Journal of Engineering Science and Technology
Vol. 13, No. 3 (2018) 755 - 765

SEISMIC ATTRIBUTES SELECTION AND POROSITY PREDICTION USING MODIFIED ARTIFICIAL IMMUNE NETWORK ALGORITHM

SAAD ZAID SAAD*, MUNA HADI SALEH

Electrical Engineering Department, University of Baghdad, Baghdad, Iraq
*Corresponding Author: saadalshamary99@gmail.com

Abstract

In petroleum field, well logs such as porosity, permeability, resistivity, etc. play a major role in hydrocarbon reservoir studying, but these logs require expensive and slow measurement methods so that seismic attributes can be used as predictors to predict the logs depending on the already measured logs. This research presents a proposed modification for Artificial Immune Network (aiNet) algorithm to achieve better performance. The modified algorithm was used to select the best combination of attributes that leads to the best possible prediction by maximizing Correlation Feature Selection (CFS) objective function that leads to low correlated attributes between each other and high correlated with the porosity. The prediction is accomplished by using Artificial Neural Network learned by the modified aiNet.

Keywords: Artificial immune system, Correlation feature selection, Feed-forward artificial neural network, Immune network theory, Optimization.

1. Introduction

In petroleum industry, porosity log is the best parameter because it is accurate, fast, and require easy measurements as compared to the other logs, but studying core properties depending only on the usual methods is time consuming and expensive method [1]. To decrease the time and costs, feed forward Artificial Neural Network (ANN) can be used to predict the porosity using the seismic attributes as predictors. Many methods of learning for ANN are used and the most popular are Back-Propagation ANN (BP-ANN) [1].

ANN is a new technique of porosity prediction and it provides a nonlinear input/output mapping, since seismic attributes and porosity have a complex nonlinear relation it became necessary to use ANN for porosity prediction [2].

Nomenclatures

A	Original clone
$A\sim$	Mutated clone
B	Mutation rate
b_n^k	The bias of the *n*th neuron of the *k*th layer
d	Euclidian distance between two cells
f	Fitness vector
$f\sim$	Normalized fitness vector
r	Correlation coefficient
W_{ij}^k	The weight between the *i*th neuron of the *k*th layer and the *j*th neuron of the *(k+1)*th layer

Abbreviations

AIS	Artificial Immune System
ANN	Artificial Neural Network
aiNet	Artificial Immune Network
BP-ANN	Back-Propagation Artificial Neural Network
CFS	Correlation Feature Selection
GA	Genetic Algorithm
MSE	Mean Squared Error
PCA	Principle Component analysis

The first an important issue is the selection of seismic attributes as an input data to learn and test ANN. Principle Component Analysis (PCA) is widely used in seismic attributes reduction, but it is a general unsupervised method the final reduced data are can be very low correlated with the porosity [3]. A supervised method called Correlation Feature Selection (CFS) can be used to select the attributes that achieve high porosity prediction by selecting a combination of attributes that high correlated with the porosity and low correlated with each other [4].

Artificial Immune System (AIS) and ANN both are biologically inspired, and AIS is inspired by the biological immune system of organisms [5]. Immune system has a special mechanism of body defending against foreign bodies, and immune system is self-adapted because there is no central unit that control it [5]. AIS can be used in different application such as pattern recognition, machine learning, and optimization. In optimization, AIS is a new technique as compared to other biologically-inspired methods [6]. Immune system has two types of cells that responsible for defending the body, and these cells are T-cell and B-cell, each cell has a different mechanism to recognize and kill the invader, but each cell has special bodies on its surface called antibodies, these antibodies are responsible of recognizing invader's (antigen's) Epitopes that located on the surface of the antigen [7]. Immune system has three main concepts, clonal expansion, hypermutation, and suppression. These concepts lead to develop three main models of AIS and they are clonal selection principle, immune network theory, and negative selection [7].

2. Literature Review

Wong et al. [8] introduced two approaches to predict the porosity using back propagation neural network. The methods are genetic and nongenetic and the performances of them are critically evaluated. They proposed a new methodical mechanism to optimize the neural network configuration depending on weight visualization curve, so artificial neural network training time was highly reduced. In the example problem, the genetic approach provides excellent porosity prediction to that based on a nongenetic approach.

De Castro and Timmis [9] presented the adaptation of immune network to be compatible with optimization requirements. Immune network was originally proposed to carry out data compression and clustering, and in this paper the model was adapted to find the optimum solution of multimodal function. The algorithm was described, theoretically and compared with the clonal selection algorithm.

Dorrington and Link [2] introduced a new method of best attribute subset selection depending on the cross validation correlation of the neural network output. The best subset was selected to achieve maximum cross validation correlation and minimum number of attributes using Genetic Algorithm (GA). Binary coded GA was used and each gene represented the state of the attribute (i.e., selected or not).

3. Immune Network Model

Immune network theory is based on all the three concepts (clonal expansion, hypermutation, and suppression), and the main feature of this model is immune system cells can recognize and eliminate each other in a process called suppression so that the number of cells (population) is dynamic. Clonal expansion is a process of creating new clones (cells) of each cell and that increases the population so that suppression process is needed to control the number of cells [9]. Suppression is simply the process of eliminating similar cells and preserving the best cell. Hypermutation is a dynamic mutation depends on the affinity of the cell to the invader, and the cells with high affinity have a low mutation rate and vice versa.

3.1. Artificial immune network algorithm

Artificial Immune Network (aiNet) algorithm was proposed [9] as an optimization algorithm. The following terms is needed to develop an optimization version:

- Cell: Each cell is real coded and represents an expected solution.
- Fitness: Fitness of cell represents the associated value of the objective function.
- Affinity: Affinity is the measure of similarity between two cells and it can be measured by finding the Euclidian distance between the two cells.

The procedure of aiNet algorithm is shown below [9]:

1. Initial population: Generate random P cells and calculate the fitness of each one.
2. While (average error > t_1)

2.1. Clonal expansion: Make N_c copies of each cell.

2.2. Normalize the fitness of the cells to [0,1] range to create $f^{\sim}$.

2.3. Hypermutation: Apply hypermutation to each cell by applying the following two equations:

$$A_i^{\sim} = A_i + \alpha * \mu \quad (1)$$

$$\alpha = \frac{1}{B} * e^{-f^{\sim}(i)} \quad (2)$$

where μ a random number between 0 and 1, B is a parameter controls the mutation rate, A_i is the original i'th clone, $A_i^{\sim}$ is the i'th mutated clone, and $f^{\sim}(i)$ is the normalized fitness of i'th clone.

2.4. Calculate the fitness of the new mutated population.

2.5. Calculate the average error between the fitness of the present iteration and the fitness of the previous iteration.

3. Suppression: Calculate the Euclidian distance (d) or the affinity between each two cells as shown in Eq. (3). If the Euclidian distance is smaller than t_2, remove the cell with lower fitness.

$$d(A_i, A_j) = \sqrt{\sum_{k=1}^{n} (A_i(k) - A_j(k))^2} \quad (3)$$

4. Replace the worst k cells by random cells to save the diversity.

Go to 2.1.

3.2. Modified artificial immune network algorithm

The disadvantages of aiNet optimization version that need to be modified are shown below:

- All population undergoes clonal expansion even bad antibodies (or cells).
- All cloned antibodies go through hypermutation process.
- Hypermutation process does not add diversity to the population, it only searches for the neighbor solutions because antibodies fitness is normalized to [0 1] range so that α maximum value is 1/B as shown in Eq. (2). Since μ is a random vector its maximum value is 1, the mutation in Eq. (1) has a maximum expectable value of 1/B and B is a large number greater than 1. If B takes values smaller than 1, the mutation rate will be high, and the problem is if the mutation is high and all antibodies' elements are undergo mutation, the mutation will be as a generation of new random antibodies so that no convergence to the solution will be existed.

The proposed modifications are as shown below:

- Another stage of mutation before the original while loop is added. The new mutation is different from the original mutation, the new mutation is a hypermutation but not all cell's elements undergo mutation, and the number of elements (variables) that go through hypermutation process is depend on the fitness of the cell. For the best cell, only one element goes through hyper

mutation by replacing it with new random number. The number of selected elements changes linearly with fitness value.

- Only fixed rate of best antibodies are selected to be cloned in clonal expansion process.
- Only fixed rate of best antibodies are selected to be mutated in hyper mutation processes (stage 1and stage 2).

The block diagram of the modified aiNet is shown in Fig. 1.

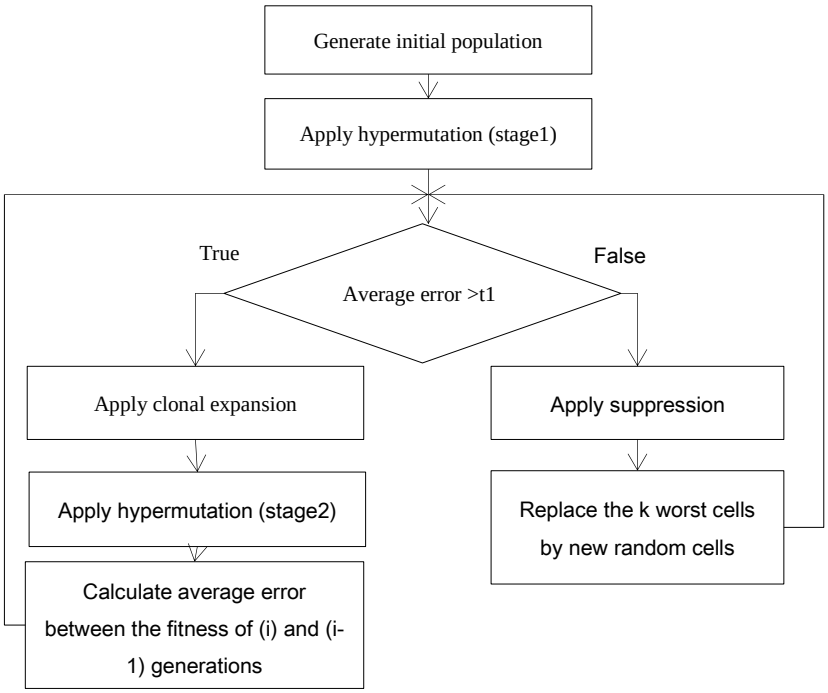

Fig. 1. Modified aiNet block diagram.

4. Correlation-Based Attributes Selection

In machine learning, data reduction for classification issues can be accomplished depending on correlation principles [4]. Correlation Feature Selection (CFS) is a supervised method it reduces the number of features (attributes) depending on specific criteria to give minimum redundancy in the selected features and maximum relation with the class [4].

In CFS method the criterion of selecting a good features subset is a good subset should be highly correlated with the class and the features should have low correlation between each other. CFS criterion is shown in Eq. (4) [4].

$$CFC = \frac{k * \overline{r_{ic}}}{\sqrt{k + k * (k-1) * \overline{r_{ij}}}} \tag{4}$$

where k is the number of selected features, $\overline{r_{ic}}$ is the average correlation between the selected features and the class, and $\overline{r_{ij}}$ is the average correlation between the selected features. By converting the average in Eq. (4) to summation, CFS criterion will be as shown in Eq. (5).

$$CFS = \frac{\sum_{i=1}^{k} r_{ic}}{\sqrt{k + 2 * \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} r_{ij}}} \tag{5}$$

The correlation (r) can be evaluated using different methods, Pearson's equation of evaluating the correlation coefficient is used and it is as show in Eq. (6) [10].

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{6}$$

where n is the number of attribute's samples, x_i is i'th sample of attribute x, $\bar{x}$ is the average of attribute x, y_i is i'th sample of attribute y, and $\bar{y}$ is the average of attribute y.

Attributes subset selection is based on finding a subset achieves maximum CFS criterion shown in Eq. (5). Modified aiNet was used to maximize CFS criterion and finding the global optimum solution regardless of initial subset. Antibody length is the total number of attributes and it is binary coded, 1 means that the associated attribute is selected and 0 means that the associated attribute is not selected as shown in the follow example:

[0 1 0 1 1 1 0 0 0 0]

The above example is a binary coded antibody with length of 10 represents the total number of attributes, and the number of selected attributes (k) is 4 and they are indicated by 1. The objective function value (CFS) of the above antibody is founded by finding the correlation between each one of selected attributes and the class and finding the correlation of the 4 selected attributes then applying Eq. (5).

5. Modified aiNet Based ANN

AIS algorithms can be used to learn ANN by adjusting the weights and the biases of the network to achieve minimum mean squared error [11]. The antibody (or cell) is real coded and it represents the weights and the biases of the network so that the length of each antibody is equal to the number of weights and biases. The learning is performed by applying AIS optimization algorithms to adjust the antibodies to find the minimum mean squared error (maximum –ve mean squared error) [10]. The real coded antibody will be as shown below:

$$\left[W_{11}^{1} \ldots\ldots W_{ij}^{k} \ldots\ldots b_{1}^{2} \ldots . b_{n}^{k}\right]$$

where W_{ij}^{k} is the weight between the *i*th neuron of the *k*th layer and the *j*th neuron of the *(k+1)*th layer, and b_{n}^{k} is the bias of the *n*th neuron of the *k*th layer. Figure 2 shows the block diagram of AIS based ANN.

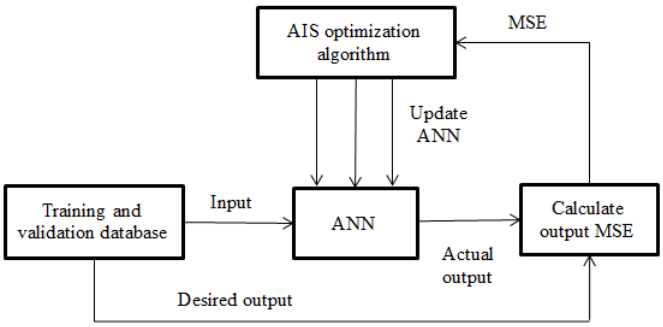

Fig. 2. AIS based ANN.

6. Simulation Results

Modified aiNet is applied to find the best subset of attributes by maximizing the objective function shown in Eq. (5). The algorithm was applied to 26 instantaneous attributes and Fig. 3 Shows the values of CFS objective function at each generation.

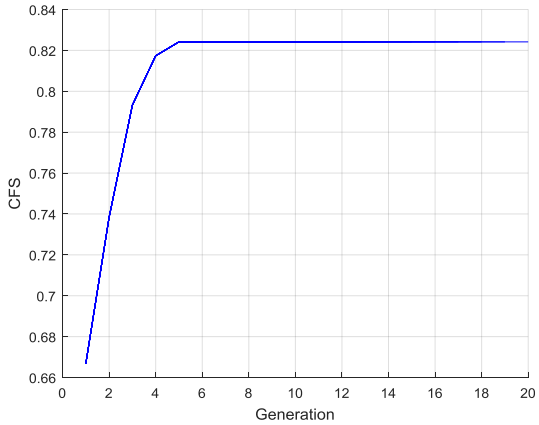

Fig. 3. Modified aiNet based CFS.

After 7 generations the best subset was founded with CFS of 0.8242, and the best subset consists of 4 attributes. Figure 4 shows the best four selected attributes and the porosity.

The selected attributes were used as an input to ANN learned by modified aiNet algorithm to predict the actual porosity. The settings of the used feed-forward ANN and the algorithm are as shown below:

- 4 input neurons.
- Single hidden layer with 5 neurons.
- Single output neuron.
- Each neuron has tangent sigmoid activation function.
- The objective function is mean squared error (MSE).

- Cells are real coded with a length of 31 variables (25 weights and 6 biases).
- Initial population size is 100 cells.
- ¼ of cells go through clonal expansion and the number of clones of each cell is 10 clones.
- ¼ of cells go through hyper mutation.
- The diversity rate is 20% of the population.
- Other parameters such as mutation decay, affinity threshold, and error change threshold can be adjusted to control the population change.

Each attribute has a length of 97 samples so that the data is divided into 50 samples for training, 20 samples for validation, and 27 samples for test. Multi-run simulation is used, and each run has 15 generation. ANN training and MSE minimization are shown in Fig. 5.

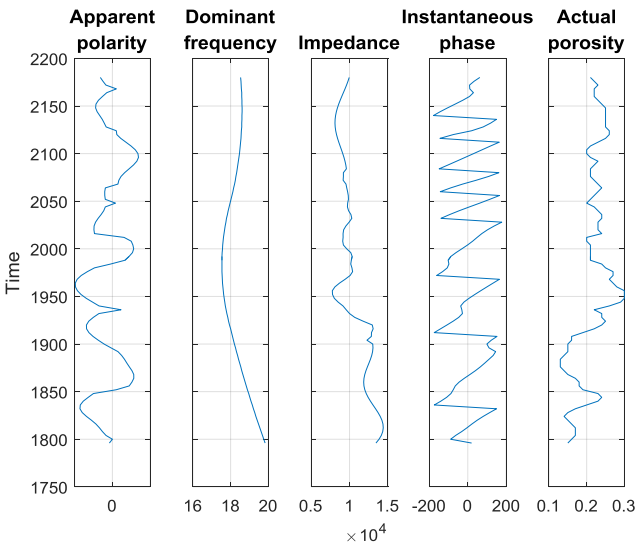

Fig. 4. Selected attributes and target porosity.

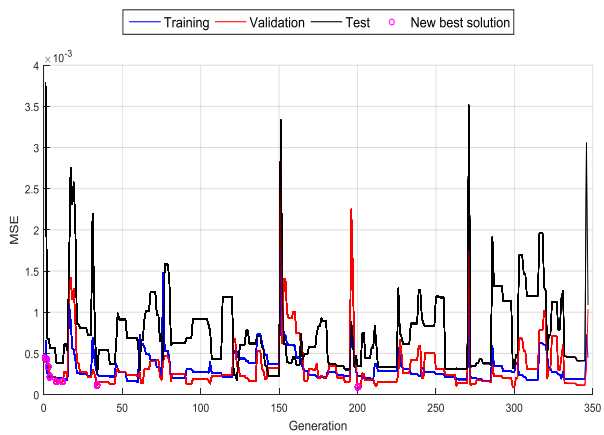

Fig. 5. ANN training using modified aiNET.

The best solution was founded in generation 203. Figure 6 shows the change of population size during the 350 generation, and Fig. 7 shows the values of the predicted and the actual porosity for the three parts of data.

The values of MSE and the correlation between the actual and predicated porosity are shown in Table 1.

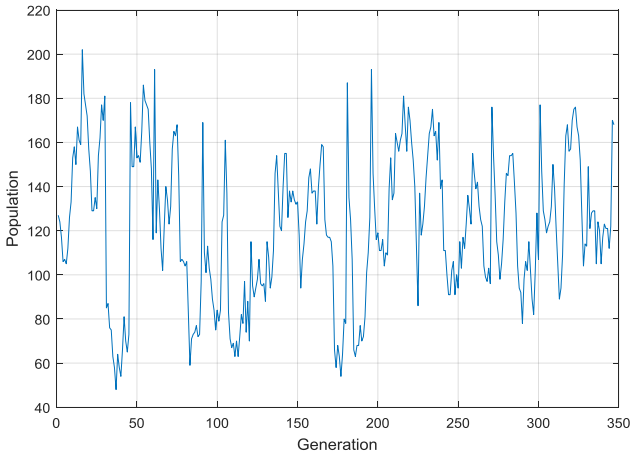

Fig. 6. Modified aiNet population value after each generation of ANN training.

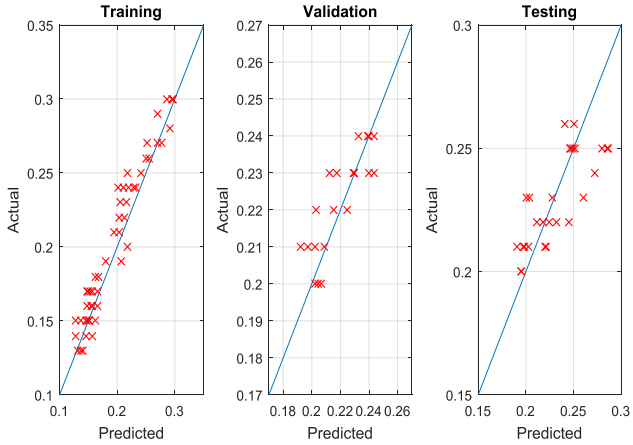

Fig. 7. The actual an the predicted porosity.

Table 1. Test model Specifications and test conditions.

	MSE	Correlation
Training	0.00022216	0.9699
Validation	0.00008853	0.8412
Test	0.00034579	0.7823

7. Discussion and Conclusion

By using random search optimization algorithms the problem of initial weights and biases is not completely solved because finding global solution for the learning data does not mean finding global solution for the testing or validation data. For example, there are two solutions both have the same leaning mean squared error, but the second solution has better testing or validation mean squared error, now the goal is to find the second solution, the optimization algorithm can find both solutions in equal chances, if it finds the first solution, the first solution will dominate the population so that there will be no chance to find the second solution. To solve this problem a refreshing for the population every some generations is needed, refreshing means changing the population and applying new run to give chance to the other solutions to dominate the population to give different choices of solutions. For this reason multiple runs are used to search for a solution with good validation MSE.

Modified aiNet showed that it is too fast and more accurate as compared to the original aiNet algorithm for more than one reason, and the first reason is the new mutation gives the algorithm more diversity and bigger steps toward the solution, the second reason is clonal expansion and hypermutation is limited only to the group of top cells of the population instead of applying these processes to the entire population. Modified aiNet does not solve the complexity of aiNet, and the complexity is how to find the corrects parameters and thresholds settings that control population growth (or shrinking) because the population change is sensitive to the all parameters so that the future scope of this research is to modify and enhance the algorithm to make it simpler and solve the complexity of population control.

References

1. Hamidi, H.; and Rafati, R. (2012). Prediction of oil reservoir porosity based on BP-ANN. *Proceedings of the International Conference on Innovation Management and Technology Research (ICIMTR).Malacca, Malaysia*, 241-246.
2. Dorrington, K.P.; and Link, C.A. (2004). Genetic-algorithm/neural-network approach to seismic attribute selection for well-log prediction. *Geophysics*, 69(1), 212-221.
3. Hongjie, L.; Bing, L.; Taoping, L.; Yong, L.; and Chunhe, W. (2014). Seismic attribute reduction method and its application. *Information Technology Journal*, 13(14), 2326-2014.
4. Hall, M.A. (1999). *Correlation-based feature selection for machine learning*. Ph.D. Thesis. The University of Waikato, Hillcrest, New Zealand.
5. Weiguo, X. (2010). The weather prediction method based on artificial immune system. *Proceedings of the International Forum on Information Technology and Applications (IFITA)*, 386-389.
6. Bernardino, H.S.; and Barbosa, H.J. (2009). Artificial immune systems for optimization. In *Nature-Inspired Algorithms for Optimisation*. New York: Springer, 389-411.

7. De Castro, L.N.; and Timmis, J. (2002). Artificial immune systems: A novel paradigm to pattern recognition. In: *Artificial Neural Networks in Pattern Recognition*, 1, 67-84.
8. Wong, P.M.; Gedeon, T.D.; and Taggart, I.J. (1995). An improved technique in porosity prediction: a neural network approach. *IEEE Transactions on Geoscience and Remote Sensing*, 33(4), 971-980.
9. De Castro, L.N.; and Timmis, J. (2002). An artificial immune network for multimodal function optimization. *Proceedings of the Congress in Evolutionary Computation* (*CEC'02*), 1, 699-704.
10. Dubey, V.K.; and Saxena, A.K. (2016). Hybrid classification model of correlation-based feature selection and support vector machine. *Proceedings of the IEEE International Conference on Current Trends in Advanced Computing* (*ICCTAC*), 1-6.
11. Coelho, P.H.G.; and Neto, L.B. (2013). Neural equalizer trained by the artificial immune system learning algorithm. *Proceedings of the International Joint Conference on Neural Networks* (*IJCNN*), 1-4.