Artificial Intelligence: Excellent Key for Developing E&P Oil Industry in Iraq

Wathiq j. Almudhafer, SPE, Iraqi South Oil Company
Maitham A. Shaheed, Ph.D. Student, School of Computer Science, Manchester University

Abstract

A Multidisciplinary study for increasing oil recovery has been made in the present paper. This work has been adopted in the Upper Sandstone member/Zubair formation in South Rumaila Oil Field. The work was achieved by using optimization techniques for determining the optimal future reservoir performance regarding to infill drilling. Two different methods of Genetic Algorithm used to optimize the number and locations of infill wells. The first method is simple adaptive genetic algorithm and the second one is the breeder adaptive. The main parameters depended in this study is the cumulative oil production obtained from the output of reservoir simulation software. These two methods of GA depend on using Net Present Value (NPV) as economic analysis as objective function. The optimal number of infill wells is three wells which have maximum cumulative oil production and maximum value of NPV. The same results from two GA methods have been obtained. The locations of these optimal infill wells located in the crest of the oil field.

Introduction

Production and management of oil and gas in today's highly competitive environment require the use of high tech tools. These tools provide the means by which the cost of exploration, production, and management of hydrocarbon resources may be reduced. Engineers find themselves in a never ending race to catch up with new advancements in information technologies. Employing computers in the workplace, incorporating sophisticated simulation models in decision-making processes, and digital control and
monitoring of equipment that were regarded as state of the art only a few years ago, are now normal day-to-day procedures. The phrase "Advanced Technologies" has a highly dynamic meaning. In recent years, Genetic Algorithm, Neural Networks and Fuzzy Logic set theory with its application in artificial intelligence has assumed the new meaning of the phrase "Advanced Technologies". These tools are providing engineers and scientists with the foundation upon which intelligent machines can be developed (4).

In the present study, only Genetic Algorithm has been adopted to increase oil recovery for the main pay in South Rumaila Oil Field. GA offers an efficient search method and can be used as powerful optimization tools introduced by John Holland in 1975(1). Potential solutions generated randomly (population in terms of GAs consist of a number of individuals represented by chromosomes) to a problem compete with each other in order to achieve increasingly better results by applying a set of operators: Selection, Crossover (Recombination), and Mutation. These operators mimic the genetic reproduction in biological sense similar to Darwin’s theory of Natural Selection (1).

**Mechanics of Neural Networks Operation**

An artificial neural network(4) is a collection of neurons that are arranged in specific formations. Neurons are grouped into layers. In a multi-layer network there are usually an input layer, one or more hidden layers and an output layer. The number of neurons in the input layer corresponds to the number of parameters that are being presented to the network as input. The same is true for the output layer. It should be noted that neural network analysis is not limited to a single output and that neural nets can be trained to
build neuro-models with multiple outputs. The neurons in the hidden layer or layers are mainly responsible for feature extraction. They provide increased dimensionality and accommodate tasks such as classification and pattern recognition. Figure 1 is a schematic diagram of a fully connected three layered neural network.

Fig.1 Schematic Diagram of a Three-Layer Neuron Network

In a typical neural data processing procedure, the database is divided into three separate portions called training, calibration and verification sets. The training set is used to develop the desired network. In this process (depending on the paradigm that is being used), the desired output in the training set is used to help the network adjust the weights between its neurons or processing elements \((16, 41)\). During the training process the question arises as when to stop the training. How many times should the network go through the data in the training set in order to learn the system behavior? When should the training stop? These are legitimate questions, since a network can be over trained. In the neural network related literature over-training is also referred to as memorization. Once the network memorizes a data set, it would be incapable of
generalization. It will fit the training data set quite accurately, but suffers in generalization. Performance of an over-trained neural network is similar to a complex non-linear regression analysis \((16,24,41)\).

**Applications of ANN in the Oil and Gas Industry**

Common sense indicates that if a problem can be solved using conventional methods, one should not use neural networks or any other virtual intelligence technique to solve them. For example, balancing your checkbook using a neural network is not recommended. Although there is academic value to solving simple problems, such as polynomials and differential equations, using neural networks to show its capabilities, they should be used mainly in solving problems that otherwise are very time consuming or simply impossible to solve by conventional methods \(^4\).

Neural networks have shown great potential for generating accurate analysis and results from large historical databases. The kind of data that engineers may not consider valuable or relevant in conventional modeling and analysis processes. Neural networks should be used in cases where mathematical modeling is not a practical option. This may be due to the fact that all the parameters involved in a particular process are not known and/or the inter-relation of the parameters is too complicated for mathematical modeling of the system. In such cases a neural network can be constructed to observe the system behavior (what types of output is produced as a result of certain set of inputs) and try to mimic its functionality and behavior \(^{4,24,25}\).

The ANN was applied widely in reservoir characterization. Neural networks have been utilized to predict or virtually measure formation characteristics such as porosity, permeability and fluid saturation from conventional well logs \(^{33,35}\).

Using well logs as input data coupled with core analysis of the
corresponding depth, these reservoir characteristics were successfully predicted for a heterogeneous formation in West Virginia.

**Fuzzy Logic**

The human thought, reasoning, and decision-making process is not crisp. We use vague and imprecise words to explain our thoughts or communicate with one another. There is a contradiction between the imprecise and vague process of human reasoning, thinking, and decision-making and the crisp, scientific reasoning of black and white computer algorithms and approaches. This contradiction has given rise to an impractical approach of using computers to assist humans in the decision-making process, which has been the main reason behind the lack of success for traditional artificial intelligence or conventional rule-based systems, also known as expert systems. Expert systems as a technology started in early 1950s and remained in the research laboratories and never broke through to consumer market \(^{(24)}\).

In essence, fuzzy logic provides the means to compute with words. Using fuzzy logic, experts no longer are forced to summarize their knowledge to a language that machines or computers can understand. What traditional expert systems failed to achieve finally became reality (as mentioned above) with the use of fuzzy expert systems. Fuzzy logic comprises of fuzzy sets, which are a way of representing no statistical uncertainty and approximate reasoning, which includes the operations used to make inferences \(^{(24,41)}\).

Fuzzy set theory provides a means for representing uncertainty. Uncertainty is usually either due to the random nature of events or due to imprecision and ambiguity of information we have about the problem we are trying to solve. In a ran-
of a random process, the outcome of an event from among several possibilities is strictly the result of chance. When the uncertainty is a product of randomness of events, probability theory is the proper tool to use. Observations and measurements can be used to resolve statistical or random uncertainty \(^{(25)}\). For example, once a coin is tossed, no more random or statistical uncertainty remains.

Most uncertainties, especially when dealing with complex systems, are the result of a lack of information. The kind of uncertainty that is the outcome of the complexity of a system is the type of uncertainty that rises from imprecision, from our inability to perform adequate measurements, from a lack of knowledge, or from vagueness (like the fuzziness inherent in natural language). Fuzzy set theory is a marvelous tool for modeling the kind of uncertainty associated with vagueness, with imprecision, and/or with a lack of information regarding a particular element of the problem at hand. Fuzzy logic achieves this important task through fuzzy sets. In crisp sets, an object either belongs to a set or it does not. In fuzzy sets, everything is a matter of degrees. Therefore, an object belongs to a set to a certain degree. For example, the price of oil today is 24.30$ per barrel. Given the price of oil in the past few years, this price seems to be high, but what is a high price for oil? A few months ago, the price of oil was about 10.00$ per barrel. Everybody agrees that 10.00$ per barrel is low. Given how much it costs to produce a barrel of oil in the United States, one can say that the cutoff between low and high for oil price is 15.00$ per barrel. If we use crisp sets, then 14.99$ is low and 15.01$ is high. However, imagine if this was the criterion that was used by oil company executives to make a decision. The fact is, while 15.01 $ is a good price that many people will be happy with, 16.00$ is better and 20.00$ is even better. Categorizing all these prices as high can be quite
misleading. Fuzzy logic proposes the following fuzzy sets for the price of oil as shown in Fig.2.

![Fuzzy Sets Representing the Price of Oil](image)

The most popular (although not yet standard) form of representing fuzzy set and membership information is as follows:

\[ \mu_A(x) = m \]

This representation provides the following information: the membership \( F \) of \( x \) in fuzzy set \( A \) is \( m \). According to the above figure, when the price of oil is 20.00$ per barrel, it has a membership of 0.15 in the fuzzy set “Good” and a membership of 0.85 in the fuzzy set “High”. Using the above notation to represent the oil price membership values,

\[ \mu_{Good}(20.00) = 0.15 \quad \mu_{High}(20.00) = 0.85 \]

**Application of FL in Petroleum Industry**

Fuzzy logic has been used in several petroleum engineering related applications. These applications include petrophysics and reservoir characterization, among others.
hanced recovery \cite{34,35}, infill drilling \cite{36}, decision making analysis \cite{37}, and well stimulation \cite{38,39}. In this section we review an application that incorporates fuzzy logic in a hybrid manner in concert with neural networks and genetic algorithms.

In this example of use of the intelligent systems in petroleum engineering, neural networks, genetic algorithms, and fuzzy logic are used to select candidates for re-stimulation in the Frontier formation in the Green River Basin \cite{39}. As the first step of the methodology, neural networks are used to build a representative model of the well performance in the Frontier formation.

**Application of GA in the Petroleum Industry**

There are several applications of genetic algorithms in petroleum and natural gas industry. The first application in the literature goes back to one of Holland’s students named David Goldberg. He applied a genetic algorithm to find the optimum design for gas transmission lines \cite{2}. Also Genetic algorithms have been used in reservoir characterization \cite{3,4,5} the stimulation candidate selection in tight gas sands \cite{6}, distribution of gas-lift injection \cite{7}, petrophysics \cite{8}, well test analysis \cite{9,10}, and hydraulic fracturing design \cite{11,12}, determining the Value of Reservoir Data \cite{13}, and modeling \cite{14}, Nonconventional Well Deployment \cite{15}, and other petroleum problems \cite{16,17,18,19}.

In the current study, each well represents one gene in the chromosome. The chromosome used in this study consists of eight genes, i.e. each chromosome represents the total number of wells to be optimized. The initial population would be a collection of non limited number of the chromosomes. Actually, the number of chromosomes in the initial population is nearly equal or more than the number of optimized wells.
**Encoding methods**

The decision variables in the current study are the number and locations of proposed wells. The true representation of well locations according to their dimensions is called phenotype. These wells and their number should be converted to genetic terms (genotype) by encoding. The types of encoding are binary, integer, and real valued etc (1).

The genetic algorithm starts with creating initial population of chromosomes. The length of chromosome is equal to the number of proposed wells. Each chromosome consists of eight genes (proposed wells). If the well is selected, the encoding will be 1, if not it will be 0. The selection of genes takes place randomly (23, 25). Fig. 3 shows the binary encoding.

<table>
<thead>
<tr>
<th>well1</th>
<th>well2</th>
<th>well3</th>
<th>well4</th>
<th>well5</th>
<th>well6</th>
<th>well7</th>
<th>well8</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1, y1</td>
<td>x2, y2</td>
<td>x3, y3</td>
<td>x4, y4</td>
<td>x5, y5</td>
<td>x6, y6</td>
<td>x7, y7</td>
<td>x8, y8</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 3: A chromosome in a Genetic Algorithm

**Population and Initialization**

A population consists of a number of individuals each representing a solution for a given problem. This number called population size. A designer of the GAs chooses it. Every chromosome consists of genes which often referred to as the genotype while the decoding creates phenotype based on a genotype.

**Evaluation**

The evaluation criterion is done through an objective function that characterizes an individual’s performance in the problem domain. In the natural world, this would be
an individual’s ability to survive in its present environment. Thus, the objective function establishes the basis for selection of pairs of individuals that will be mated together during reproduction \(^{(20, 24)}\).

**Selection**

Selection is a genetic operator that chooses a chromosome from the current generation’s population for inclusion in the next generation’s population.

**Crossover / Recombination**

Crossover is a genetic operator that combines (mates) two chromosomes (parents) with probability \((P_c)\) to produce a new chromosome (offspring). The idea behind crossover is that the new chromosome may be better than both of the parents if it takes the best characteristics from each of the parents Fig. 4.

**Mutation**

Mutation \(^{(20)}\) is a random change of one or more genes. This can result in entirely new gene values being added to the gene pool. Every chromosome is simply scanned gene by gene and with a mutation rate \((P_m)\) a gene is changed / swapped, i.e. 0 to 1 and 1 to 0. The probability for a mutation is usually kept small, such that we can expect one muted gene per chromosome. As shown in Fig. 5.
Replacement Policy

The Replacement operator removes few relatively poor individuals from populations and replace it with offspring have higher fittest \(^1, 25\).

Stopping criteria

There are many different ways to determine when to stop running the GA and return the best solution. The first method regarding the simple GA method is to stop after a given number of generations and this criterion has been adopted in the current program. Already maximum of generation is one hundred (Max-gen=100). Another is to stop after the GA has converged that is, all individuals in the population are identical as in the breeder GA. The GA can also be halted if the solution...
quality of the population does not improve within a specified number of generations.

**Objective Function Calculation**

In petroleum engineering, the objective function is normally the total oil production or the net present value over a certain time period \(^{27}\). Based on the analysis of a given problem, some parameters that may have significant influence on oil production history and the potential profit are chosen as the decision variables to be optimized. Depending on the constraints provided by the problem or by practical analysis, the value or the range of each decision variable can be determined, and then the domain of the parameter space is specified. After the objective function and all the decision variables are determined, the optimization problem can be formulated as a maximization problem subjected to certain constraints \(^{28}\). The methodology of objective function calculation determines the general framework of the optimization. The steps of calculation can be summarized as follows:

1) By notice the special contour maps for pressure, oil saturation, thickness, permeability, and porosity, determine the region that reasonable to propose possible new well locations \(^{26}\).

2) Implementation the reservoir simulator by setting the new well locations.

3) GA creates an initial population and evaluates fitness of the individuals (objective function evaluation).

4) Then GA selects the individuals based on their probabilities determined by their fitness values.

5) After that the genetic operators (crossover, mutation, and replacement) are adopted in order to achieve the optimal solution.
Net Present Value Formulation

The net present value is defined as the revenues from produced oil and gas sales, after subtracting the costs of disposing produced water and the cost of injecting water and the initial costs \(^{(8),(9)}\). The initial costs represent the capital expenditures. The result is the net cash flow:

\[
\text{Net Cash Flow (t)} = \text{Oil Production (t)} \times \text{Oil Price} + \text{Gas Production (t)} \times \text{Gas Price} - \text{Water Production (t)} \times \text{Water Handling Cost} - \text{Water Injection (t)} \times \text{Water Injection Cost} - \text{OPEX} - \text{CAPEX} \quad (2)
\]

Where:

- Oil price: ($ per STB).
- Gas price: ($ per MSCf).
- Water handling cost: ($ per bbl).
- Water Injection Cost: ($ per bbl).

Eq. (4) can be written in the following form:

\[
NPV = \sum_{t} \frac{NCF(t)}{(1 + i)^t} \quad (3)
\]

Where:

- NPV: net present value.
- NCF: net cash flow.
**Decision variables**

The decision variables in the current problem are the (i, j) coordinates of the wells to be drilled and the number of wells. The dimensions of the problem depend on the number of wells. Since each well has two variables to be optimized which are i and j, the dimension of the problem is equal to power the number of wells, i.e. when the number of optimized wells is eight, the number of iterations is $2^8$ ($2^8=256$).

**Constraints**

The constraints for the current study are the well constraints, which are the water cut (WC), gas-oil ratio (GOR), and bottom hole pressure (BHP) as shown in Table 1. These constraints are treated in the simulator by setting their values in the input files of the simulator. In addition to the locations of the old wells, the locations in east and west flank from the aquifer, and there is no more than well in one grid are also treated as constraints.

### Table 1 Well Constraints

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum, WC</td>
<td>45</td>
<td>percent</td>
</tr>
<tr>
<td>Maximum, GOR</td>
<td>800</td>
<td>SCF/STB</td>
</tr>
<tr>
<td>Minimum, BHFP</td>
<td>2700</td>
<td>psia</td>
</tr>
</tbody>
</table>
Problem description

A commercial finite difference petroleum reservoir simulator was used as the evaluation function, which allowed for the evaluation of detailed information about the reservoir behavior. The decision variables in this case were the (i, j) coordinates of the wells to be located. The hybrid algorithm can handle any number of wells; however, computational issues limit this number. The dimensions of the problem depend on the number of wells. Since each well has two variables to be optimized, the dimension of the problem is equal to twice the number of wells. The objective was to maximize cumulative oil production recovered from the producing wells after two years of production. In an actual case study, the objective function should be the net present value of the process and should incorporate revenues such as oil and gas sales, as well as expenses such as operating, water-handling and injection costs. There is no constraint on the objective function to be used, however in this study the cumulative oil production was used for an easier interpretation of results. Production wells were controlled by total liquid flow rate. The control was switched to bottom hole pressure when the reservoir was no longer able to supply the target flow rate. Water was injected into the injection wells where the water injection rate was specified.

Adaptive GA Program

Adaptive genetic algorithm has been adopted in the current study. Adaptive genetic algorithm is one of genetic algorithm types. It is called adaptive because it generates new population at each iteration and it changes the values of crossover and mutation probabilities also at each iteration. The genetic computer program searches about the optimal solution by breeding population at
each iteration and changing with the values of probability of crossover and mutation without keeping the optimal solution at each iteration. The genetic program is coupled with the simulation program. The genetic algorithm takes two parents from population randomly and produces two children (offspring) by applying the genetic operators such as crossover, mutation, and replacement on the two parents. After that the program sorts the population from best to worst. The resulted chromosome represents the optimal wells at this iteration. Then the GA program re-input the optimized wells in the input file of simulator (SimBest II) (29). After that the simulator is run in order to repeat the genetic operators. This operation is completed by coupling the two programs: simulation and optimization. The genetic program has been written by Pascal under DOS as well as the simulation program is operating under the same environment. The genetic algorithm flow chart is shown in figure 6.

**Simple GA Results**

The optimal solution is the optimal number and locations of infill wells according to maximization of the net present value (NPV). The search result concludes that the optimal number of wells is three (26). The GA program specified to search about the optimal solution from eight wells.
The genetic algorithm parameters at the optimal solution are shown in Table 2.

### Table 2 Simple and Breeder Genetic Algorithm Parameters

<table>
<thead>
<tr>
<th>GA Parameters</th>
<th>Simple (26)</th>
<th>Breeder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Size (popsize)</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Crossover Probability (Pc)</td>
<td>0.6</td>
<td>0.55</td>
</tr>
<tr>
<td>Mutation Probability (Pm)</td>
<td>0.4</td>
<td>0.45</td>
</tr>
<tr>
<td>Maximum Generations (maxgen)</td>
<td>100</td>
<td>53</td>
</tr>
</tbody>
</table>

The locations of optimal well locations on the grid map are (6, 7), (7, 6), and (6, 5) as shown in Fig. 8.

![Simple Genetic Algorithm Results](image)

Fig. 7 Optimal Number of Wells by Simple Genetic Algorithm Program
**Breeder GA Results**

In this method, the genetic computer program searches about the optimal solution by breeding population at each iteration and changing with the values of probability of crossover and mutation but with keeping the optimal solution at each iteration. Therefore you see in Fig. 8 all the values of optimal choice at each iteration are increasing and this is the main difference between simple and breeder GA and there isn’t difference in results, the optimal number and locations of infill wells are the same in the simple and breeder genetic algorithm.
Breeder Genetic Algorithm Results

Fig. 8 Optimal Number of Wells by Breeder Genetic Algorithm Program

Fig. 8 Optimal Locations of Wells by Genetic Algorithm Program
Conclusion

1. Two methods of Genetic Algorithm to optimize number and locations of new infill wells have been developed. The obtained results were the same in the two methods.

2. Using of the net present value as objective function in GA program is found better than using the cumulative oil production because the net present value depends on the economic analysis for determining the optimal future reservoir through infill drilling.

3. Because of the water flooding of most of the east and west flank.

4. The locations of optimal infill wells were located in the crest of the reservoir.

Acknowledgement

I wish to express my appreciation and thanks to Prof. Dr. Mohammed S. Al-Jawad, the ex-Head of Petroleum Engineering Department, Baghdad University and Mrs. Dunia Al-Shamma, Senior Chief Reservoir Engineer, Reservoir Directorate, Iraqi Ministry of Oil for their guidance and support.
References


9. Yin, Hongjun, Zhai, Yunfang, “An Optimum Method of Early-Time Well Test Analysis --Genetic Algorithm, paper SPE 50905 presented at the Inter-
national Oil & Gas Conference and Exhibition in China held in Beijing, China, 2-6 November, 1998.


29. Al-Mudhafer, Wathiq J., Al-Jawad, Mohammed S. and Al-Shamma, Dunia A. (2010) "Reservoir Flow Simulation Study for a Sector in Main Pay-South Rumaila Oil Field" SPE 126427, presented in SPE Oil and Gas India Conference and Exhibition (OGIC) which was held in Mumbai, India, January 20-22.

30. Prof. Dr. Mohaghegh, S. D. "West Virginia University" Personal Communication.


