

## The Application of Evolutionary Algorithms in the Artificial Neural Network Training Process for the Oilfield Equipment Malfunctions' Forecasting

I.S. Korovin, Head of the neural network systems laboratory  
Scientific Research Institute of Multiprocessor Systems  
of the South Federal University, Taganrog, Russia  
korovin\_yakov@mail.ru

M.V. Khisamutdinov, Researcher  
Scientific Research Institute of Multiprocessor Systems  
of the South Federal University, Taganrog, Russia  
L-V-P@yandex.ru

A.I. Kaliaev, Researcher at the neural network systems laboratory  
Scientific Research Institute of Multiprocessor Systems of the South Federal University, Taganrog, Russia  
anatoly.kaliaev@gmail.com

**Abstract**—The paper describes an evolutionary approach to artificial neural network (NN) training, which is used to determine the state of oil-production equipment. A new artificial NN weight coefficient coding method using multi-chromosomes is proposed. The genetic operators of crossover and mutation applied to multi-chromosomes are examined. A genetic algorithm structure of artificial NN training based on the developed genetic operators is proposed. A comparison of the proposed approach to NN training with existing ones has been carried out.

**Keywords**—neural network, genetic algorithm, oilfield equipment, forecasting, malfunction, mutation, crossover

### I. INTRODUCTION

Today, one of the most vital problems for the oil companies in Western Siberia is the problem of prime cost reduction of the produced output, due to significant manufacture of oilfields and the deterioration of oilfield equipment. The most possible way to solve the above mentioned problem of reduction of failures in the oil wells fund work is to decrease the number of outages in shifts by means of carrying out qualitative and timely procedures of diagnostics, forecasting and online management. Despite the relatively large amount of automated control systems in the market of information technologies for the petroleum field, the effect of their usage is rather poor because of several basic reasons (we speak about Russian Federation):

- A. *Low degree of adaptation (as all systems are abroad developed without taking into consideration domestic industrial specifics).*
- B. *High cost of software so as of its after sales service.*
- C. *Inapplicability or limited functionality of data analysis approaches, methods and of which the applied systems.*

Therefore, the use of automated systems based on the Data Mining (so as Big Data, Knowledge Database

discovery) methods is effective and justified. In the works [1, 2] new methods of oilfield equipment diagnostics based on intelligent data handling technology with the use of multilayer artificial NN technology have been proposed, however the problem of oilfield equipment state forecasting has not been solved yet (we take into consideration our research).

Research based on the results, depicted in [1,2], have revealed that classical NN architecture can be successfully applied in classification tasks; however speaking about forecasting the given NN architecture is slightly applicable due to narrowly focus of the search space. It is known that one of the probable effective approaches to expanding the search space is the complication of NN architecture [3] in the form of hierarchical NN [4], and the application of Kalman recursive filter to forecast oilfield equipment states [5, 6]. However, existing methods of hierarchical NN training are inefficient and require significant time for their proceeding, thus the evolutionary approach to hierarchical NN training, which will increase the performance of the developed decision support systems for oilfield equipment operators, is proposed in this paper.

### II. MATERIALS AND METHODS

The main problem of the existing approaches to NN training is the retraining problem related with the NN weights coefficients adjustment method [1]. A possible solution of this problem is an application of hybrid approach to NN training, combining classic [2] and evolutionary approaches [7]. In the conditions of uncertainty evolutionary methods, including genetic algorithms, have a better chance to achieve the demanded results. Classic genetic algorithm operates a binary system, although recently some works are published in which genetic algorithm operators perform operations on sets of real numbers. This can significantly increase a capability of the described algorithms application [8].

Let us consider the mathematical model of the assigned task. We have a vector  $X$  with dimension  $N$ , which is a set of parameters, describing the oilfield equipment state. There is

also a vector  $Y$  with dimension  $K$ , which reflects the required result of recognition, indicating the equipment state. It is necessary to find a weighting matrix  $P$ , whose elements are real numbers within the interval  $[0, 1]$  that would fulfill the equation:

$$X \cdot P = Y.$$

It is necessary to proceed a weight matrix  $P$  setting using a genetic algorithm. In the context of the current task the matrix under consideration has the following dimensions: the number of rows is equal to  $K$ , which represents all types of states that describe the oilfield equipment work mode; the number of columns  $N$  is determined by the size of the dot-matrix field allotted for each state.

In the view of obviousness and simplicity of understanding each column of weight matrix should be considered as chromosomes, resulting in the presence of  $N$  chromosomes, which will aggregate each individual [8, 9]. However, this approach is valid if the number of inner layer neurons does not exceed the dimension  $N$  of the input vector  $X$ .

Let us consider the general case of individual coding (NN weight matrix). In the developed genetic algorithm each individual is represented as a set of chromosomes that describe the links am the between NN layers. The structure of the weight matrix in the form of multi-chromosomes is illustrated in Figure 1, where  $W$  – is the number of NN hidden layers,  $T^W$  – the number of neurons in the hidden layer  $W$ .

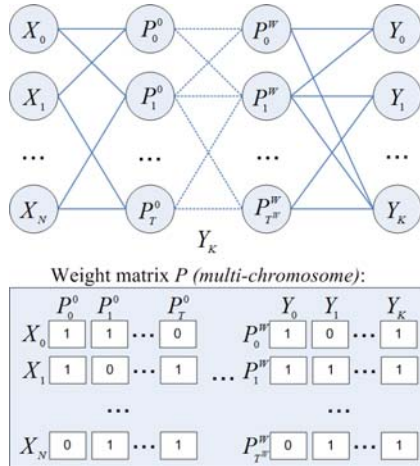


Figure 1 - Weight matrix representation in the form of multi-chromosomes.

Thus, the storage cell number, necessary for one individual SumGen, can be calculated by the formula:

$$SumGen = N \cdot T^0 + \sum_{i=0}^{W-1} T^i \cdot T^{i+1} + T^W \cdot K. \quad (1)$$

The effectiveness of any genetic algorithm operation is directly proportional to the crossover operation efficiency.

Furthermore, the genetic algorithm performance depends above all on the crossover success [10]. In the context of the current task a well-ordered crossover operator has been implemented, which provides a phased genetic material transformation, getting closer to the optimal solution. Figure 2 demonstrates the process of a getting new individuals with the use of the well-ordered crossover. There are two parental chromosomes:  $H1$  and  $H2$ .

Real numbers between 0 and 1 are genetic material. The well-ordered crossover works as follows. Originally "dividing point" is determined by random select. At the next point, the first descendant  $new\_H1$  inherits left side of the parental chromosome  $H1$ . Filling the remaining genes of a new chromosome is proceeded by the information stored by the other parent  $H2$ . The algorithm examines chromosome  $H2$  from the outset and performs genes extracting which differ from the genes already present in the descendant. A similar procedure is performed to get the second descendant  $new\_H2$ . The second descendant  $new\_H2$  inherits the left side of the parental chromosomes  $H2$ . Filling the remaining genes of the derived chromosome is proceeded by the information stored by the other parent  $H1$ .

Algorithm produces chromosome  $H1$  analysis from the first gene and provides well-ordered genes extraction which differ from the genes already present in the descendant over more than desired value. The work of each crossover operator in the population results in appearance of two new individuals. To control the amount of crossover operations crossing coefficient is used, determining the fraction of produced descendants in each iteration.

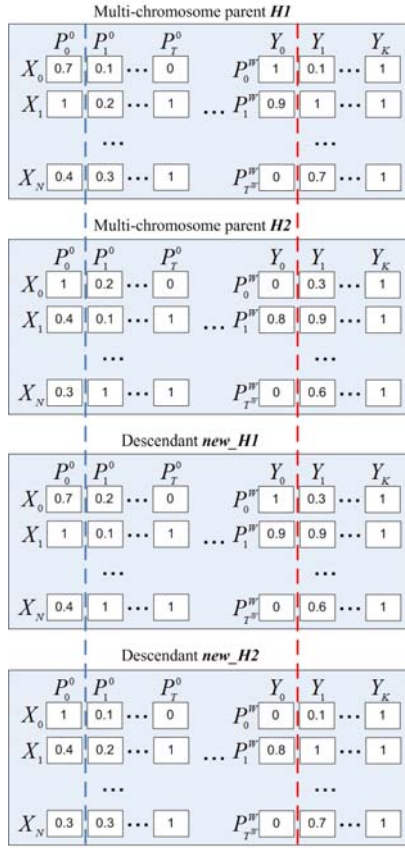


Figure 2 - The principle of the well-ordered crossover operation for multi-chromosomes.

The high value of the crossing coefficient allows to increase the number of search space fields and reduces the risk of getting to a local optimum. However, too high value of this parameter will increase the time of the algorithm operation, and result to the excessive research of search space unpromising areas.

The next stage of the genetic algorithm is a mutation. A mutation is a change that leads to the appearance of qualitatively new properties of genetic material [11]. Mutations occur randomly and cause abrupt changes in the structure of genotype [11].

In the context of the optimization task solution gene mutations are of the greatest significance that involve in the most cases one or more genes. The mutation can look any way, whether it's genes exchanging positions or another gene value copying, etc. Then it is necessary to choose mutation type in each genetic algorithm. In the genetic algorithm under consideration genes store real numbers from 0 to 1. According to this, the mutation operator has to make specific changes to the genetic material, i.e. to change the values of certain genes without taking into consideration already existing ones.

The essence of the developed mutation operator is as follows. In the chromosome under consideration a genes number is chosen by random selection. The mutation coefficient determines the mutation intensity. It estimates the

fraction of genes mutated in the current iteration, counting on their total amount. If the mutation coefficient is too small, it will lead to a such situation where an array of useful genes will not exist in the population. At the same time, the use of high coefficient of mutation intensity will bring a plurality of random perturbations and increase search time significantly. The descendants will cease to resemble their parents, the algorithm will no longer be able to be trained on the basis of hereditary trait preservation. The selected genes are transformed and that causes the current gene value change on some small value. This value is chosen so that, after changing the value of the  $i$ -th gene, it should not exceed the interval  $[0, 1]$ .

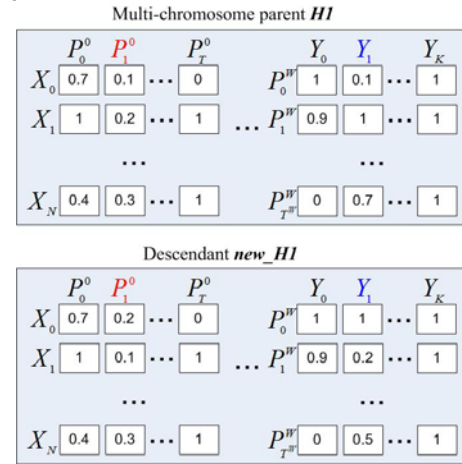


Figure 3 - The principle of the mutation operator work for multi-chromosomes.

Figure 3 shows a mutation process for multi-chromosomes. Genes in the locus  $P_1^0$  and  $Y_1$  have successfully received new values and it will lead to changes in reading of mutating individual fitness function. The use of mutations allows to bring a new genetic material into the population. This will increase the search space that is necessary for efficient optimum search. Of course, it makes sense to use the classical mutation operator, which is based on a random change of the gene order. This also leads to quite good results in the early stages of a genetic algorithm operation.

Let us consider the generalized structure of hierarchical NN training method on the basis of an evolutionary approach with the use of the proposed genetic operators (Fig. 4).

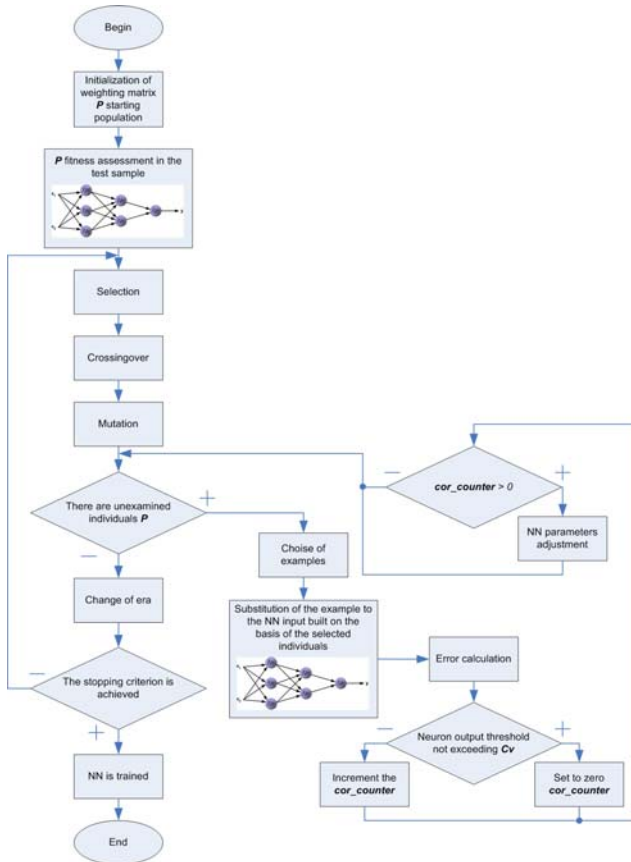


Figure 4 - Hierarchical neural network training method structure based on an evolutionary approach.

The underlying training method is the one with the accumulation of the NN parameter vector [8] adjustments. The basic idea an evolutionary approach to NN training application is to store a variety of alternative NN weight matrices, which encoding is performed using multi-chromosomes.

The first stage of the method is to initialize starting population of weight matrix  $P$  individuals, where each individual is represented as multi-chromosomes (see Fig. 1). The first gene initialization is carried out on the basis of a shotgun experiment. Next is the assessment of each population  $P$  individual fitness using a test training sample. The training quality criterion in this case is the minimum total deviation of the NN recognition result of the test data from the sample.

After the formation of the individuals of the first epoch genetic operators realization is implemented - selection, crossover and mutation on the basis of the above proposed approaches (see Fig. 2 and Fig. 3).

Each individual quality in the population is improved further on using training method with NN parameter vector adjustment accumulation. For each individual a predetermined iterations' quantity of weight matrix adjustments is performed.

The received set of modified individuals passes a selection in order to maintain a given number of individuals

in the population. Thus, a change of era occurs and a new generation of evolutionary algorithm begins.

At the end of each generation a stopping criterion achievement is checked, which consists in the examination of all individuals in the population, namely: if there is at least one individual, neurons output threshold of which has not exceeded  $C_v$  in the training process on the test data, the algorithm is stopped. Otherwise, a transition to a new generation of the evolutionary algorithm occurs.

### III. EXPERIMENTAL RESEARCH AND BENEFITS

To evaluate the characteristics of the proposed method of NN training, standard test data from the site UCI Machine Learning Repository [12] have been taken. The given benchmark data are used to compare the classification quality of the proposed intelligent data analysis approaches, including neural networks and genetic.

The following table provides a set of standard test data, the hierarchical NN training results with the given parameters, as well as a comparison results of the proposed approach with the well-known ones to the NN training on the basis of genetic algorithm (GA) and Bayesian optimization algorithm (BOA) [13]. Testing methodology is analogous to that considered in [13].

TABLE I. EXPERIMENTAL RESEARCH RESULTS OF THE PROPOSED APPROACH

Test set name	Number of classes	Number of fields	Number of records	Test sample set recognition results (%)		
				GA	BOA	Hierarchical NN, evolutionary training
Breast Cancer	2	9	699	95.84	96.42	94.98
Credit-Australian	2	6	653	86.18	85.84	85.74
Credit-German	2	7	1000	70.30	70.14	70.50
Heart-Cleveland	2	6	303	88.78	89.37	89.32
Housing	3	12	506	64.18	66.24	65.51
Ionosphere	2	34	351	82.22	84.22	84.29
Iris	3	4	150	67.73	93.60	87.63
Pima-Diabetes	2	8	768	76.04	75.88	76.04
Segmentation	7	19	2310	64.32	63.66	67.59
Sonar	2	60	208	84.90	83.55	83.12
Vehicle	4	18	846	76.64	78.62	79.87
Wine	3	13	178	87.41	93.48	92.57
Random21	2	21	2500	94.03	94.09	93.25
Redundant2	2	21	2500	95.82	95.72	94.63

As the table shows, the proposed method of NN hierarchical training is not inferior to existing approaches to NN training on the majority of test sets, and exceeds them owing to a large search space due to hierarchical NN structure in tests with a sufficiently large number of fields and classes (such as Segmentation, Vehicle). Further improvement of the proposed approach is possible by the

application of parallel NN training operations, on the basis of approach adaptation proposed in [14].

#### IV. CONCLUSION

Thus, speaking about oil and gas industry industrial objects' state forecasting task, the proposed approaches, methods and algorithms possess the following major advantages:

- In comparison with existing weight matrix coding methods the proposed encoding method enables to represent any artificial NN architecture including hierarchical one.
- Method of evolutionary hierarchical NN training in comparison with existing approaches to training provides a solution of NN retraining task using multi-chromosome population that significantly improves the quality of emergency situations forecasting in the oilfield industrial equipment work.
- The main idea of the proposed method of hierarchical NN weight matrix genetic search is to use a directed training on the basis of NN parameter vector accumulation's adjustment, thereby achieving a high rate of solution search, and therefore it becomes possible to increase the performance of decision support systems for operators of oilfield equipment, developed on the basis of the proposed method.

© Iakov Korovin, 2013

#### REFERENCES

- [1] Korovin Y.S., Tkachenko M.G., Kononov, S.V. Operational diagnostics of oil-field equipment based on intelligent data handling technology. *Oil Industry-2012*, -vol. 9, -pp. 116-118.
- [2] Korovin Y.S., Tkachenko M.G. The use of neural network analysis in the oil and gas industry // *Proceedings of the Southern Federal University. Technical sciences-2010*, -vol. pp. 172-178.
- [3] Viharos Zs.J., Monostori L., Vincze T., Training and application of artificial neural networks with incomplete data, *Lecture Notes in Computer Science*, Springer-Verlag Gmb, 2002. Vol. 2358, p. 649.
- [4] Aksenov S.V., Novosel V.B. The organization and the use of neural networks (methods and techniques) / under general editorship of V.B. Novoseltseva. - Tomsk: Publishing House of NTL, 2006. – 128 p.
- [5] Feraud R., Clerot F., Simon J.L., Pallou D., Labbe C., Martin S. Kalman and neural network approaches for the control of a vp bandwidth in an atm network. *Lecture Notes in Computer Science*. Springer-Verlag Gmb. 2000. Vol. 1815, p. 0655.
- [6] Perez-Ortiz Ju.A., Schmidhuber Ju., Gers F.A., Eck D. Improving long-term online prediction with decoupled extended kalman filters. *Lecture Notes in Computer Science*. Springer-Verlag Gmb. 2002. Vol. 2415, p. 1055.
- [7] Rabunal Ju.R., Dorado Ju., Pazos A., Rivero D., rules and generalization capacity extraction from ann with gp, *Lecture Notes in Computer Science*, Springer-Verlag Gmb, 2003. Vol. 2686, pp 606-613.
- [8] Mishchenko V.A., Korobkin A.A. Using genetic algorithms for training neural networks. *Modern problems of science and education-2011*, -Vol 6, pp. 116-119.
- [9] Tenenev V.A. The solution of multicriterion optimization with genetic algorithms. *Intelligent systems in manufacturing-2006*, -Vol. 2, pp. 103-109
- [10] Maor-Shoshani A., Reuven N. B., Tomer G., Livneh Z. Highly mutagenic replication by DNA polymerase V (UmuC) provides a mechanistic basis for SOS untargeted mutagenesis // *Proc. Natl. Acad. Sci. USA* — 2000. — 97. — P. 565—570.
- [11] Rutkowski L. *Methods and techniques of artificial intelligence*. - Moscow: Hotline-Telecom, 2010. — P. 520. — ISBN 5-9912-0105-6.
- [12] UCI Machine Learning Repository, URL:<http://archive.ics.uci.edu/ml/>
- [13] Cantu-Paz E., Pruning neural networks with distribution estimation algorithms, *Lecture Notes in Computer Science*, Springer-Verlag Gmb, 2003. Vol 2723, pp.790-800.
- [14] Schuessler O., Loyola D., Parallel training of artificial neural networks using multithreaded and multicore cpus, *Lecture Notes in Computer Science*. 2011. Vol. 6593 LNCS. № PART 1. pp. 70-79.