

Novel Multi-Gen Multi Parameter Genetic Algorithm Representation for Attributes Selection and Porosity Prediction

Muna Hadi Saleh, PhD
Assistant Professor
Department of Electrical Engineering/
College of Engineering / University of Baghdad

Hadeel Mohammed Tuama
Department of Electrical Engineering/
College of Engineering/ University of Baghdad

ABSTRACT

Many applications require a careful selection of attributes or features from a much larger set of data. This attributes selection problem need to optimized. In order to tackle this problem this paper proposes a binary-real code multi-gen multi-parameter genetic algorithm for attributes selection from large seismic data and prediction of effective porosity. Genetic Algorithm (GA) uses three selection methods for this purpose, mean square error and correlation coefficient are two witness criteria to choose the best subset of attributes that minimize the error and give high prediction of porosity.

Keywords

Genetic Algorithm, Multi-Gene, Multi-parameter, Attributes selection, Attributes prediction.

1. INTRODUCTION

One components of the seismic data are the seismic attributes which are obtained by many ways; measurement, computation, and others. Seismic Attributes were introduced as a section of the seismic interpretation in premature 1970's. Since then numerous new attributes were derived and computed [1]. Today, seismic attributes are used largely for petro physical prediction of reservoirs, lithological and other methodologies which have been developed for many applications to wider hydrocarbon exploration [2]. Porosity refers to the void space part of the rocks. Characterization of reservoir porosity is a very complicated job due to its ingrained heterogeneity [3]. However, its role in petroleum industry, especially in economic success of a reservoir growth, makes the reservoir porosity of large attention.

Core analysis is the oldest and still practiced technique for precise measurement of porosity. However, it is so cost, time, and people terrible. Therefore, this type of data is scars in hydrocarbon fields. On the opposite, classical well log data are available in almost all wells. These logs include invaluable inherent information about the hydrocarbon formations. Many researchers have tried to estimate reservoir properties from conventional well log data and seismic attributes [4]. Genetic algorithms are general purpose search algorithm which use principles inspired by natural populations to find best solutions for many problems. GA is one of the nature inspired algorithm that have been used widely in solving problems requiring an advanced form of heuristic search over a solution space, besides numerical and combinatorial optimization problems. The basic idea is to maintain a population of chromosomes, which represent elect solutions to the problem [5]. In the operation of sexual reproduction, the parents contribute their genetic characteristics by providing half of their child's chromosomes. Then the chromosomes will bend an operation is called crossover, which let the parent's chromosomes cut and then recombine again into new

chromosomes. The gene set are contributed by both parents. Hence, this child get a merge characteristics from both parents. The chromosomes and the genes are referred to the Genotype. While the new properties result from those genes, are referred to as a Phenotype. Sometimes genetic mutations happen, these mutations are altering in the original sequence of genes and can lead to many effects like evolution, health problems, or may have no effect [6].

Each chromosome in the population has a related fitness to locate the chromosomes which are used to form a new chromosome in the competition process. GA has large measure of success in both search and optimization problems. The reason for these success is their ability to utilize the information accumulated about the unknown large search space so that to alignment subsequent searches into useful subspaces, i.e., their adaption. This is the key feature, particularly in large, unknown, and ambituses search spaces, where traditional search tools (enumerative, heuristic, ...) are unsuitable, displaying valid approach to the problems needed an efficient and active search techniques [5].

2. RELATED WORK

Genetic algorithms found a large area of applications for which to generate useful solutions. These applications did not spare any direction such as physics, mathematics, medical, economics, computer, bioinformatics, etc. In the remainder of this section, a brief literature reviews of applications utilizing genetic algorithms for obtaining solutions is presented. Kalyanmoy Deb and Amarendra Kumar, 1995. Explain however the success of binary-coded genetic algorithm in various optimization problems, real-coded genetic algorithm are finding some attention primarily in solving continuous search space problems. Real-coded genetic algorithm differ from the binary-coded genetic algorithm in the coding of the problem variables. Since the problem variables are used directly in real-coded genetic algorithm, there lied a need for developing new, yet efficient, crossover and mutation operators [7]. Jihoon Yang and Vasant Honavar, 1998. Demonstrate the probability of genetic algorithm for feature subset selection, the select of attributes according to accuracy, cost and time [8]. F. HERRERA, et al., 1998. Explained different kinds of real code genetic algorithm's crossover and mutation [5]. Kevin P. Dorrington and Curtis A. Link, 2004. Explain the role of binary genetic algorithm to select the best combination of attributes, then enter the result from genetic algorithm to neural network to predict porosity [9]. Natalia Soubotcheva and Robert R. Stewart, 2004. Utilized stepwise regression method to select the less subset attributes error then enter the result to different types of neural network [10]. Barani R. and Sumathi M., 2013. Utilized a multi-gene, multi-parameter Genetic Algorithm with binary code representation for proposed feature-level image fusion method which fuses

multi-focus images based on feature values at block-level [11]. Ursula Iturrarán-Viveros, Jorge O. Parra, 2014. Utilized gamma test method to select the best combination of seismic attributes, then the result investment to neural network to get high porosity prediction [12]. Adi A. Maaita, et al., 2015. Presented a generic solution which utilized both multi-gene set chromosomes, and an adaptive gene mutation rate scheme for solving complex solution-space search problems that involve multiple sub-solutions [6].

3. THE PRINCIPAL METHODOLOGY

In reflection seismology, a seismic attribute is a quantity extracted or derived from seismic data that can be analyzed in order to improve information that might be more subtle in a traditional seismic image, leading to a good geological or geophysical interpretation of the data. Examples of seismic attributes can include measured time, amplitude and frequency. Seismic attributes have come a long way since their introduction in the early 1970s and have become an important part of seismic interpretation projects. Today, they are being used widely for lithological and petro physical prediction of reservoirs and various methodologies have been developed for their application to broader hydrocarbon exploration and development decision making [2]. Uses genetic algorithm to select the best attributes and find a relationship between the log and seismic data at the well locations, it uses this relation to "predict "or estimate a volume of the log property at the other location of the seismic volume (see Figure 1) [10].

Genetic algorithm is a class of stochastic global search techniques based on biological evolution principles, several have applied genetic algorithm to geophysical optimization problems such as seismic attributes [9]. Genetic algorithm represents parameters as an encoded binary string and works with the binary strings to minimize the cost, while the other works with the continuous parameters themselves to minimize the cost [13]. Genetic algorithms have had a great measure of success in search and optimization problems. The reason for a large part of their success is their ability to exploit the information accumulated about an initially unknown search space in order to bias subsequent searches [5].

Fixed-length and binary coded strings for the representation of the solutions have dominated genetic algorithm research since there are theoretical results that show them to be the most appropriate ones, and as they are amenable to simple implementation. But the genetic algorithm's good properties do not stem from the use of bit strings [14]. For this reason, the path has been laid toward the use of non-binary representations more adequate for each particular application problem. One of the more important ones is the real number representation into useful subspaces. which would seem particularly natural when optimization problems with parameters in continuous search spaces are tackled. So genetic algorithms based on real number representation are called real-coded genetic algorithms (RCGAs) [5]. Binary genetic algorithm solves many optimization problems that stump traditional techniques, but if you are attempting to solve a problem where the values of the parameters are continuous and you want to know them to the full machine precision, in such a problem each parameter requires many bits to represent it. If the number of parameters is large, the size of the chromosome is also large. so 1s and 0s are not the only way to represent a parameter. One could, in principle, use any representation conceivable for encoding the parameters. When the parameters are naturally quantized, the binary genetic algorithm fits nicely.

However, when the parameters are continuous, it is more logical to represent them by real code. In addition, since the binary genetic algorithm has its precision limited by the binary representation of parameters, using real code instead easily allows representation to the machine precision. This continuous genetic algorithm also has the advantage of requiring less storage than the binary genetic algorithm because a single number represents the variable instead of *Nbits*. The continuous genetic algorithm is inherently faster than the binary genetic algorithm, because the chromosomes do not have to be decoded prior to the evaluation of the cost function [13].

4. BINARY-REAL CODE ATTRIBUTES REPRESENTATION

The genetic algorithm attribute selection technique requires an initial set of attribute combinations. Each attribute combination is named a chromosome and the entire set of attribute combinations is designated a population. Each chromosome is described by a code of zeros and ones, with one representing a selected attribute and zero representing an attribute not selected [9]. Each binary digit in the chromosome is a gene and each gen consist of large number of parameters. For example, given a *p* total number of attributes, the number of possible combination of seismic attributes to be use 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. 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), the concept of the population that contain $2^p - 1$ combinations help us to clearly indicate any particular combination of attribute [12].

After representing attributes by binary code put in each 1 string the attribute that return to same column with its own values, then apply the procedure of genetic algorithm with real code genetic algorithm's crossover and mutation, as illustrated in Figure 2. Correlation coefficient, root mean square error and linear regression with multiple attributes were used to evaluate the result of genetic algorithm and the accuracy of porosity prediction, Correlation coefficient *R* measures the relationship of the process inputs on the output. It is the degree or extent of the relationship between two parameters Eq.1. These studies are used to examine if there is a predictive relationship of the input on the process [15].

$$R = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

Root-Mean-Square Error (*RMSE*) is a frequently used measure of the differences between predicted by a model or an estimator and the values actually observed Eq.2 [16].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x-y)^2}{n}} \quad (2)$$

Where:

x: represented predicted values.

y: represented actual values.

n: represented total number of values

Linear regression is a technique for choosing a line to represents the relationship between two variables, based on a set of observed values of the variables, while multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable *x* is associated with a value of the

dependent variable. Consider the problem of predicting porosity with three attributes plus a DC component w_0 [17]:

$$\phi_i = w_0 + w_1 I_i + w_2 F_i + w_3 E_i$$

Where: ϕ_i : porosity, I_i, F_i and E_i : attributes, all at the i^{th} sample, where $i=1 \dots N$, this can be written as a series of linear equations

$$\phi_1 = w_0 + w_1 I_1 + w_2 F_1 + w_3 E_1$$

$$\phi_2 = w_0 + w_1 I_2 + w_2 F_2 + w_3 E_2$$

$$\phi_N = w_0 + w_1 I_N + w_2 F_N + w_3 E_N$$

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ I_1 & I_2 & \dots & I_N \\ E_1 & E_2 & \dots & E_N \\ F_1 & F_2 & \dots & F_N \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} \Rightarrow P = AW$$

This can be solved by least squares minimization to give

$$W = (A^T A)^{-1} A^T P$$

$$\begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 & \dots & 1 \\ I_1 & I_2 & \dots & I_N \\ E_1 & E_2 & \dots & E_N \\ F_1 & F_2 & \dots & F_N \end{bmatrix} \begin{bmatrix} 1 & I_1 & E_1 & F_1 \\ 1 & I_2 & E_2 & F_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & I_N & E_N & F_N \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & \dots & 1 \\ I_1 & I_2 & \dots & I_N \\ E_1 & E_2 & \dots & E_N \\ F_1 & F_2 & \dots & F_N \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix}$$

These weighting coefficients (w_0, w_1, w_2, w_3) minimize the total prediction error (E^2)

$$E^2 = \frac{1}{N} \sum_{i=1}^N (\phi_i - w_0 - w_1 I_i - w_2 E_i - w_3 F_i)^2 \quad (3)$$

4.1 Genetic algorithm procedure

Fitness: it is cost function quantifies the optimality of solution. So the particular solution may be ranked against all the other solution. the function depicts the closeness of a given solution to the desired result. we have to minimize the average some square error between the measured porosity from well log and the output of genetic algorithm [13].
selection: it is important operator in genetic algorithm and represent the stage of a genetic algorithm in which individual genomes are chosen from a population for later breeding (using the crossover operator).in this study used three types of selection roulette wheel selection, tournament selection and random selection [18].

Crossover: After parents have been selected by the selection method, crossover takes place. Crossover operator mates the parents to produce offspring. Crossover consists of swapping chromosome parts between individuals but is not performed on every pair of individuals, as illustrative in Figure 3, The crossover frequency is controlled by a crossover probability (pc). [19].in this study Arithmetic crossover method (AMXO) is used, in arithmetic crossover, two parents produce two

offspring. Thus the parents are as defined by equation (4,5). The offspring are arithmetically the following [21]

$$z_i^{(1)} = \beta_i y_i^{(1)} + (1 - \beta_i) y_i^{(2)} \quad (4)$$

$$z_i^{(2)} = \beta_i y_i^{(2)} + (1 - \beta_i) y_i^{(1)} \quad (5)$$

Where: β are uniform random numbers Furthermore, β is constant in uniform arithmetical crossover or it may vary with regard to the number of generations made in non-uniform arithmetical crossover [6].
Mutation: Mutation can create a new genetic material in the population to maintain the population diversity. It is changing a random part of string representing the individual [20]. Mutation is regulated with the mutation probability, pm . The mutation operator provides a possible mutation on some selected chromosomes. The mutation operator provides a possible mutation on some selected elements of some chromosomes. The chromosome before mutation is given by:

$$y = (y_1, \dots, y_i, \dots, y_n) \quad (6)$$

The element of the chromosome to be mutated is denoted by y_i
the chromosome after mutation is given by:

$$y' = (y_1, \dots, y'_i, \dots, y_n) \quad (7)$$

5. SIMULATION AND RESULT

At the first stage of this study, quality of dataset was computed, the result showed bad correlation and error. Therefore, it was necessary to improve result by genetic algorithm. Table 1 represent the result of binary-real code genetic algorithm applied to 10 seismic attributes in MATLAB and the accuracy of porosity prediction from seismic attributes and well log determined by evaluated the correlation coefficient, root mean square error and linear regression with multiple attributes. See Figure (4...9).

6. CONCLUSION

This study followed a binary-real code genetic algorithm seismic attributes representation to construct relationship between seismic attributes and well log data. Results indicated tournament selection method is capable of producing precise prediction of porosity from seismic attributes because it has high correlation coefficient, low E^2 and low $RMSE$ but it is take more time, more number of iterations to reach to optimal result, also roulette wheel selection method has high average correlation coefficient and less time compared with the tournament selection method.

Table 1. Results of 10 seismic attributes

Measurements	Roulette wheel	Tournament	Random
RMSE	28.3312	27.8445	28.9104
R	0.9460	0.9609	0.9101
E ²	0.0021	0.0015	0.0034
Best Cost	0.15399	0.13909	0.18995
Time in Minutes	3.8487	6.8566	1.0255
No. of Iteration	444	704	101
Best Chromosome	Chrom.3	Chrom.3	Chrom.3

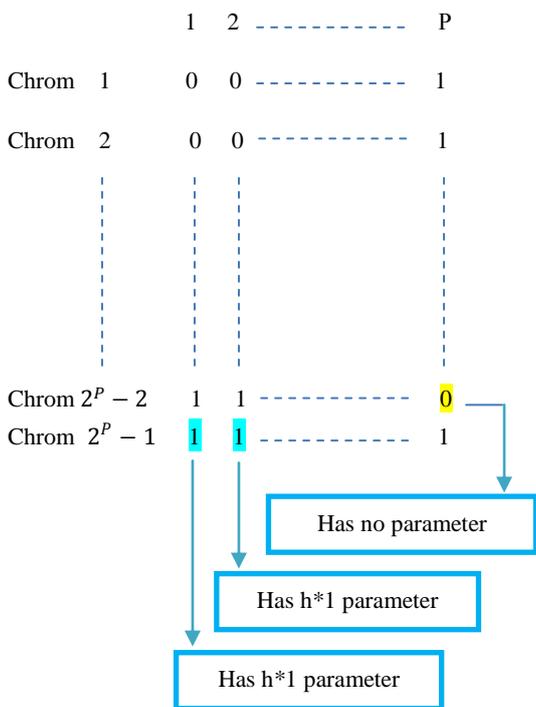


Fig 2: Binary-real code for attributes representation

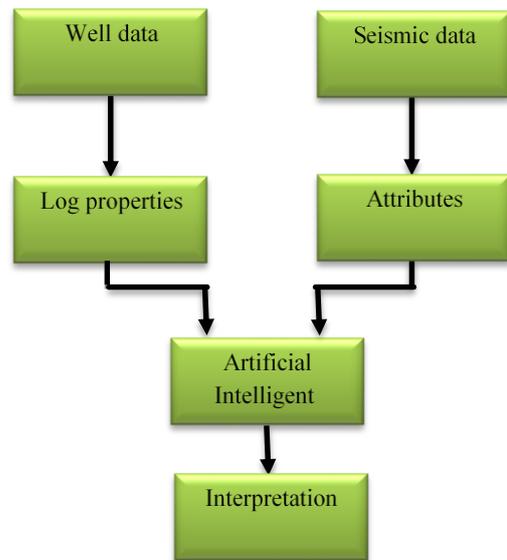


Fig 1: The data driven interpreters by artificial intelligent

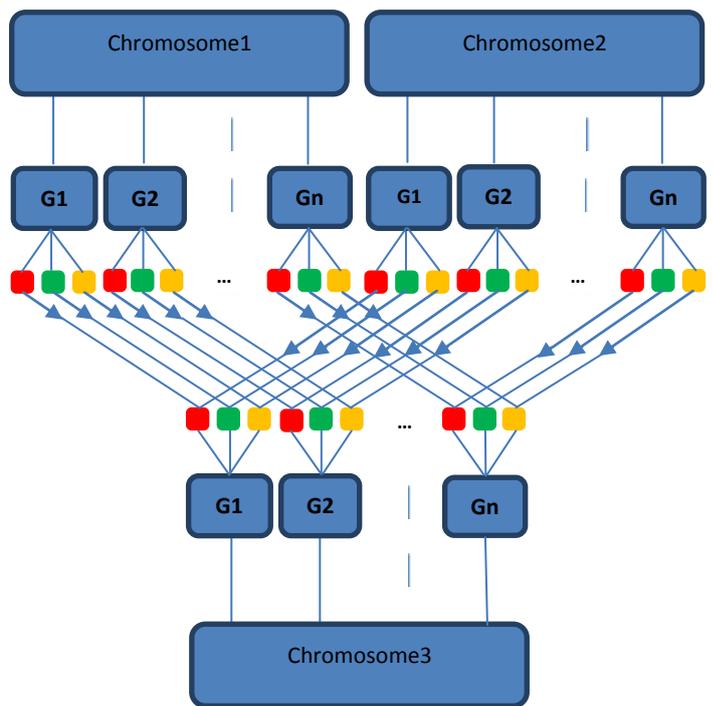


Fig 3: Real code genetic algorithm

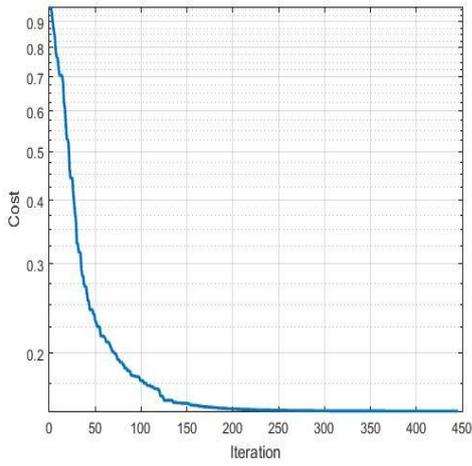


Fig 4: Roulette wheel selection

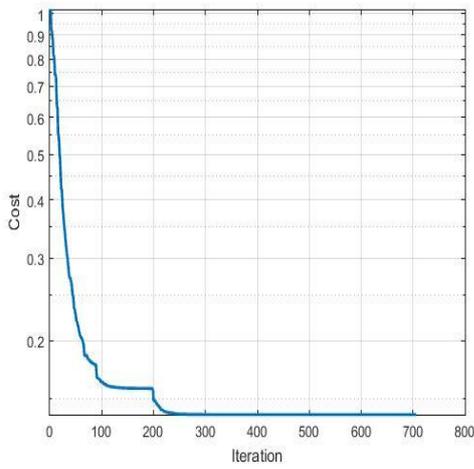


Fig 6: Tournament selection

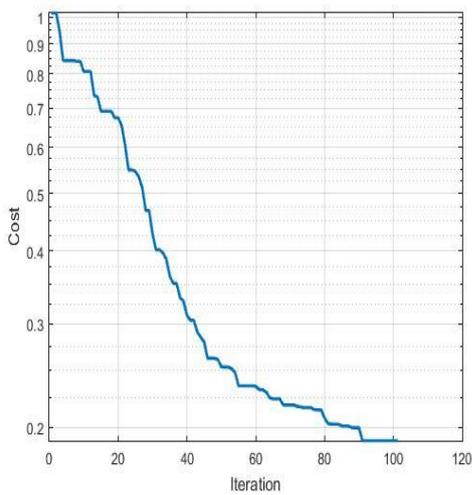


Fig 8: Random selection method

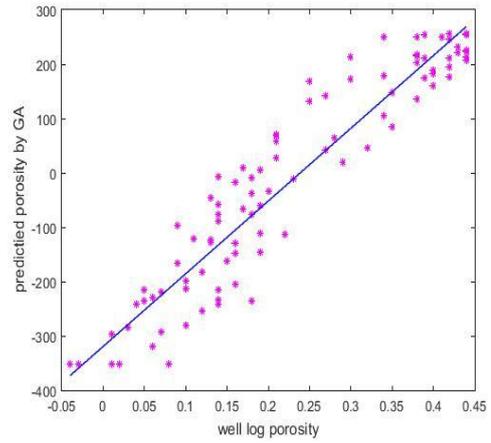


Fig 5: Cross plot showing the correlation coefficient between actual and predicted porosity by roulette wheel selection method

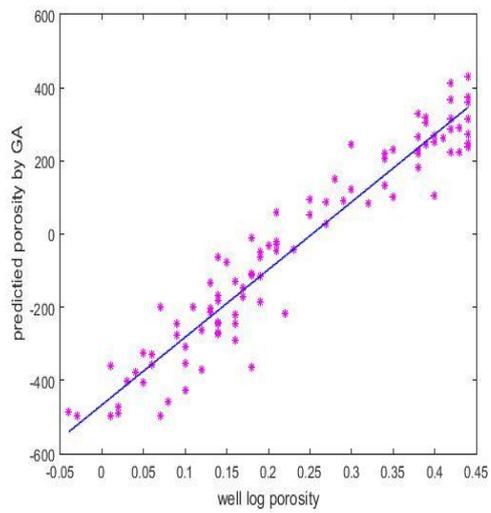


Fig 7: Cross plot showing the correlation coefficient between actual and predicted porosity by tournament selection method

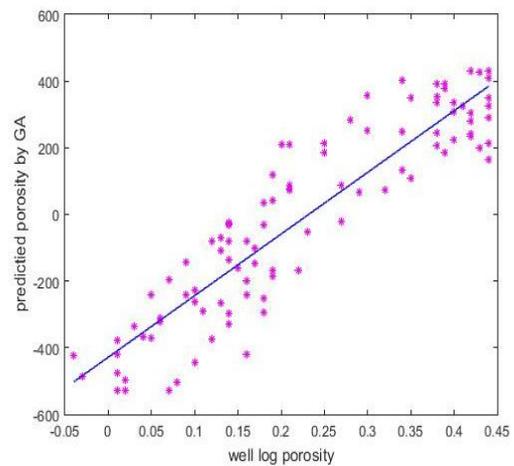


Fig 9: Cross plot showing the correlation coefficient between actual and predicted porosity by random selection method

7. REFERENCES

- [1] D. Subrahmanyam, P.H. Rao, 2008. Seismic Attributes- A Review, International Conference and Exposition on Petroleum Geophysics, P. 398.
- [2] Satinder Chopra and Kurt Marfurt, 2006. Seismic Attributes – a Atomising Aid for Geologic Prediction, Allied Geophysical Laboratories, University of Houston, Houston, Texas, USA.
- [3] Al-Qahtani, F. A. 2000. Porosity Distribution Prediction Using Artificial Neural Networks. College of Engineering and Mining Recourses, West Virginia University, M.Sc. Thesis, 1-3.
- [4] Asoodeh, M., Bagheripour, P. 2012. Prediction of Compressional, Shear, and Stoneley Wave Velocities from Conventional Well Log Data Using a Committee Machine with Intelligent Systems. *Journal of Rock Mechanics and Rock Engineering* 45, 45-63.
- [5] F. HERRERA, M. LOZANO and J.L. 1998. VERDEGAY, Tackling Real-Coded Genetic Algorithms, *Artificial Intelligence Review* 12: 265–319.
- [6] Adi A. Maaita, Jamal Zraqou, Fadi Hamad and Hamza A. Al-Sewadi, 2015. A Generic Adaptive Multi-Gene-Set Genetic Algorithm (AMGA), *International Journal of Advanced Computer Science and Applications*", Vol. 6, No. 5.
- [7] Kalyanmoy Deb, Amarendra Kumar, 1995. Real-coded Genetic Algorithms with Simulated Binary Crossover: Studies on Multimodal and Multiobjective Problems, *Complex Systems* 9, 431-454.
- [8] Jihoon Yang and Vasant Honavar, 1997. Feature Subset Selection Using a Genetic Algorithm. May, 3
- [9] Kevin P. Dorrington and Curtis A. Link, 2004. Genetic-algorithm/neural-network approach to seismic
- [10] attribute selection for well-log prediction", *Geophysics*, vol. 69, No. 1 (january-feburaty), P. 212–221, 14 FIGS., 2 TABLES.
- [11] Natalia Soubotcheva and Robert R. Stewart, (2004). Predicting porosity logs from seismic attributes using geostatistics, Vol. 16.
- [12] Barani R. and Sumathi M., A 2013. Multi-gene, Multi-parameter Genetic Algorithm for Block-Based Feature-Level Image Fusion", CSIP 2013, pp. 188–201. Elsevier Publications.
- [13] Ursula Iturrarán-Viveros, Jorge O. Parra, 2014. Artificial Neural Networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well log data, *Journal of Applied Geophysics* 107, 45–54.
- [14] Randy L. Haupt and Sue Ellen Haupt, 2004. PRACTICAL GENETIC ALGORITHMS, Wiley.
- [15] Antonisse, J. 1989. A New Interpretation of Schema Notation that Overturns the Binary Encoding Constraint. Proc. of the Third Int. Conf. on Genetic Algorithms, J. David Schaffer (Ed.), (Morgan Kaufmann Publishers, San Mateo), 86–91.
- [16] David Shen, WCI, Inc., 2016. Computation of Correlation Coefficient and Its Confidence Interval in SAS, Paper 170-31. www2.sas.com/proceedings/sugi31/170-31.pdf
- [17] 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.*, 7, 1247–1250, 2014.
- [18] Daniel P. Hampson, James S. Schuelke, and John A. Quirein," Use of multiattribute transforms to predict log properties from seismic data", *GEOPHYSICS*, VOL. 66, NO. 1, P. 220–236, JANUARY-FEBRUARY 2001.
- [19] Noraini Mohd Razali, John Geraghty, 2011. Genetic Algorithm Performance with Different Selection Strategies in Solving TSP", Vol II.
- [20] Riikka Peltokangas and Aki Sorsa, 2008. Real-Coded Genetic Algorithms and Nonlinear Parameter identification, No. 34, April.
- [21] Arumugan M.S. and Rao M.V.C. 2004. Novel Hybrid Approaches for Real Coded Genetic Algorithm to Compute the Optimal Control of a Single Stage Hybrid Manufacturing Systems. *International Journal of Computational Intelligence*, Volume 1, Number 3, 189-206.
- [22] Kaelo P. and Ali M.M., 2007. Integrated Crossover Rules in Real Coded Genetic Algorithms. *European Journal of Operational Research*, Volume 176, Issue 1, 60-76.