Australian Journal of Basic and Applied Sciences, 4(5): 764-772, 2010
ISSN 1991-8178

Comparison of Optimized Neural Network with Fuzzy Logic for Ore Grade Estimation

Pejman Tahmasebi, Ardeshir Hezarkhani

Department of Mining, Metallurgy and Petroleum Engineering, Amirkabir University, Hafez Ave. No. 424, Tehran, Iran

Abstract: This paper presents a new approach to improve the performance of neural network method to estimate the grade values. The true value of ore body grade which is calculated based on the accurate data is a challenge of the mining industry. The main goal of the following investigation would be the performance comparison of various learning algorithms in neural network that could apply for ore grade estimation. Up to now, there is not presented procedure to determine the network structure for some complicated cases, therefore; design and production of neural network would be almost dependent on the user's experience. To prevent this problem, neural network based on genetic algorithm (ANN-GA) was applied to present the advantages, e.g. assay estimations. To show the performance of this method, three prevalent estimation methods such as artificial neural network (ANN) and fuzzy logic (FL). One of the most important problems in neural network designing is the topology and the value parameter accuracy that if those elements selection was correctly and optimally, the designer would achieve a better result. To test this new method, it was evaluated by a case study. First, the parameters and topology of neural network determined without applying the genetic algorithm (trial and error method) and in order to consider the effect of genetic algorithm on results, it was applied GA to obtain the parameters (the input number, number of neurons and layers in the hidden layers, the momentum and the learning rates) and then network performance. The results indicate that this method could improve the network performance rather than ANN and FL, also the mean square error (MSE) and R values decreased and increased respectively.

Keywords: Global learning algorithm, genetic algorithm, neural network optimization, fuzzy logic, grade estimation.

INTRODUCTION

The success of a mining project depends on the accuracy of the total reserve estimation as well as the deposit grade. Geostatistics and Neural Network (NN), the two prevalent techniques that are most frequently used today to estimate the ore grade (Ke, 2002; Yama and Lineberry, 1999). Geostatistical technique such as Ordinary Kriging works under the stationary condition assumption. Moreover, it is a linear model based on the local neighborhood structure. On the other hand, neural network is a non-linear model free estimator, which is robust in noisy and extreme conditions. Neural network performs better when there is a non-linear spatial trend data, which violates the stationary assumption of ordinary Kriging technique. Each of these has advantages and disadvantages (Cargill *et al.,* 1977; Carras, 1987; David, 1977). Several studies reported successful implementations of the neural network technique to estimate the spatial attributes. For example, Wu and Zhou (1993)] applied neural network for copper reserve estimation. Rizzo and Dougherty (1994) used these techniques to characterize the aquifer properties. Koike and others (Koike *et al.,* 2001) used the neural network to determine the principal metal contents of Hokuroku district in Northern Japan. Koike and Matsuda (2003) also used this technique to estimate the impurity content of a limestone mine namely, SiO_2, Fe_2O_3, MnO, and P_2O_5. Samanta and others (Samanta *et al.,* 2004) have also used neural networks (along with geostatistics) for grade estimation in bauxite and a gold deposit (Samanta *et al.,* 2004). In the bauxite deposit, neural networks and geostatistics showed almost equivalent performance, while for the gold deposit the neural network performed better than the geostatistics (ordinary kriging). The major reason for this rapid growth and diverse is their ability to approximate virtually any function in stable and efficient way.

Corresponding Author: Pejman Tahmasebi, Department of Mining, Metallurgy and Petroleum Engineering, Amirkabir University, Hafez Ave. No. 424, Tehran, Iran

Along the other Artificial Intelligent (AI) tools, fuzzy set was also applied in mining and geosciences. There are two types of the application available which included: a. integration with geostatistics for variogram selection (fuzzy-variogram) for uncertainty decreasing which by fuzzy arithmetic and prior information they assessment the completeness of mineral exploration, and b. reserve estimation: several researchers have applied this technique for reserve and grade estimation (Pham, 1997; Tutmez *et al.,* 2007). For example, Tutmez and others (Tutmez *et al.,* 2007) proposed a fuzzy modeling algorithm and estimates the reserve parameters based on spatial variability.

Several problems are involved with neural network designing. It is because of some reasons such as unsuitable network topology, wrong number of hidden layers, learning rate, momentum rate, the weights and biases values etc. All of these problems could separately affect the network performance. In some literature, it is mention that back-propagation has the ability to be applied with variety of hidden neuron layers but also; it was shown that one layer of hidden layer for most of the problems is enough (Hornik *et al.,* 1989; Funahashi, 1989; Cybenko, 1989; Hartman *et al.,* 1990). Furthermore, in some complex, very noisy dataset and when there is a few training data are available, one layer of hidden neurons would not be enough.

All of the discussed parameters have an effect on neural network performance. The last problem which may happen by these parameters is leading the network to local minima. The importance of each parameter (step size, momentum rat, number of hidden layers and etc.) is discussed severally in the literature.

Therefore, because of those problems it is necessary to avoid the network local minima and it should be tried to minimize its error which is based on the applied parameters and topology. Then, the local and global minima are the problems that the designer should paying attention to and try to use the methods so that it does not trap in local minima (Shang and Wah, 1996). In many of the previous studies, local learning algorithm has been used.

For example, quickprop algorithmp (Wu and Zhou, 1993), standard backpropagation algorithm (Yama and Lineberry,1999) and back-propagation algorithm with momentum (Koike and Matsuda, 2003) are all local learning algorithms. Although Singer and Kouda (Samanta *et al.,* 2004) used simulated annealing along with conjugate gradient method for network training to escape from local minima and Samanta and others (Samanta *et al.,* 2004) also applied the NOVEL and annealing algorithms to elevation the network performance. In the last investigation, it was only regularized the weights and the other parameters such as; the number of element processing in the middle layer, learning and the momentum rate were not optimized. Therefore, by applying their methods, the achieved final results did not show any improved performance. To find the optimal structure for a neural network, significant amount of time and effort is required, since the optimal structure is including the optimal weight, neuron, learning rate, momentum rate, processing elements (PEs) and layer numbers. In this study, the attempt is to optimizing those elements in order to ore grade estimation. Finally, the obtained results of this new methodology with ANN and FL were investigated for comparison.

Genetic Algorithm:

Genetic algorithm firs time was introduced as a powerful and effective optimization technique by John Holland. The GA principal is based on natural selection and human genes.

In pas few years, GA has been attention more that the other optimization methods. These attentions are because of the simple implementation and its ability to search or the optimal solution for complex problems. In general GA has five basic components which summarized by Michalewicz (2004):

1. Expression the problem in genetic rules
2. Creations an initial population for solution
3. A fitness function which based on that the process going on
4. Applying genetic operation which by using them the children and reproduction would be done
5. Determination the values for parameters of GA.

One of the basic elements of GA is the chromosome which should be used to code the variables. All chromosomes are composed of smaller elements which called "gene". These genes would be presented by alphabet or real-value. Each chromosome represents a potential solution to the problem which should be solved. All of the chromosomes values should be evaluate a standard which named "fitness". Therefore, the solutions which are selected and convey to next population (offspring) would be selected based on their fitness and also this process will be continued to produce fitter chromosomes through genetic operators (see Fig. 1).

After transformation chromosomes to new generation, some of them because of their low fitness would be eliminated. Also, two type of the transformation which used as genetic operator: *mutation* that allows the chromosomes to take part in selection again which are possible by changing the chromosomes values, and, crossover, which creates new chromosomes by combining parts from two chromosomes that these new

chromosomes called offspring. As these iteration would be proceed with more generation, there may not be much improvement in the population fitness and the chromosomes will be recognized as the final chromosomes which presents an optimal solution to the problem.

In this paper, genetic algorithm has been applied in order to optimizing the neural network's parameters. These parameters were the input vectors, the number of hidden Processing Elements (PEs), the learning rates and the momentum coefficient.

Designing of Neural Network by Genetic Algorithm:

As discussed the parameters which commonly are effect on network performance and should be optimized are: the input columns, the number of hidden PEs, number of memory taps, and the learning rates. Also, according to genetic algorithm which is based on evolutionary process, it was tried to select the best parameters for network in order to minimize its error and escape from trapping in local minima.

Genetic algorithm by using its three operators (selection, crossover and mutation) is able to find the best solution for this minimization problem (Fig. 1, 2). Also, this searching process will be continued until condition that named as the specified termination criterion which is a condition that the genetic algorithm's search would be ceased.

In this study, the criterion used to evaluate the fitness of each potential solution is the lowest cost achieved during the training run. The mean squared error is simply two times the average cost. The formula for the mean squared error is:

$$MSE = \sum_{j=0}^{P} \sum_{i=0}^{N} (d_{ij} - y_{ij})^2 / N \quad (1)$$

where P= number of input processing elements, N= number of exemplars in the data set y_{ij}= network output for exemplar i at processing element j and d_{ij}= desired output for exemplar i at processing element j.

The start of genetic algorithm will be by the smallest element of this tool which is named "*gene*". The values which these genes will carry are the network parameters that should be optimized. Actually, each of those genes is a representative of the network's parameter. Therefore, all of these genes will organize a chromosome. In the other hand, a chromosome is made up of several genes.

At the first step of genetic algorithm process, an initial population would be randomly created. It is obvious that this population itself is made of several chromosomes and each chromosome is made up of several genes which their values set randomly in the first trying. Then, the network will be trained be these initial values for the population and their importance would be cleared. It means that all of the chromosomes have a value which by using this values, it would be possible to compare the different chromosome, because, the aim of using genetic algorithm is to reach the best network's parameters. These parameters here are the number of input vectors, the number of elements in hidden layer, momentum coefficient and learning rate values that set by chromosomes in populations. All of the used parameters have some boundaries. For example, the momentum had a lower bound of zero and an upper bound of one.

The primary chromosome presented as a haploid chromosome which itself is made up of several numerical genes. Since, the parameters which should be optimized are four parameters, therefore, each of the chromosomes are making up four genes. The first, second, third and fourth gene represents the number of input, the number of neurons in the middle layers, the learning rate and momentum step respectively. Actually, by using these parameters network would be trained.

As mentioned earlier, genetic algorithm has three parameters which their values could have a great effect on the results. Therefore, the parameter's value should be selected carefully. Therefore, based on this strategy several combinations of those parameters in order to finding the best form of their value tested. Therefore, in the present research, uniform cross-over and uniform mutation operators were applied and their probabilities were adjusted at 0.9 and 0.01, respectively. Also, it was used the roulette wheel method for selection operator. Actually, each of these probabilities has their reasons. For example, the aim of using crossover is to combine to chromosomes in order to create a new chromosome. It is because of that may be this new chromosome obtain a better fitness rather than their original chromosome separately. Therefore, according to its probability this operator will act in the genetic algorithm evaluation. Obviously, this probability because of its discussed rule should be set fairly high (0.9 is the first good choice).

Another probability is the mutation probability which deal with chromosomes alteration. Actually, based on this operator, it is probable to change the gene's value in a chromosome which could lead to create a new chromosome and it will be added to genetic algorithm's competition. The aim of using this operator is also

because of to extend the genetic algorithm searching scope which may lead to a better solution. It also helps the genetic algorithm to avoid from ceasing in the local minima. Therefore, because of its importance, its probability should be set fairly low (0.01 is a good first choice).

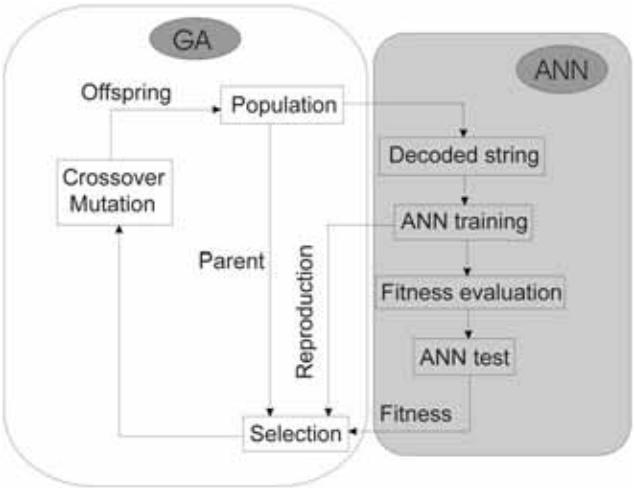

Fig. 1: Flowchart of proposed method.

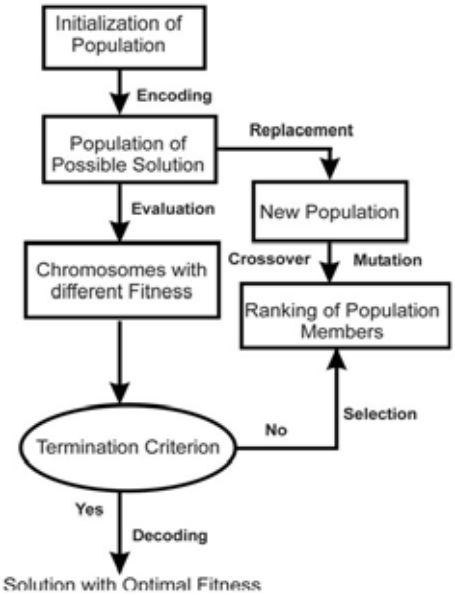

Fig. 2: Block diagram of Genetic Algorithm process.

Finally, the genetic algorithm would be started with an initial population which is made up from several chromosomes (here around 60 chromosomes) and genes (binary string) which carrying the different values for different network's parameters. During the mentioned process, called generations, by using those operators the chromosome will go forward and the best chromosome which has high fitness will go to next generation. This protracted process until reach the best solution which is the network's parameters would be continued.

RESULTS AND DISCUSSION

In this paper, the ANN-GA modeling is illustrated using a case study. This case study is an iron-ore deposit, where the dataset (Table 1) is partially taken from Clark (1979). The simulated deposit is a low grade sedimentary iron ore, with an overall average of about 35% Fe, a standard deviation of 5% Fe. According to

pervious study on this dataset, this area has been simulated by three different methods which will be discussed bellow. Also, Kriging because of the insufficient amount of data for the construction of a semivariogram was not applied. The testing stag because the lack of data in order to dividing them into different section was fairly different. In this study, first all of the available data except one of them has been used for training and this process has been applied for all of the data. It means that in several steps, testing and training has been performed. This process has been repeated for all of the points.

Neural Network Based on Trial and Error Method:

In this study, the inputs were coordinated data and the output was the grade value. In the local algorithms, the only way to determine the optimal structure would be trial and error method. According to the results, the network optimal structure was found to be 2-6-1. This structure has the lowest MSE (Mean Square Error) and R value (correlation coefficient). It is clear that the best and the desired values for the mentioned parameters are 0, and 1 respectively. For more than a single layer, ANNs were also used throughout this stage, but the results were unsatisfied. We have tested this network by using a set of data. These results are shown the Fig. 3, which represent the performance of the network for testing data.

Fuzzy Logic:

Pham (1997) introduced the grade estimation using fuzzy c-means and a fuzzy inference system (FIS). He also used the same data which used in this study. In his paper, based on a collection of cluster centers obtained from fuzzy c-means, a fuzzy rule base and fuzzy domains were established to compute grades at these cluster centers Pham (1997). The results of this application are shown in Table 2 and Fig. 4 in which the MSE and R values of different clusters and the best results are presented respectively. Also, Tutmez and others (Tutmez *et al.,* 2007) applied a new approach which used fuzzy modeling algorithms and estimates the reserve parameters based on spatial variability. They proposed a new cluster index approach in which used a new spatial measure function (point semimadogram). The results from their study are also shown in Table 2. According to Table 2, the second applied method presented a better performance than the ordinary fuzzy method. The difference is because of spatial variability and a new cluster validity index for estimating mineral reserves.

Genetic Algorithm for Designing:

In this section, GA was applied to determine the final structure of network and the other parameters. By applying the GA, the optimal structure and best parameters were obtained, and it means that by using these parameters, the network performance is more accreted. Therefore, the results that obtained from the GA, shows that the input and output number are constant (2 and 1 respectively) but the number of hidden layer founded to be 2. These parameters show that the ANNs, using two hidden layers, illustrate the best performance. Obtained architecture is 2-24-17-1. These obtained parameters, which have been merged by minimizing the fitness function "based on MSE value", was applied in the final network structure. Since the goal of GA is to minimize the network MSE.

The maximum epoch was set on 10000, due to avoid overtraining (over fitting) and also to prevent the continuing processing. It is because, in certain epoch, the network reaches a fixed MSE and at this point it is better to stop its activity. In this application, after 5000 epochs the network reached to fixed MSE and no more improvements happens. As mentioned before, the fitness function here is MSE value, which it means, after using a chromosome; the fitness function will calculate all of the generations and the best results “contained the minimum MSE" would be selected. This selection would be the best selected network structure. The number of generations was 100, which in this case study, after 65^{th} generations, MSE was reached to the lowest possible error values and the network designing was based on that obtained structures (Fig. 5). In this figure it is obvious that, after which generation it was reached to minimum MSE.

As discussed earlier, genetic algorithm by using its operators tries to find the best solution. Also, this will be continued until it reaches the lowest possible MSE during the network training. In the other hand, the MSE converge is a criteria which used to evaluate the genetic algorithm performance to finding the network's parameters. After it was insured that the optimal structure has been founded, in this step it would be permitted to use the testing data for evaluate the network performance in the testing data set (Fig. 6). Table 3 shows the results of testing data for the new obtained structure by genetic algorithm.

The networks have desired performance and when it uses the genetic algorithm it gets the improved results. The parameter properties that have been obtained by genetic algorithm are: 24 neurons in the first and 17 neurons in the second hidden layers, with learning rates of 0.48 and 0.47 in the first and second layer

respectively. Also, the optimum momentums were 0.038 and 0.05 in the first layer and second layer respectively. This network structure was detected by 85th generation that its responsible was the 25 network chromosomes.

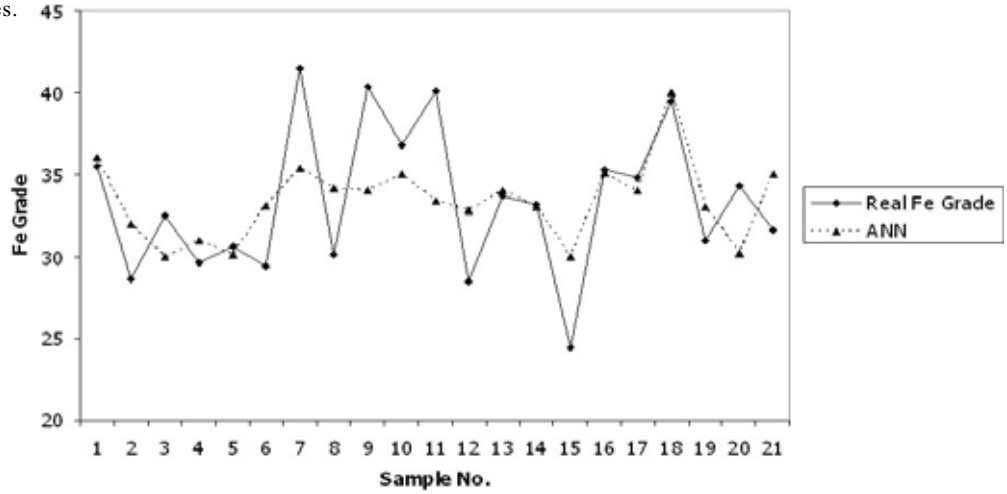

Fig. 3: Comparison between real and estimated grade by applying ANN.

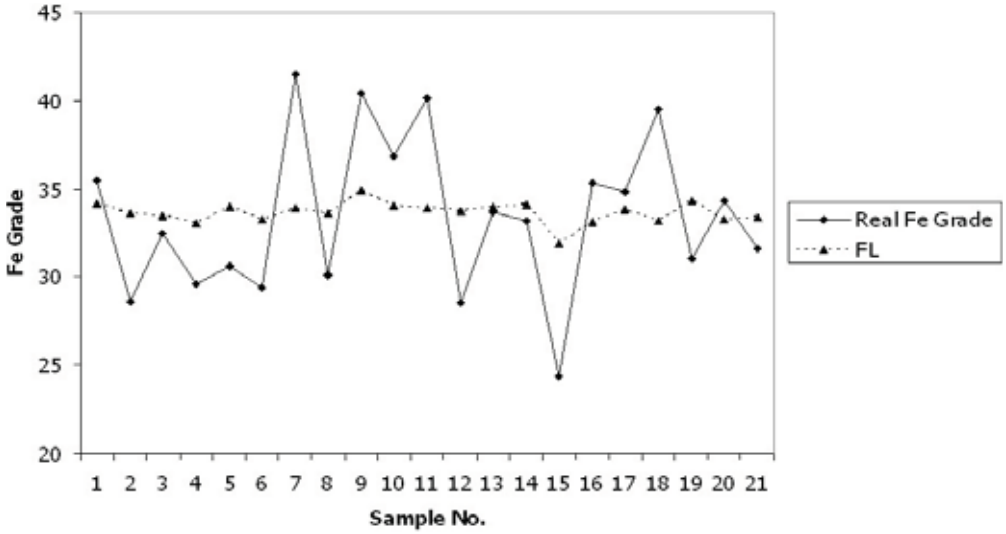

Fig. 4: Comparison between real and estimated grade by applying different FL methods.

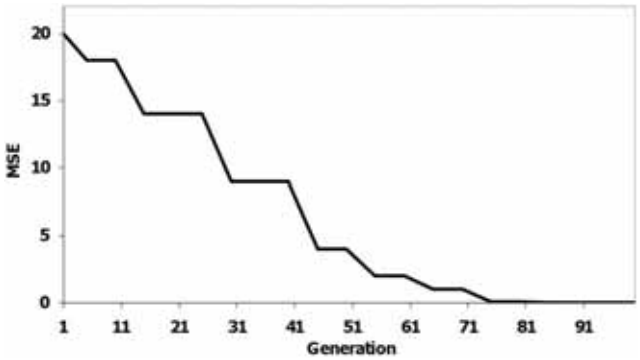

Fig. 5: Best fitness (MSE) versus generation for grade estimation in which best fitness in each of the different generation is shown.

The ANN-GA technique is expected to provide a significant improvement when the unseen (new or test) data come from new conditions. Since, most of the mining and geological activities are mixed with some fuzzy condition, the performance of the ANN-GA could be an excellent chose compared to the other methods such as ANN and FL. The comparison between these methods is shown graphically in Fig. 7 in which three different methods as explained, are compared (Fig. 7). According to the results, the new approach could to predict the grade accurately more than the other methods. High accuracy and low error is because of ANN and GA combination which make a powerful tool for grade estimation.

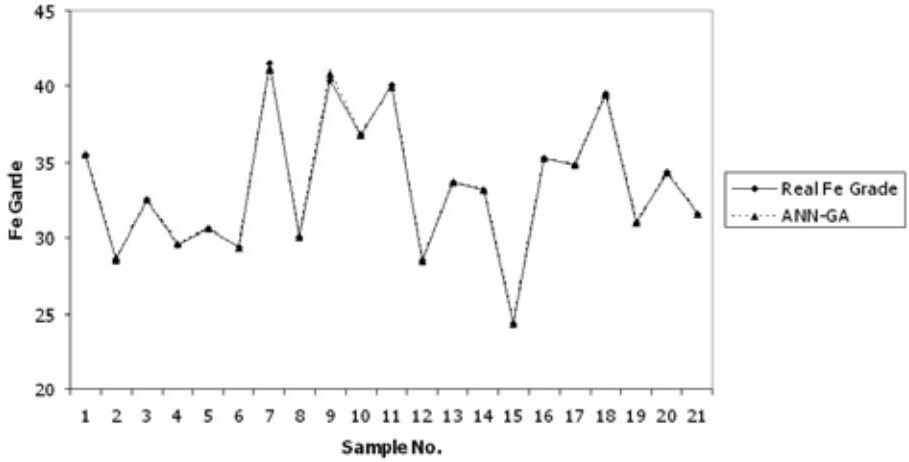

Fig. 6: Comparison between real and estimated grade by applying ANN-GA.

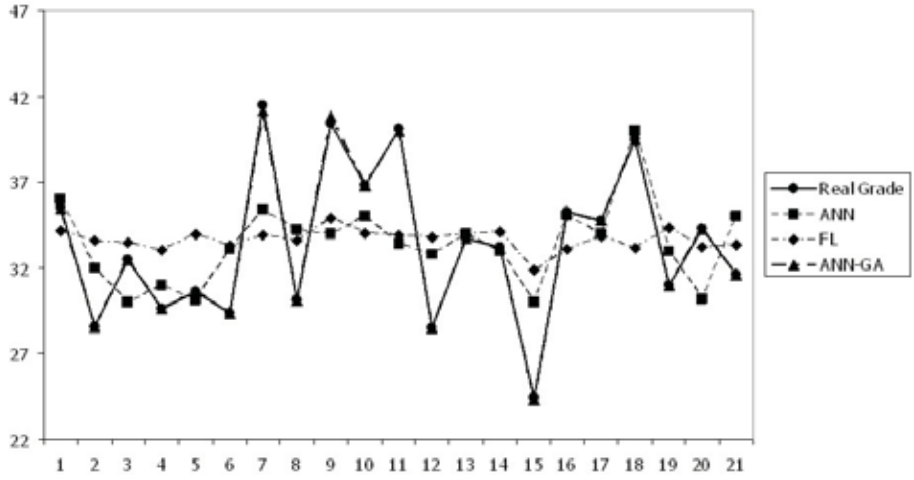

Fig. 7: Comparison between ANN, FL and ANN-GA performance for grade estimation.

Table 1: Initial dataset from Iron ore deposit (Clark, 1979).

Sample No	Easting	Northing	Fe %	Sample No	Easting	Northing	Fe %
1	10	40	35.5	12	15	135	28.6
2	55	145	29.4	13	125	20	41.5
3	175	50	36.8	14	260	115	33.2
4	235	15	33.7	15	365	60	34.3
5	285	110	35.3	16	345	115	31
6	20	105	32.5	17	25	155	39.6
7	50	40	30.6	18	155	15	10.4
8	145	125	30.1	19	220	90	28.5
9	205	0	40.1	20	265	65	24.4
10	390	65	31.6	21	325	105	39.5
11	10	40	35.5	22	15	135	28.6

Table 2: Performance comparison between Pham (1997) and Tutmez (2007) for grade estimation by applying FL.

	Pham (1997)	Tutmez and others (2007)
MSE	17.0569	1.3924
R	0.5138	0.94

Table 3: Performance of ANN-GA model for testing stage.

Statistics	Value
MSE	0.0001764
R	0.9995
Min absolute error	1.98455E-07
Max absolute error	0.186786

Conclusions:

In this study, the application of genetic algorithm in artificial neural network for grade estimation was investigated. It was used the genetic algorithm as a method to minimize the errors and to achieve the optimal structure. For this purpose the network parameters such a number of neurons in hidden layers, learning rate and the momentum were optimized. In the common methods, like trial and error, user should spend a long time to find the optimal structure and in most of the situation it is dependent of the input data and their spatial relations. But in addition, this method needs a long time and the results are unreliable. As it is shown, the obtained results indicate a fine performance of genetic algorithm to optimize the network and improve it to predict the parameters accurately. Genetic algorithm to optimize the networks for grade estimation was applied for a well-known case study in which the network performances to grade estimation were increased. Most of this optimization operation is to escape the trapping in local and reach to global minima. By this method, it is possible to optimize the network architecture for grade estimation. Solving these problems and their spatial relation is very difficult. The results also indicate that this method could improve the network performance in the case study and the mean square error (MSE) and R values decreased and increased respectively. The proposed methodology was also compared with other prediction methods like artificial neural network (ANN) and fuzzy logic (FL) which shows significant improvement compared to both prediction methods. The aim of this comparison is to show the performance and the advantages of the applied method results compared to the other ones. According to obtained results, the combined neural network with genetic algorithm could be a powerful tool for mining engineers and geologists to apply that in the most earth science problems to achieve a better result.

REFERENCES

Cargill, S.M., R.F. Meyer, D.D. Picklyk and F. Urquidi, 1977. Summary of resource assessment methods resulting from the International Geological Correlation Program Project 98: Mathematical Geology, 9(3): 211-220.

Carras, S.N., 1987. Concepts for estimating ore reserves and a comparative view of resource estimation methods: Resources and Reserves Symp. (AuslMM Sydney Branch, Australia), pp: 73-80

Clark, I., 1979. Practical Geostatistics. Applied Science Publishers, London, pp: 281.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function, Math. of Control, Signal and System, 2(4): 303.

David, M., 1977., Geostatistical ore reserve estimation: Elsevier Scientific Publ Co, Amsterdam, pp: 364.

Funahashi, K.I., 1989. On the approximate realization of continuous mappings by neural networks. Neural Networks, 2(3): 193.

Hartman, E.J., J.D. Keeler and J.M. Kowalski, 1990. Layered neural networks with Gaussian hidden units as universal approximations. Neural Computation, 2(2): 210.

Hornik, K., M. Stinchcombe and White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks, 2(5): 359.

Ke, J., 2002. Neural network modeling of placer ore grade spatial variability: Unpublished Doctoral Dissertation, University of Alaska Fairbanks, pp: 251.

Koike, K. and S. Matsuda, 2003. Characterizing content distributions of impurities in A limestone mine using a feed forward neural network: Natural Resources Research, 12(3): 209-223.

Koike, K., S. Matsuda and B. Gu, 2001. Evaluation of interpolation accuracy of neural kriging with application to temperature-distribution analysis, Mathematical Geology, 33(4): 421-448.

Michalewicz, Z., G.A. Vignauax and M. Hobbs, 1991. A non-standard genetic algorithm for the nonlinear transportation problem, ORSA Journal of Computing, 3(4): 307-316.

Pham, T.D., 1997. Grade estimation using fuzzy-set algorithms. Mathematical Geology, 29(2): 291-305.

Samanta, B., S. Bandopadhyay and R. Ganguli, 2004. Data segmentation and genetic algorithms for sparse data division in Nome placer gold grade estimation using neural network and geostatistics: Mining Exploration Geology, 11(1-4): 69-76.

Shang, Y. and B.W. Wah, 1996. Global optimization for neural network training: IEEE Computation, 29(3): 45-54.

Tutmez, B., E.A. Tercan and U. Kaymak, 2007. Fuzzy modeling for reserve estimation based on spatial variability. Mathematical Geology, 39(1): 87-111.

Wu, X. and Zhou Y., 1993. Reserve estimation using neural network techniques. Computer & Geosciences 19(4): 567-575.

Yama, B.R. and G.T. Lineberry, 1999. Artificial neural network application for a predictive task in Mining: Mining Engineering, 51(2): 59-64.