ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Hybrid Soft Computing for PVT Properties Prediction

Lahouari Ghouti[1] and Saeed Al-Bukhitan[1] *

1- King Fahd University of Petroleum and Minerals - Department of Information and Computer Science
KFUPM Box 1128. Dhahran 31261. Saudi Arabia First Author's school

Abstract. Pressure-Volume-Temperature (PVT) properties are very important in the reservoir engineering computations. There are many approaches for predicting various PVT properties based on empirical correlations and statistical regression models. Last decade, researchers utilized neural networks to develop more accurate PVT correlations. These achievements of neural networks open the door to data mining techniques to play a major role in oil and gas industry. Unfortunately, the developed neural networks correlations are often limited and global correlations are usually less accurate compared to local correlations. Recently, adaptive neuro-fuzzy inference systems have been proposed as a new intelligence framework for both prediction and classification based on fuzzy clustering optimization criterion and ranking. In this paper, a genetic-neuro-fuzzy inference system is proposed for estimating PVT properties of crude oil systems.

1 Introduction

Reservoir fluid properties are very important in petroleum engineering computations such as material balance calculations, well test analysis, reserve estimation, inflow performance calculations, fluid flow in porous media, evaluation of new formation for potential development, numerical reservoir simulations, design of production equipment, and planning future enhanced oil recovery projects. These properties are determined from laboratory studies on samples collected from the bottom of the wellbore or at the surface. Such experimental data are not always available or very expensive to obtain. The solution is to use either empirically derived correlations or the equations of state models to predict the PVT properties.
The relationships of PVT properties for oil, gas, and water are traditionally estimated using empirical studies. Ideally, those properties could be measured in laboratories. The problem with those measurements is the availability of those laboratories and the right samples collected from the well-bore or well surface. To overcome those problems, several correlations between those properties and well data had been developed based on available data for deferent regions in the world. During the last several years, neural network has been used to have a better prediction models than the empirical ones and they have shown a significant prediction improvement. These achievements open the door the machine learning techniques to play a major rule in petroleum engineering.

* This research is funded by King Fahd University of Petroleum and Minerals.

1.1 Problem Statement

Pressure-Volume-Temperature (PVT) properties are very important in the reservoir engineering computations. There are many approaches for predicting various PVT properties based on empirical correlations and statistical regression models. Last decade, researchers utilized neural networks to develop more accurate PVT correlations. These achievements of neural networks open the door to data mining techniques to play a major role in oil and gas industry. Unfortunately, the developed neural networks correlations are often limited and global correlations are usually less accurate compared to local correlations. Recently, adaptive neuro-fuzzy inference systems have been proposed as a new intelligence framework for both prediction and classification based on fuzzy clustering optimization criterion and ranking. In this paper, a genetic-neuro-fuzzy inference system is proposed for estimating PVT properties of crude oil systems.

1.2 Literature Review

1.2.1 Empirical Correlations

Standing [1] proposed the use of graphical correlations for the determination of bubble point pressure (*Pb*) and the oil formation volume factor (*Bob*). In his proposed technique, Standing used 105 experimentally measured data points from 22 different crude-oil and gas mixtures from California oil fields. Average relative errors of 4.8% and of 1.17% were reported for *Pb* and *Bob*, respectively. An empirical equation, based on Henry's law, for estimating the bubble point pressure was developed by Lasater [2]. A total of 137 crude-oil and gas mixtures from North and South America were used to estimate the empirical correlations. An average error of 3.8% was reported. However, no correlation estimations were reported for the *Bob* variable. In [3], Vazuquez and Beggs reported two sets of correlations. 600 data points were used from various locations all over the world to estimate correlation values for the *Pb* and *Bob* variables. Two different types of correlations were presented, one for crude oils with °API > 30 and the other for crude oils with °API < 30. An average error of 4.7% was reported for the correlation estimates of the *Bob* variable. Glaso [4] used a total of 45 oil samples from the North Sea to develop his empirical correlations to calculate the *Pb* and *Bob* values. Average errors of 1.28% and 20.43% were reported for the bubble point pressure and the formation volume factor, respectively. In [5], Al-Marhoun used 160 experimentally determined data points from the PVT analysis of 69 Middle Eastern hydrocarbon mixtures to develop his correlations. Average errors of 0.03% and 20.01% were reported for the *Pb* and *Bob* variable, respectively.

1.2.2 Prediction of PVT Properties Using Artificial Neural Networks

Garb and Elsharkawy [6] were the first to propose the use of artificial neural networks (ANNs) to estimate the *Pb* and *Bob* variables. The proposed ANN models were tested on Middle East crude oils. Two hidden layers were used to model each property separately. The *Pb* model had eight neurons in the first layer and four neurons in the second. The *Bob* model had six neurons in both layers. Both models were trained using 498 data sets collected from the literature and unpublished sources. The models were tested using 22 data points only. Results showed improvement over the empirical correlation methods with reduction in the average error for the *Pb* and *Bob*. In [7], a radial basis function neural network model (RBF) was proposed by Elsharkawy to predict the PVT properties for crude oil and gas. El-Sebakhy [8] identified the reservoir engineering properties based on support vector machine (SVM) and handled the over fitting and local minima in the standard neural networks. As discussed above, there are two critical issues with SVM regarding both high dimensional and uncertain situations, which are common in reservoir engineering problems. Comparative studies were carried out to compare the performance of neural networks to numerical correlations and statistical regression models for bubble point pressure and oil formation volume factor for their accuracy and flexibility in representing hydrocarbon mixtures from different locations worldwide.

2 Proposed Hybrid Model

2.1 Neuro-Fuzzy Inference System

Fuzzy logic (FL) is a well established soft computing technique [9]. It is a design method that can be effectively applied to problems that, because of complex, nonlinear, or ambiguous models, cannot be easily solved using traditional engineering analytical techniques. Fuzzy logic comprises of fuzzy sets, which consist of different means for representing non-statistical uncertainty and approximate reasoning, which includes the operations used to make inferences. Fuzzy theory is a theoretical framework having fuzzy sets and fuzzy logic as its core; it started with the fuzziness concept and its expression in the form of fuzzy.

2.1.1 Fuzzy Inference Systems

The fuzzy theory has found many applications in a variety of fields such as plant process control, auto-immunization, pattern recognition and decision-making. It is an excellent 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 systems perform well on uncertain information, very similar to the way human reasoning does. Moreover, the information in prediction or pattern classification problems is imprecise rather than precise in nature, and fuzzy set theory allows us to properly model this vague information. All the basic concepts, such as, concepts of fuzzy set theory, including fuzzy relations, fuzzification and defuzzification, construction of membership functions, and fuzzy arithmetic are shown in details [9]. Generally, the rule-based fuzzy modeling technique can be classified into three categories, namely the linguistic (Mamdani-type), the relational equation, and the Takagi, Sugeno and Kang (TSK) [9]. In linguistic models, both the antecedent and the consequence are fuzzy sets, while in the TSK model the antecedent consists of fuzzy sets but the consequence is made up of linear equations. Fuzzy relational equation models aim at building the fuzzy relation matrices according to the input-output process data determined. We are going to focus on the use of the neuro-fuzzy systems with the TSK model for predicting the PVT correlations of crude oil systems, because of TSK needs less rules and its parameters can be estimated from numerical data using optimization methods such as least-square algorithms.

2.2 Adaptive Neuro-Fuzzy Inference System

The neuro-fuzzy inference system is a hybrid forecasting/classification framework, which learns the rules and membership functions from data. It is a network of nodes and directional links. Associated with the network is a learning rule, for instance, backpropagation. This hybrid network learns a relationship between inputs and outputs. This type of networks covers different approaches, namely, the Mamdani and TSK types [9]. The construction of TSK fuzzy model from numerical data proceeds in three steps: fuzzy clustering, setting the membership functions, and parameter estimation [9].

2.3 Genetic-Neuro-Fuzzy System

2.3.1 Genetic Algorithm (GA)

Genetic Algorithm (GA) is a search mechanism based on the principle of natural selection and population genetics that are transformed by three genetic operators: selection, crossover and mutation. Each string (chromosome) is a possible solution to the problem being optimized and each bit (or group of bits) represents a value or some variable of the problem (gene). These solutions are classified by an evaluation function, giving better values, or fitness, to better solutions. Each solution must be evaluated by the fitness function to produce a value. Different crossover and mutation rates are used for processing of optimization of genetic algorithms.

2.3.2 Hybrid Genetic Neuro-Fuzzy Inference System

There are several methods in how to combine neuro-fuzzy logic with genetic algorithm. For the current problem, the used of GA is finding the optimal subtractive clustering parameter *Radii* (ra) for all input and output variable through genetic algorithm. The hardest part of this step is selecting and implementing the fitness function.

3 Experimental Results

To evaluate performance of each modeling scheme, entire database is divided using the random selection. 70% of the data was used in building the genetic-neuro-fuzzy learning model with an internal validation of 35%. The remaining 30% of the data was used for testing. Both internal and external validation processes are repeated 100 times. Therefore, of the 782 data points, 382 were used to train the neural network models, the remaining 200 to cross-validate the relationships established during training process and 200 to test model to evaluate its accuracy and trend stability. For testing data, statistical summary to investigate different quality measures corresponding to the genetic-neuro-fuzzy scheme, RBF, feed forward neural networks system, and the most popular empirical correlations in literatures to predict both *Pb* and *Bob* properties. A performance analysis was carried out to compare the performance and accuracy of the new genetic-neuro-fuzzy model to that of the models proposed in the literature. For the ANN models, two hidden layers feed forward neural network based on back propagation (BP) learning algorithm with both linear

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

and sigmoid activation functions were used. For the RBF models, Gaussian functions are used as activation function and number of nodes is three times the number of input variables. For the genetic-neuro-fuzzy inference system, five types of implementations were tried. The least-square error was selected as the fitness function for the GA component. Fig. 1 illustrates the scatter plots of the predicted results versus the experimental data for both *Pb* and *Bob* values using the first data set. These cross plots indicates the degree of agreement between the experimental and the predicted values based on the high quality performance of the genetic-neuro-fuzzy modeling scheme. Table 1 shows the performance results of the proposed hybrid system and the existing algorithms. It should be note that while we provide in this paper the results for data set 1 only, similar performance has been achieved for the remaining two data sets.

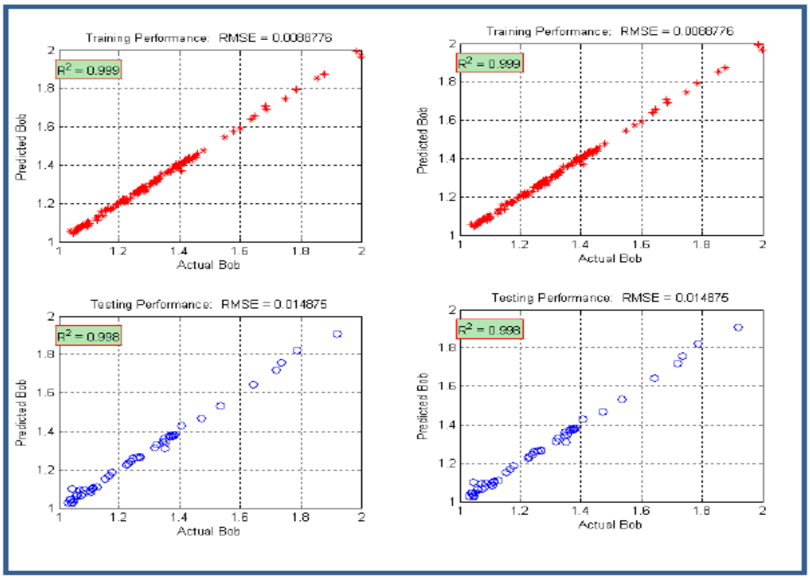

Fig. 1: ESANN 2005: Scatter plots of genetic-neuro-fuzzy interference system for dataset 1.

Factor	Standing	Glaso	Marhoun	RBF	ANN	Sys 1	Sys 2	Sys 3	Sys 4	Sys 5
E_r	5.7	-13.9	6.8	3.7	-5.6	1.4	1.9	-0.3	**4.9**	6.0
E_A	21.7	24.0	8.9	11.5	15.0	12.1	10.6	12.0	**8.2**	11.7
Ea_{min}	0.3	0.3	1.2	0.4	0.1	0.1	0.1	0.1	**0.1**	0.0
Ea_{max}	72.4	65.4	48.9	149.2	113.2	186.5	135.4	127.6	**103.1**	204.3
RMSE	563.8	464.0	173.4	123.7	182.3	131.4	127.7	157.1	**120.0**	137.8
SD	25.9	24.2	10.3	26.2	27.0	19.0	21.0	22.1	**9.0**	17.0
R^2	0.8808	0.9851	0.9970	0.9942	0.9865	0.9933	0.9939	0.9971	**0.9980**	0.9978

Table 1: Statistical quality measures for data set 1.

In the first data set of 160, the root mean square of the neuro-fuzzy scheme **4** on bubble point pressure P_b was the smallest with RMSE = 120, and the largest correlation coefficient with R^2 = 0.9980, while radial bases function network is below genetic neuro Fuzzy with RMSE = 123.7, and R^2 = 0.9942. For the formation volume factor, B_{ob}, the root mean square of the neuro-fuzzy scheme **1** was the smallest with RMSE = 0.010, and the largest correlation coefficient with R^2 = 0.9985|, while radial bases function network is below genetic neuro Fuzzy with RMSE = 0.015, and R^2 = 0.9970. It is clearly indicate from Table 1 that the proposed hybrid system achieves the smallest RMSE and bias values while having the highest correlation coefficients.

4 Conclusions

In this paper, we have proposed a novel hybrid system for the prediction of oil PVT properties. The proposed genetic-neuro-fuzzy inference system achieves better performance compared to the existing modeling schemes. More specifically, the proposed scheme outperforms both the standard feedforward neural networks and the most common empirical correlations in predicting both Pb and Bob properties.

Acknowledgment

The authors would like to thank King Fahd University of Petroleum and Minerals for supporting this research.

References

[1] M. B. Standing, A Pressure Volume Temperature Correlation for Mixture of California Oils and Gases. Drill. & Prod. Prac, API, pages 275-87, Dallas, TX (USA), 1947.

[2] J. A. Lasater, Bubble Point Pressure Correlation. SPE Paper 957-G, 1958.

[3] M. Vazuquez and H. D. Beggs, Correlation for Fluid Physical Property Prediction. SPE 6719, 1980.

[4] O. Glaso, Generalized Pressure-Volume Temperature Correlations. JPT, pages 785-95, 1980

[5] M. A. Al-Marhoun, PVT Correlations for Middle East Crude Oils. SPE Paper 13718, pages 650-66, 1988.

[6] R. B. Garb and A. M. Elsharkawy, Universal Neural-Network Model for Estimating the PVT Properties of Crude Oils. SPE 38099, 1997.

[7] A. M. Elsharkawy, Modeling the Properties of Crude Oil and Gas Systems Using RBF Network. SPE 4996, 1998.

[8] E. A. El-Sebakhy, Forecasting PVT Properties of Crude Oil Systems Based on Support Vector Machines Modeling Scheme. Journal of Petroleum Science and Engineering, 2008.

[9] L. Jong-Se and J. Kim, Reservoir Porosity and Permeability Estimation from Well Logs Using Fuzzy Logic and Neural Networks. SPE 88476, 2004.