Structural and Parametric Synthesis of predictive RBF Neural Networks using Artificial Immune Systems

Lytvynenko V.I[1], Bidjuk P.I[2]., Rogalsky A.F.[1] , Fefelov A.A.[1],Fedchuk V A.[1]

[1] *Kherson National Technical University, 73008 Kherson, Berislavske Shosse 24*

[1]*National Technical University of Ukraine "KPI", 03056, Kiev, Peremogy 37*

immun56@gmail.com, pbidyuke@gmail.com, kstu@tlc.kherson.ua, fao1976@mail.ru

Abstract. *In this paper we describe the use of clonal selection algorithm for the synthesis of radial-basis networks for solving the problem of time series prediction*

Keywords

Clonal selection algorithm, structural parametrical synthesis, RBF neural network, time series prediction

1 Introduction

Development, formal presentation and design of immune and hybrid immune systems suggests the presence of three components (Figure 1): 1) the scheme of the components of the AIS; 2) one or more measures to quantify the states of the system (affinity and measures to evaluate the fitness), and 3) immune algorithms that control the behaviour of the system.

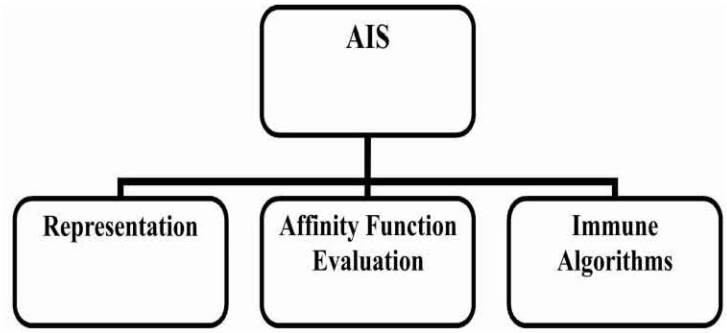


Fig.1 Structural components of AIS

The presentation accepted for immune cells and molecules, is an expanded version of the approach of the shape-space. Currently, the most frequently used are four main types of shape-space: real-valued shape-space, integer shape-space, symbolic shape-space. In addition to these forms can be used and more complex types of shape-space, such as neural networks, fuzzy neural network, fractal shape-space and shape-space of DNA. In this paper the optional choice for shape-space is selected based on RBF neural network architecture. Shape-space is a formalism designed to represent those parts of immune cells or molecules that determine the properties allowing the cell to recognize or to be recognized by other elements.

2 Theoretical Part

As shown in [1], the application of gradient methods of local search that are used to construct neuromodels in some cases is unacceptable or impossible. The generalized optimization problem of synthesis of a neural network on the training set can be formulated as follows:

$$Net = Net(\mathrm{M}, \Omega, \mathrm{B}, \Delta, \mathrm{A}), \tag{1}$$

for which

$$\zeta(Net, X, Y) \to \min, \tag{2}$$

where M is a matrix that determines the presence of synaptic connections between elements of the system (receptors, neurons); $\Omega = \Omega(\mathrm{M})$ is a matrix of weights, which correspond to those present in the network; $\mathrm{B} = \mathrm{B}(\mathrm{M})$ is biases vector of the neural network; $\Delta = \Delta(\mathrm{M})$ is a vector of the discriminant functions of elements of the neural network; $\mathrm{A} = \mathrm{A}(\mathrm{M})$ is a vector of activation function for neural network; $\zeta(Net, X, Y)$ is a criterion for determining the effectiveness of a neural network model to approximate the relationship between inputs X and their corresponding parameter vector of output values Y.

The optimality criterion for a neural network model using the mean square error:

$$\zeta = \sum_{p=1}^{m} \left(y_p - y(Net, \Psi_p)\right)^2, \tag{3}$$

where Ψ_p is a set of values for the p -th instance; $y(Net, \Psi_p)$ is a value of neural network output btained for the set of values Ψ_p.

2.1 Synthesis of computational structures using immune algorithms for solving forecasting problem

An important point in the process of forecasting is the selection of forecasting method. Depending on the properties of a time series, as well as the requirements for the process (for example, the required accuracy of the forecast or the rate of the forecasting) will be determined by the effectiveness of a method and quality of solutions obtained. In addition, a well known fact that the quality of forecasts can be improved by combining the (combination) of the results obtained by different methods. Neural network constructed on the radial basis (RBF network) is a powerful tool for approximating multidimensional nonlinear functions. They have a fairly simple architecture and have a high speed training. Block diagram of RBF network is shown in Figure 2. The RBF network consists of input, single hidden (radial basis) and linear (output) layers. The input layer consists of sensors that connect the network with the external environment. The neurons of hidden layer operate on the principle of centering on the elements of the training sample. The centers are supported by the weight matrix (W^r). The reactangular (dist) is used for calculating the Euclidean distance between input vector (X) and the corresponding center. Around each center there is a region called the radius. Radius (sensitivity of the network) is adjusted by means of the smoothing coefficients vector: $(\sigma_1, ..., \sigma_m)$. The transfer function (usually Gaussian, $f(x) = e^{-\frac{(x-c)^2}{2\sigma^2}}$), is varying in the range from 0 to 1, and it determines the output of a hidden layer. The output layer contains usual linear or sigmoidnye neurons; by adjusting their weights (W^l) we determine the network output.

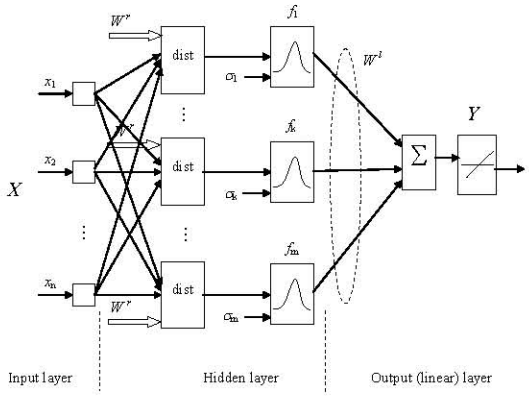


Fig. 2. Architecture of RBF network

The behavior of RBF network depends on the number and position of radial basis functions of the hidden layer. Indeed, for any real n-dimensional input vector $x = (x_1, x_2, ..., x_n)$, where $x \in X \subset \Re_n$ the output of the network will be determined as follows:

$$y_i = \sum_{k=1}^{m} w_{ik}^l f_k\left(dist\left(x, w_k^r\right), \sigma_k\right), \tag{4}$$

where, $w_{ik}^l \in W^l, i = \overline{1, p}$ is a weight of the linear layer; $w_k^r \in W^r$ are centers of radial basis functions. If as a basic function is used the Gaussian function, then:

$$f_k(x) = -\frac{dist(x, w_k^r)^2}{2\sigma_k^2}, \; k = \overline{1, m}. \tag{5}$$

In the context of approximation problem of the network configuration is to find a function, $y : \Re_n \to \Re$, satisfying equation (5) at $p = 1$. Suppose we have a sample of training data points : $X_1, ..., X_S, X_i \in \Re_n$. If the output values for each of these points of $d_1, ..., d_S, \quad d_i \in \Re$ are known, then every basis function can be centered on one point of X_i. Consequently, asymptotically the number of centers, and therefore the hidden layer neurons will be equal to the number of data points of the training sample, M. In this case there are at least two problems. First, low ability to generalize as far as the presence of excessive number of neurons in the hidden layer has a negative impact on the approximation of the new (not participating in the training data), and second, a large size of the training sample will inevitably cause problems of a computational nature. To overcome these difficulties the complexity of the network should be reduced by reducing the number of basis functions, which in turn poses a new challenge, that touches upon their optimal centering. The traditional methods of determining the centers of RBF are: the random selection of vectors from the set of training data, the application of clustering algorithms working on the scheme of unsupervised learning, the application of supervised learning. The basic idea of the work is to use an immune network for identification of centers of radial basis functions, i.e. solving the problem of recognition and clustering, that solves the problem of determining the number of input values. After that a clonal selection algorithm is used for constructing an optimal architecture of radial basis neural network (number and type of RBF-neurons in the hidden layer functions) as well as optimizing the weights and parameters of radial basis functions. As the output layer activation functions, logistic and linear function activation are used. The traditional methods of determining the RBF centers are: the random selection of vectors from the set of training data, the application of clustering algorithms working on the scheme of training without a teacher, and the use of supervised learning schemes.

Table 1. Radial-basis activation function of the inner layer

Gaussian function	Multi-quadratic function	Inverse Multi-quadratic function	Spline	Cauchy function

$f(x)=\exp\left\{-\frac{(x-c)^2}{2\sigma^2}\right\}$	$f(x)=\left(\frac{(x-c)^2}{\sigma^2}+a^2\right)^{\frac{1}{2}}$	$f(x)=\left(\frac{(x-c)^2}{\sigma^2}+a^2\right)^{-\frac{1}{2}}$	$f(x)=x^2\log(x)$	$f(x)=(1+x)^{-1}$

Following the idea of an integrated approach to solving the problem of setting parameters of a neural network, we used the clonal selection algorithm as a single (global) tool for searching the optimal values for all configurable parameters. Below is a description of the elements of AIS, which must be adapted to the task.

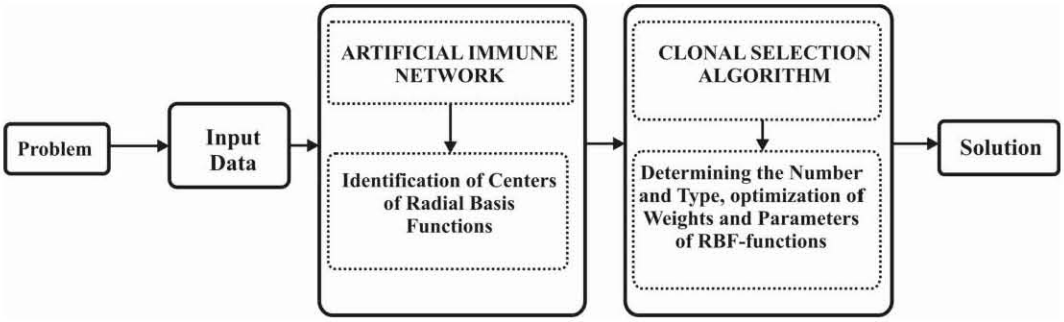


Fig. 3. A generalized scheme of the synthesis and study of radial basis neural network using clonal selection algorithm and the algorithm of artificial immune network

Based on the architecture of the neural network (Fig. 2) as adjustable parameters are the following: a) the number of neurons in the hidden layer (m); b) centers of radial basis functions (w_k^r); c) coefficients of smoothing (σ_k); d) types of basic functions of the hidden layer; e) the weight of output layer (w_{ik}^l); f) type of activation function of output layer; g) parameters of the activation function of output layer (a).

On the basis of selected parameters could be obtain a structure of an individual AIS as shown in Fig. 2.

f_1	f_2	…	f_m	w_{11}^r	…	w_{1n}^r	…	w_{m1}^r	…	w_{mn}^r	σ_1	…	σ_m	w_1^l	…	w_m^l	a

Fig. 4. The structure of the individual (antibody) AIS coding RBF-network

To encode the values by the binary system the precision (bits per value) is highlighted as a parameter setting AIS. Elements of the string $f_1, ..., f_m$ encode the status of the neurons in the hidden layer. The value of "0" corresponds to the passive or „off“ (the neuron is not involved in calculating the output value network). The value of "1" shows that the neuron is active (enabled). This scheme provides an automatic search for the optimal number of hidden elements of the RBF network. Configured AIS provides only the maximum possible number of these elements. As a target function and the function of the affinity is selected the standard error of the network on training data. As AIS learning algorithm is chosen a clonal selection algorithm with the following implementation features: (a) selection is implemented according to the principle of tournament selection, which makes it possible to manage the convergence of the algorithm and maintain the diversity of the population at the appropriate level; (b) according to the specific binary coding scheme that is proposed for mutation, whereby the probability of changing a single bit line depends not only on the affinity of antibodies in general, but also on the significance of this bit.

Consider the example shown in Fig. 5.

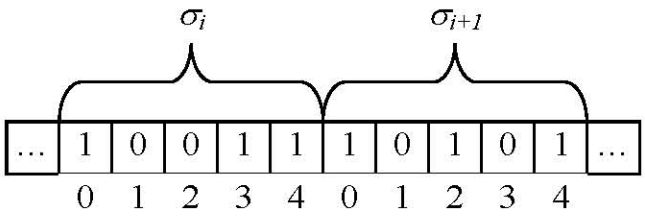


Fig. 5. Land-line antibody with binary encoding

In this example, the mutation of the bits with indices 0 or 1 obviously cause more significant changes in selected parameters than the mutation-bit number 4. When an individual reaches a sufficiently high affinity, more significant bits should be gradually excluded from the operation of mutation. This can be achieved by reducing the corresponding probabilities of the operator to the data bits. Formally, this process can be depicted as follows:

$$p_i = p_m \cdot \left(aff + \frac{i}{l}\left(aff_{\min} - aff\right)\right), \qquad (6)$$

where p_i is the probability of mutation of the i-th bits of the parameter in the string antibodies; p_m is the overall level of mutation, defined as a parameter of the algorithm; aff is the current affinity for antibodies; l is precision (bits) representation encoded parameters; $aff_{\min}$ is the minimum value of affinity, corresponding to the selected target function and the function of affinity.

Since in this case we solve the problem of minimizing an approximation error, the minimum value of antibodies affinity should be consistent with the maximum possible value of the error. Equation (6) provides the same probability of bit mutation for individuals with low affinity and increases the probability of mutation for less significant bits for individuals with high affinity (Fig. 6). The types of basic functions and the activation function of a layer are given as parameters to AIS. As a target function and the function of the affinity selected the standard error of the network on training data. As a learning algorithm is selected IMS clonal selection algorithm, which has similar features for implementation as the algorithm developed for the synthesis and training RBF networks.

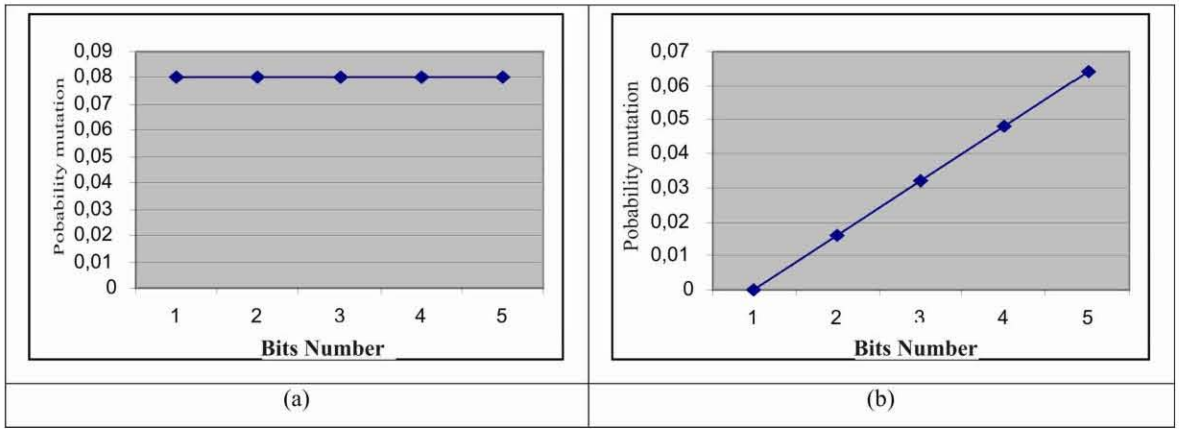


Fig. 6. Distribution of the probability of mutation encoded parameters for individuals with low (a) and higher (b) affinity

Stepwise implementation of clonal synthesis algorithm of predictive neural network models is shown in Figure 7.

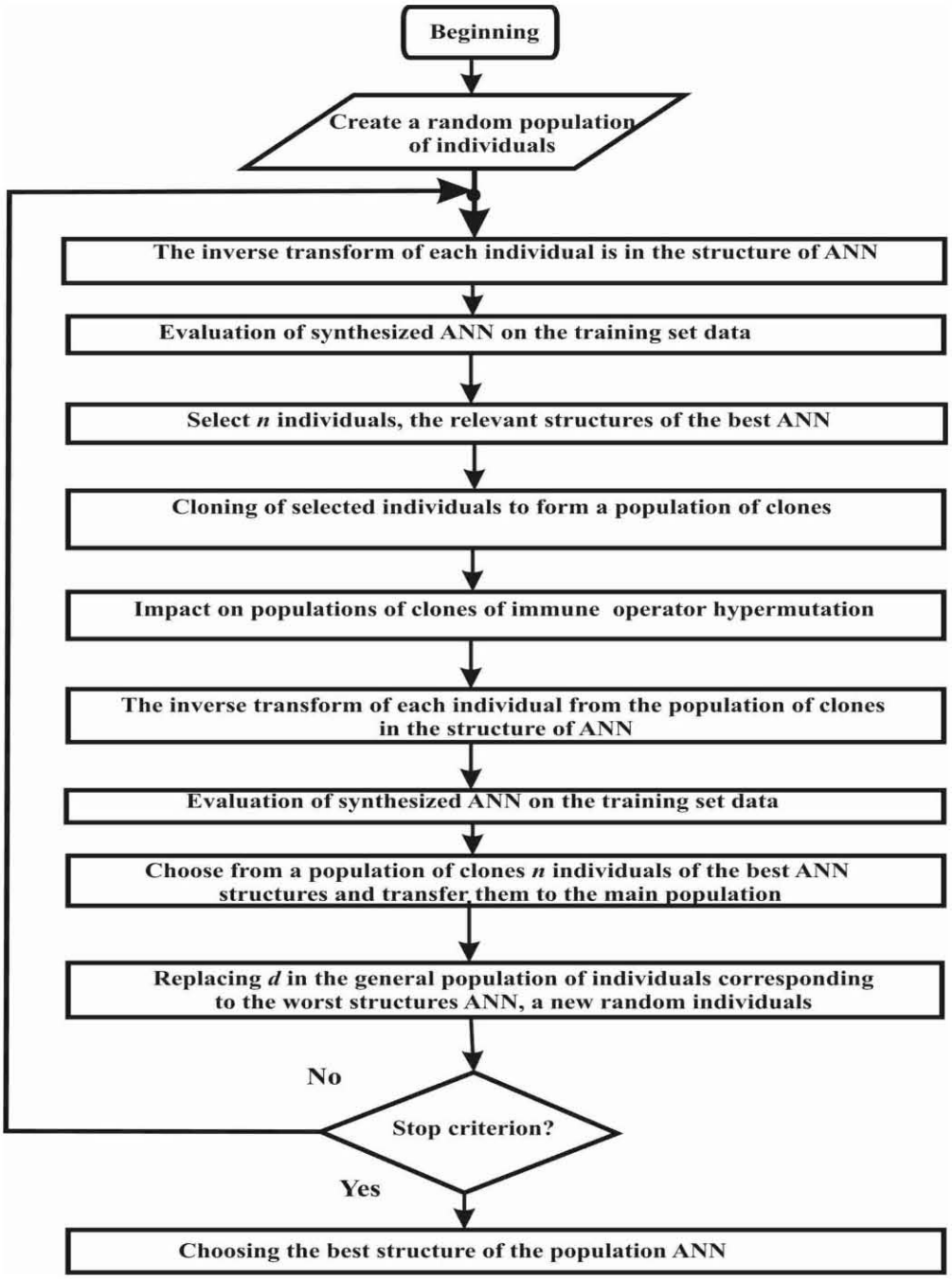


Fig. 7. A block diagram of an immune algorithm synthesis and configuration of neural networks for solving prediction

The algorithm for evaluating each solution obtained is converted back to individual lines in the structure of the neural network with the appropriate settings, which is checked for the training set to obtain the mean square approximation error.

3 Experiments

3.1 The description of experimental data

For the pilot study were selected the two time series representing real processes. First some observations on the volume of monthly sales of tickets American Airlines for 12 years; the series contains 144 observations. Its schedule is shown in Figure 8.

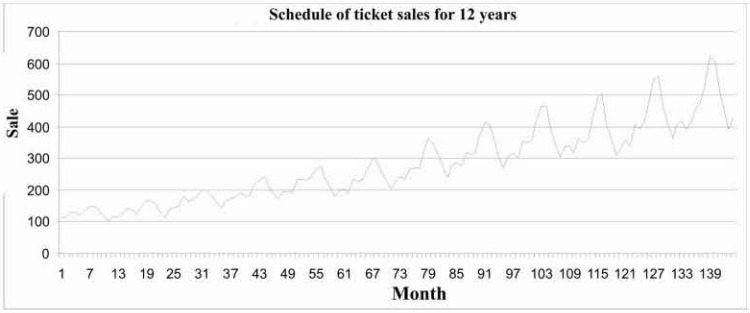


Fig. 8. Some observations on the volume of monthly sales of air tickets

The first 100 observations (70%) series were used as training sample, the remaining 44 cases (30%) were used as a test sample. The size of the minimum time lag (immersion depth d) was calculated on the basis of the partial autocorrelation function (PACF). From the PACF values $d \geq 13$, we have chosen the value $d = 14$. As a second example of the data we selected a number of observations of the daily consumption of electricity for one of the regions in Ukraine. The observations were collected during a year, so the series contains 365 values (Fig. 9).

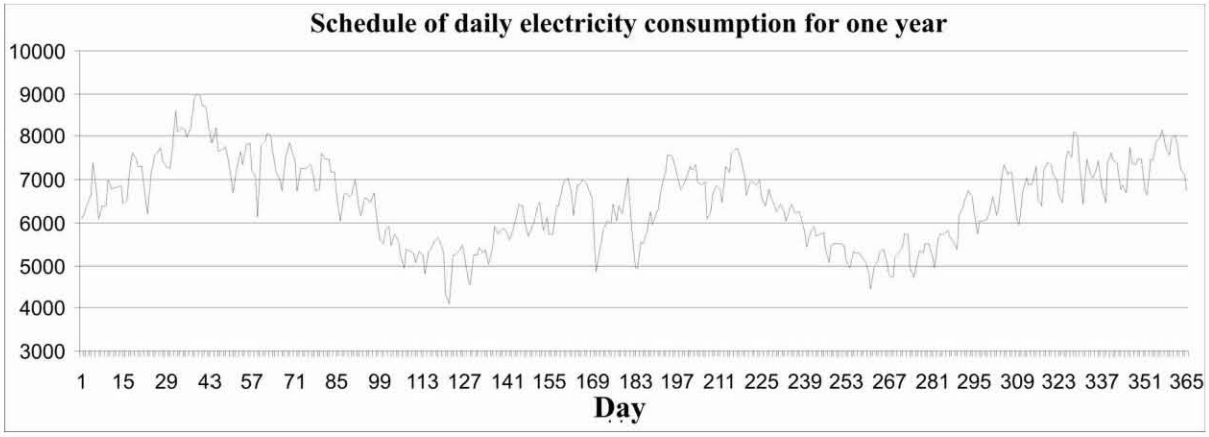


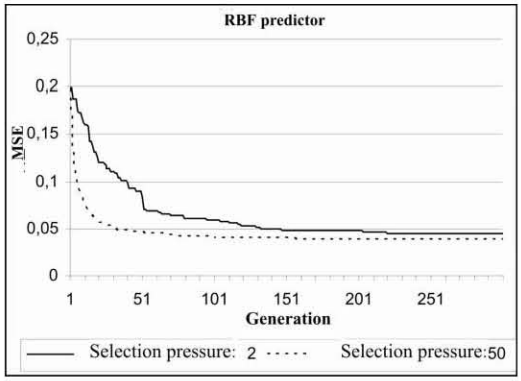
Fig. 9. A number of observations of daily consumption of electricity

To study the process we used the observations No. 260-330. The last 35 observations were used as a test sample. In this case the minimum time lag, $d \geq 15$, we have chosen the value of $d = 15$. Further on, based on initial data, we demonstrated experimentally the convergence of the developed methods and investigated the influence of main parameters on the AIS learning algorithms.

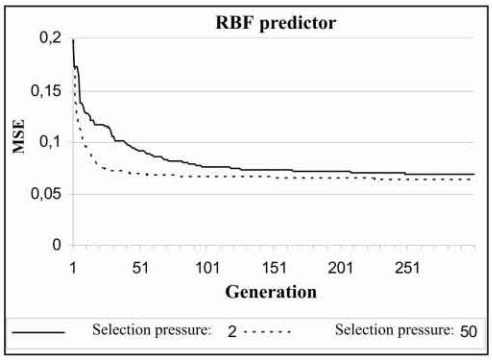
3.2 The influence of some parameters on the convergence of AIS algorithms

The experiments were set to investigate the influence of three main parameters of AIS: selection pressure, population size of clones, and the level of mutation. First, we set a high selection pressure equal to 50, the size of the population of clones equal to 300, and the level of mutation equal to 0.8. In this case the selection of the best antibodies will occur as follows. After assessing the entire population, i.e. 50, the antibodies are selected by a certain percentage of the best of them for subsequent cloning. This percentage is set by parameter "factor of selection of the best antibodies, which in this case is 0.7. Consequently, for cloning we must select, 50 * 0,7 = 35, the best antibodies. The best antibodies (i.e., those which give the smallest error of approximation on the training data) are chosen by tournament selection. The tournament selection involves a random selection from the population, the number of antibodies, specified by the parameter “selection pressure”, and the choice of one of the best out of this amount. The tournament is repeated as many times as antibodies should be selected (in this case 35 times). It follows that the larger the tournament is the less likely is a penetration of "weak" antibodies into the population of clones, and the faster should converge the immune algorithm. Then we set a minimal selection pressure, equal to 2, and compared the results obtained. Figure 10 (a) shows the graphs of convergence of the developed combined algorithm for the problem of forecasting the time series of ticket sales and Figure 10 (b) shows the similar experiments with a number of observations of daily electricity consumption. The results show that: (1) developed algorithms converge to the minimum learning error, which proves

the possibility of their use for solving the apporximation problems; (2) parameter selection pressure can control the rate of convergence of algorithms that can be effectively used to prevent premature convergence to local optima.



(a)



(b)

Fig. 10. (A) Convergence of the combined algorithms at different levels of selection pressure (for the time series of the monthly ticket sales); (b) Convergence of the combined algorithms at different levels of selection pressure (for the time series of energy consumption)

More studies have been performed touching on dependence of the time spend for learning and loss errors on the test sample from the population size of clones AIS. In these experiments the rigidity of selection for all algorithms was set equal to 20; the experimental results are shown in Table 1.

Table 1. A comparative study of the effect size of the population of clones

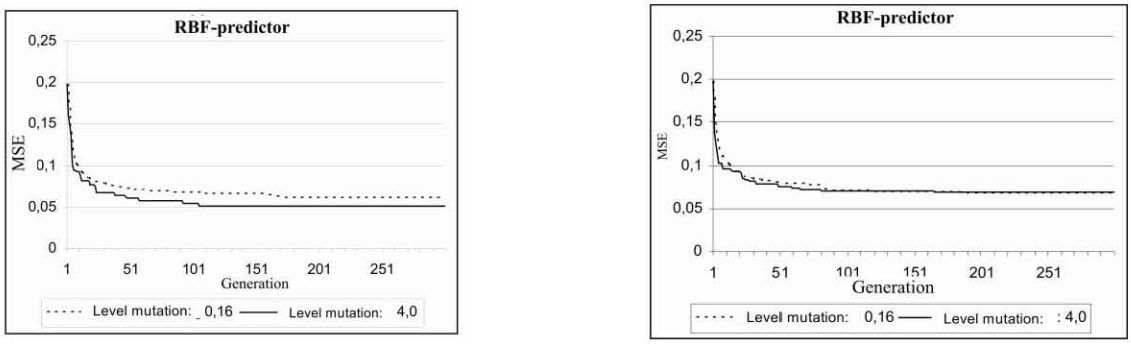
Algorithm type	Population size of clones	Data (time series)	Training time	PMSE on test sample
RBF-predictor	100	Sales tickets	35 c	12,65 %
RBF-predictor	300	Sales tickets	120 c	5,34 %
RBF-predictor	100	Electricity consumption	28 c	5,17 %
RBF-predictor	300	Electricity consumption	93 c	4,26 %

Here PMSE is the standard error in percentage, which is calculated via the formula:

$$D = \sqrt{\frac{\sum_{i=1}^{r} \frac{\left(y_i - y_i^M\right)^2}{y_i^2}}{r}} \cdot 100, \qquad (7)$$

where D is a mean square error (in percentage) for a model on training data; r is the size of test sample; y_i is the true value of a variable number; y_i^M is predicted value of a variable estimated with the model.

According to the results shown in Table 1, we can conclude that the increase in population size results in clones on one hand slows down the learning process, on the other hand in improvement of the generated models quality, judging on the value of the model error on the test sample. Furthermore, as seen from the table, combined algorithm IMS and wavelet neural network shows a higher accuracy of prediction than the combined algorithm and RBF network, while the latter has a higher rate of training. The experiments were conducted with the levels of mutations 0,16 and 4,0. The graphs of the convergence of combined algorithms for time series of the monthly ticket sales are presented in Figure 11 (a). Under the same conditions the convergence of algorithms for the time series of energy consumption is shown in Figure 11 (b).



(a) (b)

Fig. 11. Convergence of the combined algorithms at different levels of mutation: (a) for the time series of the monthly ticket sales; (b) for the time series of energy consumption

The experimental results show that for a high level of mutation, which implies a high variability of clones of the population, in most cases the training is faster, but the step character of the curve indicates a low stability of the process, thus decreasing the probability of finding the global optimum. The figure shows that the value of errors corresponding to the high level of mutation decreases rapidly at the beginning of training. But it is falling at a certain time in one of the local optima, and can not leave it in the future due to the high variability of antibodies, which leads to deterioration in the quality of training. In fact the mutation is the main driving force behind the evolution of the immune system and therefore requires more careful adjustment in accordance with the objectives set in the solution of a problem of forecasting.

4 Conclusion

The main results of this study are as follows:

- the experimental analysis of the problem of finding the settings for RBF neural networks;

- the combined methods of forecasting time series with controlled parameters is proposed, based on the synthesis of RBF networks using artificial immune systems;

- the results of computational experiments showed the higher effectiveness of the proposed combined methods.

References

[1] Subbotin, S.O, Oleinik S.A, Oleinik O.O. Not iterative, evolutionary and multi-agent methods of synthesis of fuzzy models and neural networks: Monograph / Edited by S.O. Subbotina. - Zaporozhye: ZNTU, 2009. – 375 p. (UKR).

[2] De Castro L. N., Timmis J. Artificial Immune Systems as a Novel Soft Computing Paradigm, Soft Computing Journal, vol. 7, Issue 7, July 2003.

[3] Ivakhnenko A. G.: The Polynomial theory of complex systems, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-1(1):364-378, 1971.