

Comparison of Different Types of Fitness Functions to Choose the Appropriate Attributes for Porosity Prediction

Muna Hadi Saleh
Electrical Eng. Dep.
University of Baghdad
drmunah4@yahoo.com

Hadeel M. Tuama
Electrical Eng. Dep.
University of Baghdad
calmsky.93@gmail.com

Abstract

Porosity is one of the most important reservoir characteristics because it indicates to fluid collection. Several techniques used to get good porosity prediction, so, in this study we employed seismic attributes and well log data in a genetic algorithm to get the best porosity prediction. The study attempt to enhance the performance of genetic algorithm for attribute selection and therefore porosity prediction by applying genetic algorithm on different types of fitness functions like average mean square error fitness, average correlation coefficients fitness and performance index fitness. Also, used two methods to represent attributes in genetic algorithm. Different witnesses applied to choose the appropriate fitness function that gives high porosity prediction.

Keywords: Genetic algorithm, porosity prediction, correlation coefficient

1. Introduction

Porosity is the most difficult properties to define in subsurface reservoir characterization. Among the factors necessary for the determination of a reservoir, the most significant and at the same time the most complicated to compute is porosity. In the event of a natural calamity, airlines might not stick to their original flight schedules. The difficulty of assessing them comes from the fact that porosity may vary considerably over a reservoir volume, but can only be sampled at well locations, often using various technologies at various scales of observation.

Various methods may be resolved porosity, used core samples; depending on well-log data and mathematical models are some of these methods. The particular benefit are techniques of porosity predicting from transit time analyses that make use of interval velocities obtained from seismic traces [1]. Seismic attributes are the parts of the seismic data which are acquired by measurement, computation, and other methods from the seismic data. Seismic Attributes were introduced as a section of the seismic interpretation in early 1970's. Since then many

modern attributes were derived and computed. Most of these attributes are of commercial benefit and, use of numerous of the attributes, are yet to be understood by numerous interpreters and users [2]. Well observations supply good vertical resolution of geologic layers. Porosity valuation from seismic data is a very substantial tool because it permits the determination of porosity distribution even away from drilled wells, lets a best characterization of known reservoirs in their economic and technical aspects, and supply frequently more information than normal seismic processing in the search for new hydrocarbon fields [1].

Genetic algorithm (GA) is a general purpose search algorithm which use rules inspired by natural genetic populations to develop solutions to problems. Maintaining a population of chromosomes is the basic concept of genetic algorithm, candidate solutions to the problem acted by the chromosomes, which develop over time through a process of competition and controlled diversity. Fitness was related to each chromosome in the population, selection defined which chromosomes are used to form new ones in the competition process. Genetic operators such as crossover and mutation are used to generate new chromosomes. Genetic algorithms have had a large measure of success in search and optimization problems. Genetic algorithm ability to exploit the information accumulated about a first unknown search space in order to bias next searches into beneficial subspaces is the purpose for a large part of their success.

This is their key feature, especially in large, complicated, and poorly understood search spaces, where conventional search tools are, unsuitable offering a valid approach to problems requiring effective and efficient search techniques. Various geophysical problems have applied genetic algorithm to geophysical optimization problems like seismic inversion, multiple suppression and migration velocity analysis. Seismic attributes and well log data are applied in genetic algorithm attribute selection method, using the better genetic algorithm attribute combination based on the predicted porosity volume [3]. A. Hosseini, et al., 2011. Exploited porosity as an important factor in

reservoir so they employed seismic attributes that derived from seismic data in artificial intelligent for porosity prediction [4]. Marin Golub, implemented genetic algorithm in two representation binary and floating point representation, and compare between two representations [5].

2. Attributes Representation in Genetic Algorithm

Porosity is one of important properties of reservoir, it is indicating to fluids saturation, so it constitutes an important factor to deduce the presence of oil. Employed seismic attributes and well log data in the genetic algorithm give a good prediction of the porosity and thus on fluid zones. As shown Figure (1)

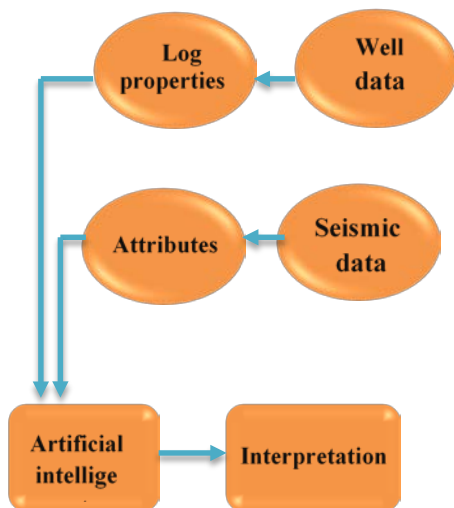


Figure 1: The data driven interpreters by artificial intelligent.

Genetic algorithm is a class of stochastic universal search technique based on biological evolution basics. Genetic algorithms represent parameters as an encoded binary variable and work with the binary variables to minimize the cost, while the other works with the continuous parameters themselves to minimize the cost. In this paper, we proposed two methods to represent chromosomes: The first representation method from our previous paper [6] that explained by the genetic algorithms attribute selection technique needs an initial set of attribute combinations. Chromosome named to each attribute combination and population represent the entire set of attribute combinations. Code of zeros and ones are described each chromosome, with selected attribute represented by one and attribute not selected representing by one. Gene referred to each contained of a large number of parameters. For example, given p total number of attributes, the number of possible combinations of seismic

attributes to be used is 2^p-1 , we define a matrix as a string of 0s and 1s with length of total number of combinations that each one wants to examine via genetic algorithm.

In case of seismic attributes, if we have p different attributes, each one contains h parameters, so a chromosome has p columns (genes) and h row (parameters). The concept of the population that contain 2^p-1 combinations help us to clearly indicate any particular combination of attribute. After representing attributes by binary code each 1 string replaced by the values of attribute that return to the same column, then apply the procedure of genetic algorithm with real code genetic algorithm's crossover and mutation, as illustrated in Figure 2.

Gene	G_1	G_2	G_{p-2}	G_{p-1}	G_p
Chromosome						
Chrom ₁	0	0	0	0	1
Chrom ₂	0	0	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Chrom _{2^{p-3}}	1	1	1	0	1
Chrom _{2^{p-2}}	1	1	1	1	0
Chrom _{2^{p-1}}	1	1	1	1	1
Has $h \times 1$ parameter						

Figure 2: Binary-Real code for attributes representation.

The second representation method described by each chromosome contained single attribute active, and the number of population equal to number of all attributes, for example given p total number of attributes, the number of possible combinations of seismic attributes to be used is p , we define a matrix as a string of 0s and 1s with length of p , where 1 represent active attribute and replaced by values of attribute. The case of seismic attributes if we have p different attributes, each one contains h parameters, so a chromosome has p columns (genes) and h row (parameters).

3. Fitness Functions Applied to Genetic Algorithm

A fitness function is a special kind of objective function that is used to concise how close a given solution is to achieving the group goals. Because of the effort involved in designing a practical fitness function, genetic algorithm cannot be considered to be a sluggish way of doing design work. Even though it is no longer the human designer, but the computer, that comes up with the last design, it is the human designer who has to design the fitness function. If this is designed

cursedly, the algorithm will either converge on an unsuitable solution, or will have hard converging at all. Fitness function must be work fast not only correlate closely with the designer's aim. Speed of execution is very important, as a typical genetic algorithm must be iterated many times in order to produce an applicable result for problem [7].

I. Average mean square error

The problem of this project is to find an optimal prediction to fit the observation data. We have to minimize the average sum square error between the predicted porosity by genetic algorithm and the observed porosity by well log data [8].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \tag{1}$$

Where:

- x_i : represented predicted values.
- y_i : represented actual values.
- n: represented total number of values.

II. Average correlation coefficient

To see whether the objective in studying the link behavior of two variables are related, there are many cases to do rather than use one to predict the value of the other. To strongly related two variables x and y , correlation coefficient R introduced as a measure of relationship between x and y [9].

$$R = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \tag{2}$$

Where:

- x: represented predicted values.
- y: represented actual values.
- n: represented total number of values.

III. Performance index

A performance index P_{m1} , derived from average correlation coefficients and the number of attributes is applied to chromosomes to determine the best subset for prediction in the population [10].

$$P_{m1} = (1 - R_{val}) + \frac{A_s}{A_n} \tag{3}$$

Where:

- R_{val} : is the average correlation coefficient.
- m: is a chromosome in the population.
- A_s : is the number of attributes in the subset.
- A_n : is the total number of attributes.

Genetic algorithm implemented in MATLAB to get high relationship between observed porosity and predicted porosity, therefore get high description of porosity locations in geological layers. The performance of prediction and selection evaluated by calculating average $RMSE$, average R and by minimize the total prediction error E^2 . Table 1

represents the result of binary-real code genetic algorithm applied to 10 seismic attributes with 2^P-1 chromosomes and Table 2 represent the result of second method representation chromosomes in genetic algorithm applied to 10 seismic attributes. The results of these Tables compared with the binary representation that illustrated in Table 3, but binary representation used for selection purpose only shown in Figures (3 to 14).

Table 1: results of 10 seismic attributes by first method of representation chromosomes

Types of fitness	Error fitness	Correlation fitness	Performance index fitness
Measurements	Roulette wheel selection	Roulette wheel selection	Roulette wheel selection
RMSE	0.1039	22.6110	13.2614
R	0.9724	0.7462	0.9689
E^2	0.0011	0.0027	0.0012
Best Cost	0.1	0.48192	0.13113
Time in Minutes	2.2919	16.3045	10.5538
No. of Iteration	341	1000	1000
Best Chromosome	Chrom.2	Chrom.1,6,3	Chrom. 2

Table 2: Results of 10 seismic attributes by second method of representation chromosomes

Types of fitness	Error fitness	Correlation fitness	Performance index fitness
Measurements	Roulette wheel selection	Roulette wheel selection	Roulette wheel selection
RMSE	1.8077	20.8881	23.7524
R	0.1364	0.8832	0.8844
E^2	0.0195	0.0044	0.0043
Best Cost	1.8077	0.91168	0.21562
Time in Minutes	0.1709	0.1962	0.0988
No. of Iteration	720	599	238
Best Chromosome	Chrom. 2	Chrom. 2	Chrom. 2

Table 3: Results of 10 seismic attributes by binary representation chromosomes.

Types of fitness	Error fitness	Correlation fitness	Performance index fitness
Measurements	Roulette wheel selection	Roulette wheel selection	Roulette wheel selection
RMSE	0.7181	26.3376	26.3376
R	0.0478	0.3781	0.3781
E^2	0.0199	0.0171	0.0171
Best Cost	0.7181	0.6220	0.7219
Time in Minutes	0.4677	0.5563	0.4436
No. of Iteration	41	41	41
Best Chromosome	Chrom.4	Chrom. 8	Chrom. 8

From Tables 1, 2 and 3 noticed that error fitness has less RMSE compared with correlation and performance index fitnesses of three representation methods. Error and performance index fitness have higher R in first representation method while in other two representation method performance index is higher than other two fitnesses. So, performance index function give good prediction for porosity. Representation methods return different number of chromosomes according to type of fitness it's used.

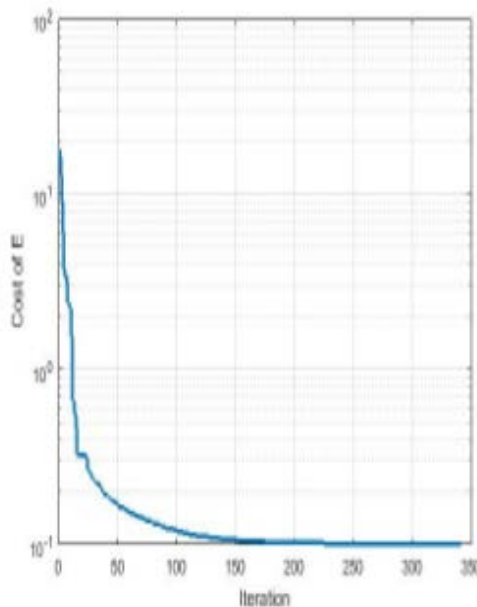


Figure 3 : Roulette wheel selection applied to error fitness with first method of representation

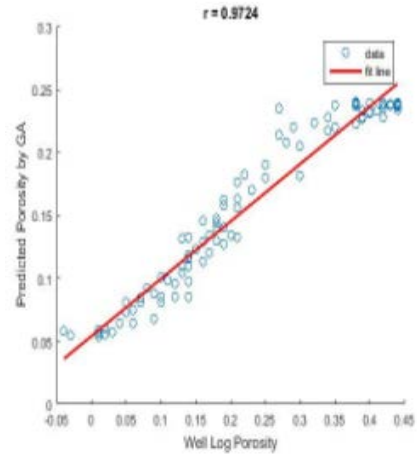


Figure 4: Cross plot showing the correlation coefficient between actual and predicted porosity by roulette wheel selection method

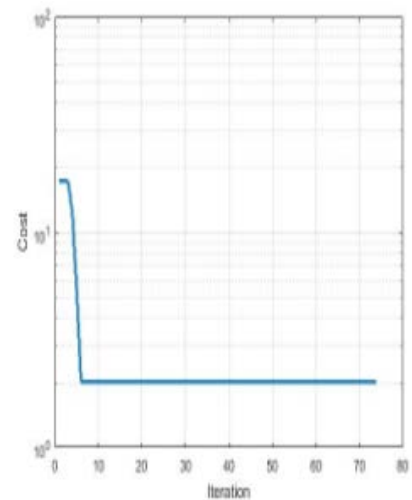


Figure 5 : Roulette wheel selection applied to error fitness with second method of representation

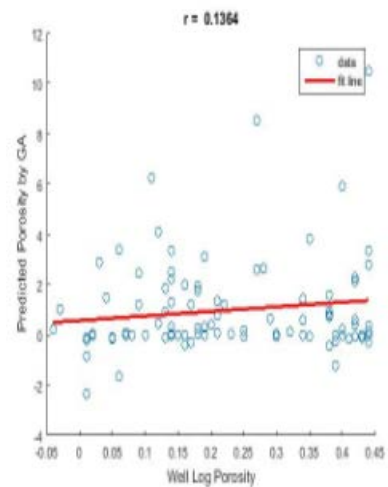


Figure 6: Cross plot showing the correlation coefficient between actual and predicted porosity by roulette wheel selection, error fitness and second method of representation

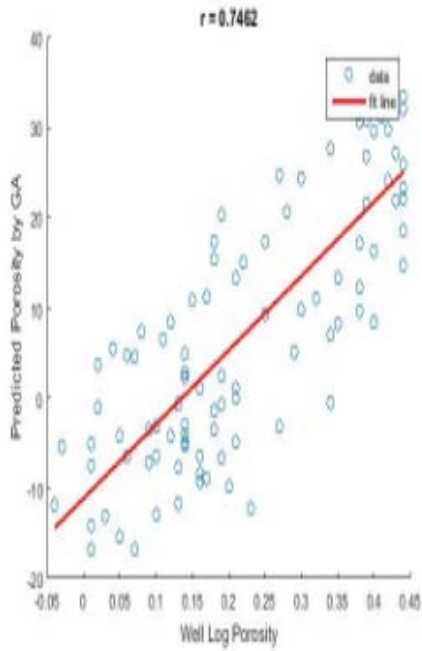


Figure 7 : Roulette wheel selection applied to correlation fitness with first method of representation

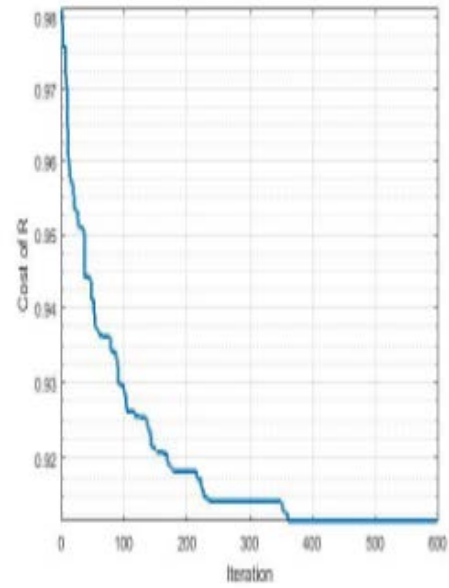


Figure 9 : Roulette wheel selection applied to correlation fitness with first method of representation

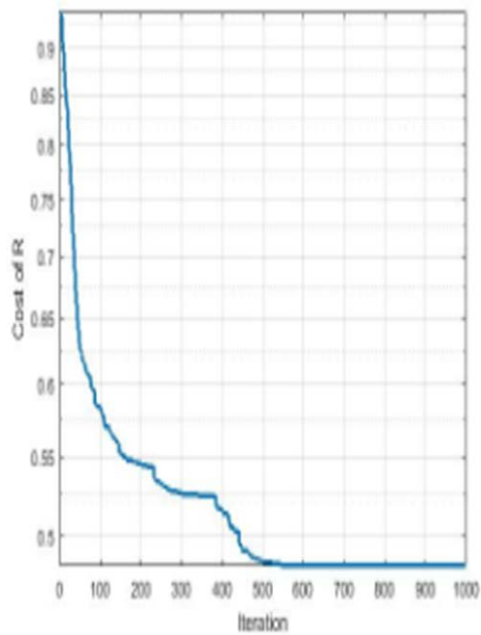


Figure 8: Cross plot showing the correlation coefficient between actual and predicted porosity by roulette wheel selection, correlation fitness and second method of representation

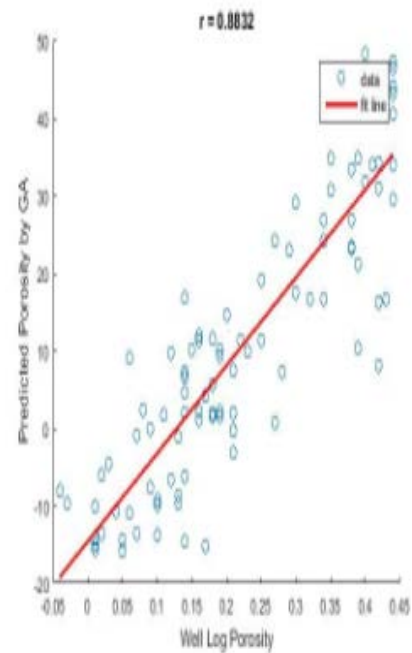


Figure 10: Cross plot showing the correlation coefficient between actual and predicted porosity by roulette wheel selection, correlation fitness and second method of representation

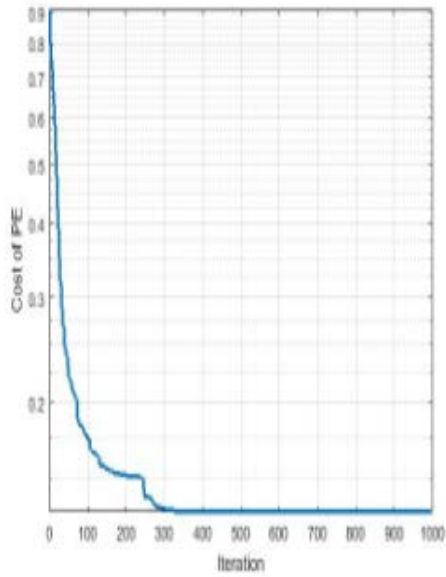


Figure 11 : Roulette wheel selection applied to performance index fitness with first method of representation

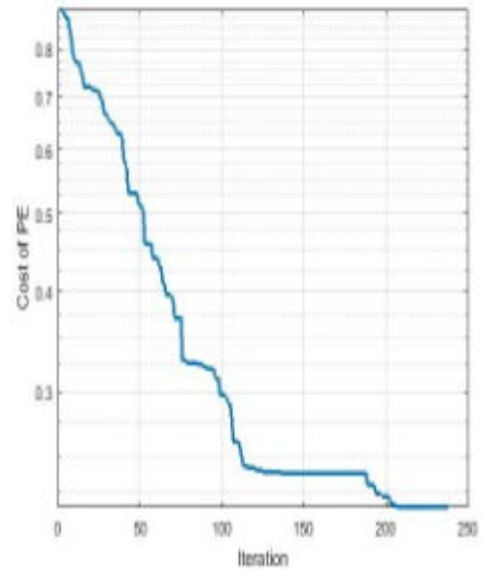


Figure 13 : Roulette wheel selection applied to correlation fitness with first method of representation

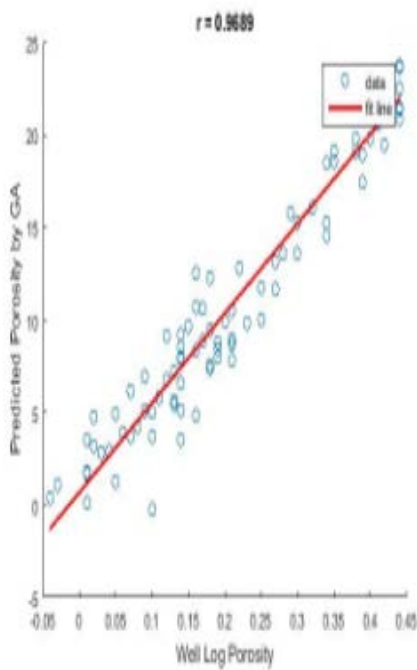


Figure 12: Cross plot showing the correlation coefficient between actual and predicted porosity by roulette wheel selection, performance index fitness and second method of representation

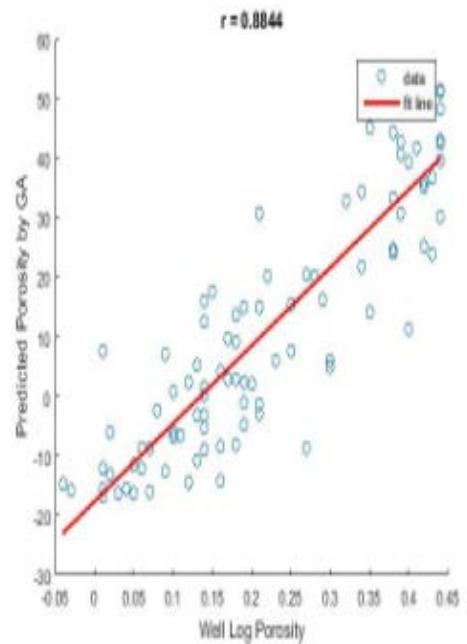


Figure 14: Cross plot showing the correlation coefficient between actual and predicted porosity by roulette wheel selection, performance index fitness and second method of representation

From Figures (3,4,5 and 6) noticed that the performance of error fitness with first method is best than second method, also has high fitting and correlation. Figures (7,8,9 and 10) show and correlation with second method while Figures (11,12,13 and 14) show that performance index has performance and high prediction with first method of representation.

5. Conclusions

From table one noticed that error fitness and performance index fitness give high correlation coefficient which mean high prediction between observed porosity and calculated porosity by genetic algorithm, but error measured by error fitness has low values in error fitness rather than performance index fitness. Also, error fitness has less number of iteration to reach to the best solution .so error fitness is appropriate fitness function for binary-real code representation. From table two noticed that correlation fitness and performance index fitness give high correlation coefficient which mean high prediction between observed porosity and calculated porosity by genetic algorithm, but they have high error values compared with error fitness. Also, this method of representation faster than first method of representation but it is less accuracy.

6. Acknowledgments

We would like to thank the Iraqi expert Mr. Jameel Radhi Jaafer to provide us with Necessary seismic information and for cooperating with us.

References

- [1] A. Adekanle and P. Enikanselu, "Porosity prediction from seismic inversion properties over 'XLD' field, Niger Delta," Am. J. Sci. Ind. Res., vol. 4, no. 1, pp. 31–35, 2013.
- [2] Subrahmanyam and P. Rao, H, 2008, "Seismic Attributes- A Review", International Conference and Exposition on Petroleum Geophysics, P. 398.
- [3] F. Herrera, M. Lozano, and J. L. Verdegay, "Tackling Real-Coded Genetic Algorithms: Operators and Tools for Behavioural Analysis," Artif. Intell. Rev., vol. 12, no. 4, pp. 265–319, 1998.
- [4] A. Hosseini, M. Ziaii, A. K. Rouhani, A. Roshandel, R. Gholami, and J. Hanachi, "Artificial Intelligence for prediction of porosity from Seismic Attributes : Case study in the Persian Gulf," pp. 168–174, 2011.
- [5] M. Golub, "an implementation of binary and floating point chromosome representation in Genetic algorithm", 2014.
- [6] Muna. H. Saleh, and H. M Tuama, " Novel Multi - Gen Multi Parameter Genetic Algorithm Representation for Attributes Selection and Porosity Prediction ", International Journal of Computer Applications, v. 141, no. 4, P. 34 - 39, 2016.
- [7] L. H. Randy and , E. H. Sue," practical Genetic algorithms", Wiley, 2004.
- [8] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature," Geosci. Model Dev., vol. 7, no. 3, pp. 1247–1250, 2014.
- [9] D. Shen and Z. Lu, "Computation of Correlation Coefficient and Its Confidence Interval," Sugi 31, pp. 1–6, 2006.
- [10] K. P. Dorrington and C. a. Link, "Genetic-algorithm/neural-network approach to seismic attribute selection for well-log prediction," Geophysics, vol. 69, no. 1, p. 212, 2004.

مقارنة بين عدة انواع من دوال الهدف لاختيار الخواص المناسبة للتنبؤ بالمسامية

هديل محمد طعمة
قسم الهندسة الكهربائية
جامعة بغداد

أ.م.د. منى هادي صالح
قسم الهندسة الكهربائية
جامعة بغداد

الخلاصة

المسامية هي واحد من اهم خواص الخزان النفطي لانها تعطي استدلال عن تجمع السوائل. هناك عدة تقنيات تستخدم للحصول على تنبؤ جيد للمسامية، في هذه الدراسة تم توظيف الخواص الزلزالية الصناعية ومعلومات تسجيلات الابار في الخوارزمية الوراثية للحصول على افضل تنبؤ للمسامية. في هذه الدراسة نحاول تحسين اداء الخوارزمية الوراثية لاختيار الخواص وبالتالي اعطاء تنبؤ جيد للمسامية عن طريق تطبيق عدة انواع من دوال الهدف مثل average mean square error, average correlation coefficients, performance index وكذلك استخدام طريقتين لتمثيل الخواص داخل الخوارزمية الوراثية، عدة معايير تطبق لاختيار دالة الهدف المناسبة التي تعطي تنبؤ عال للمسامية.