Testing data=10%,Prediction=20%
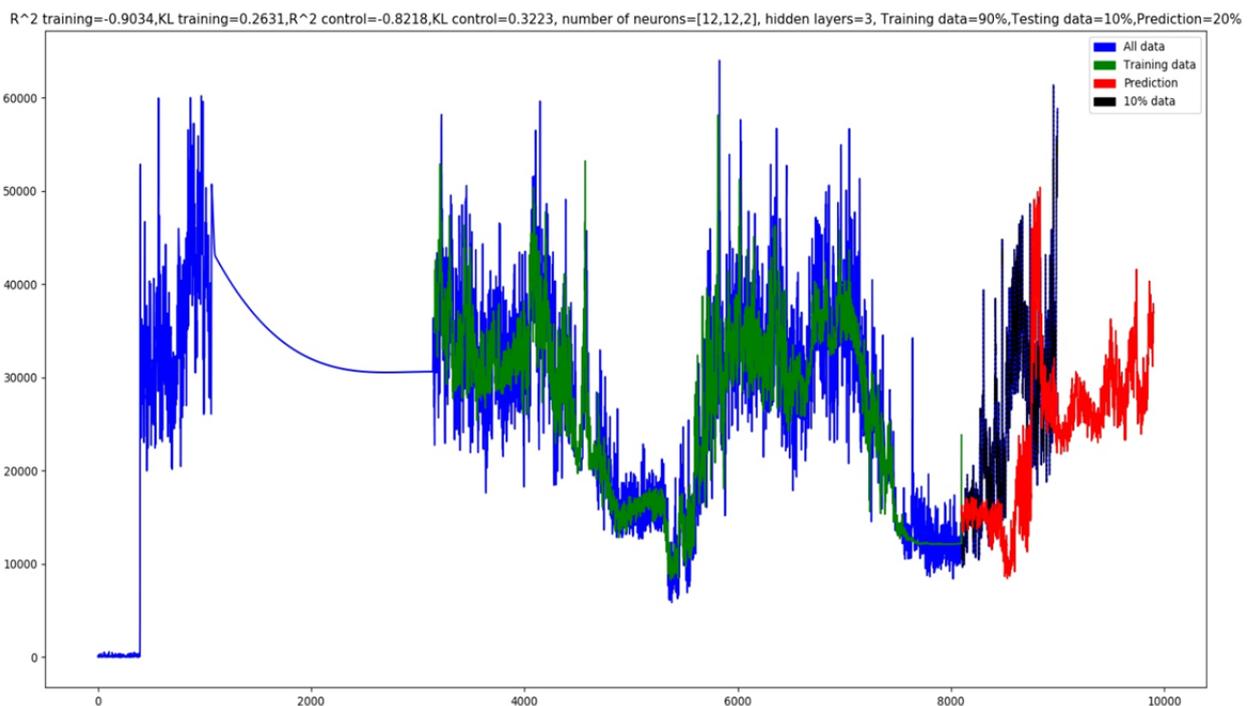
Figure 1. Forecast result with the activation function in the form of a logistic function

The blue color indicates the source data, the green one is about trained data, the black is for 20% that were cut off from the source data and the red is the predicted data.

The figure shows that the trained data dynamics accurately describes the source data. The coefficients KL and R^2 are 0.1391 and 0.7934 respectively for the trained data and 0.3388 and 0.7218 for the predicted, which is a good result, since KL is quite close to 0 and R^2 to 1. The dynamics of the predicted data corresponds to the test.

When analyzing ANN with hyperbolic tangent, the coefficients KL and R^2 are 3.2523 and 0.6794, respectively, which is significantly worse.

Therefore, in the first case, the best result for this data is the network architecture with a logistic activation function.

In the second case, let us evaluate the recurrent (LSTM) network operation.
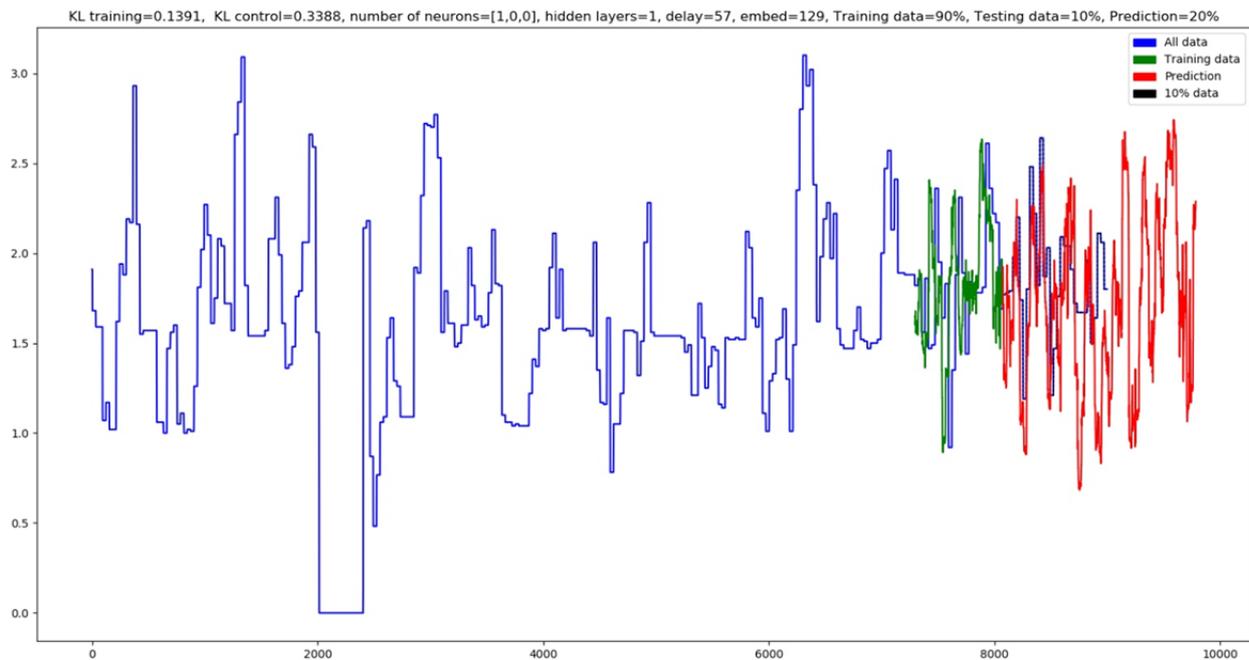
KL training=0.1391, KL control=0.3388, number of neurons=[1,0,0], hidden layers=1, delay=57, embed=129, Training data=90%, Testing data=10%, Prediction=20%

Figure 2. Forecasting result based on the enterprise's integrated data with a logistic activation function

Figure 2. shows the graphs of training and forecasting a recurrent neural network with an LSTM cell. It can be seen from the graph that the trained data dynamics quite accurately coincides with the initial ones, but significant problems arise in the forecasting implementation. The figure shows that the dynamics was not saved, and the forecast does not carry any semantic load. The coefficients KL and R^2 are 0.8991 and 0.5734, respectively, for the trained data, and for the predicted 78.6953 and 0.1129.

5. Results

As the experiments' result, it turned out that recurrent networks, which for the most part solve classification problems well, are not advisable to use for forecasting time series, since recurring ANN can retrain and lose previous settings during operation. In addition, the recurrent neural networks' operating time in the tasks of mathematical model development of complex nonlinear systems behavior is too long for practical use.

Thus, it has been experimentally shown that a multilayer perceptron is suitable as the most effective means of analyzing and forecasting data for HPF industrial safety control systems.

A multilayer perceptron allows training in almost real time, with the ability to effectively configure the neural network built on it, which, as a result, gives good predicted data dynamics.

The given conclusion was used in the neural network implementation in the "Zodiac" software for monitoring, analyzing and predicting the state of the HPF of the Russian FEC industrial safety.

References

[1]. Levin, S. E., Nagibin, S. Ya., & Shilov, V. V. (2018) Industrial Safety Monitoring of the Fuel and Energy Complex, National Security of Russia: Current Aspects: Collection of Articles of the All-Russian Scientific and Practical Conference (St. Petersburg, November 2018). *Saint Petersburg: Humanitarian National Research Institute "National Development"*.

[2]. Law, F. (1997). On Industrial Safety of Hazardous Production Facilities. Federal Law of July 21, 1997 №. 116-FZ.

[3]. Senatorov, M., Levin, S., & Nagibin, S. (2017). Distance control of process safety of Fuel and Energy Complex. *XVI International research and practical conference on "Engineering sciences – from theory to practice",* 16-24.

[4]. Presidential Decree (2018). On fundamentals of the Russian Federation state policy in the field of industrial safety for the period up to 2025 and further prospects. *Russian Federation Presidential Decree as of 06.05.2018 №. 198.*

[5]. Haykin, S. (2008). Neural networks: A Comprehensive Foundation. Second edition. *Moscow: Williams Publishing House.*

[6]. Loskutov, A. Yu. (2010). Time Series Analysis. *Moscow: Moscow State University, Faculty of Physics.*

[7]. Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980* (pp. 366-381). Springer, Berlin, Heidelberg.

[8]. Bezruchko, B. P., & Smirnov, D. A. (2005). Matematicheskoe modelirovanie i khaoticheskie vremennye ryady. *Mathematical modeling and chaotic time series Saratov: GosUNTs «Kolledzh*, 139-155.

[9]. Malinetskii, G. G., & Potapov, A. B. (2000). Modern problems of nonlinear dynamics. —Moscow: Editorial URSS, 335p. *eralized synchronization of chaos in directional coupled chaotic systems. Phys. Rev. E*, *51*(2), 980-994.

[10]. Nikolenko, S., Kadurin, A., & Arkhangel'skaya, E. (2018). Deep learning: introduction to the world of neural networks. *PITER, Saint-Petersburg & Moscow (In Russian).*

[11]. Sternberg, S. (1970). *Lectures on differential geometry.* Moscow: World.

[12]. Fraser, A. M. (1989). Reconstructing attractors from scalar time series: A comparison of singular system and redundancy criteria. *Physica D: Nonlinear Phenomena*, *34*(3), 391-404.

[13]. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115-133.

[14]. Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79-86.

[15]. Kullback, S. (1959). *Information theory and statistics*. New York: John Wiley and Sons.

[16]. Anil, K. J., Jianchang M., & Mohiuddin, K. M. (1996). Artificial Neural Networks: A Tutorial. *IEEE Computer, 29*(3), 31-44.

[17]. Vander, P. J. (2018). *Python for complex tasks: data science and machine learning.* Saint Petersburg: Peter.

[18]. Sebastian, R. (2017). *Python and machine learning.* Moscow: DMK Press.