Predicting Petroleum Reservoir Properties from Downhole Sensor Data using an Ensemble Model of Neural Networks

Fatai Anifowose
Center for Petroleum & Minerals Research Institute, King Fahd University of Petroleum & Minerals.
+966532649740
fanifowose@gmail.com

Jane Labadin
Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak.
+6(0)12 8881972
ljane@fit.unimas.edu.my

Abdulazeez Abdulraheem
Department Petroleum Engineering King Fahd University of Petroleum and Minerals, Saudi Arabia.
+966138602604
aazeez@kfupm.edu.sa

ABSTRACT

The acquisition of huge sensor data has led to the advent of the smart field phenomenon in the petroleum industry. A lot of data is acquired during drilling and production processes through logging tools equipped with sub-surface/down-hole sensors. Reservoir modeling has advanced from the use of empirical equations through statistical regression tools to the present embrace of Artificial Intelligence (AI) and its hybrid techniques. Due to the high dimensionality and heterogeneity of the sensor data, the capability of conventional AI techniques has become limited as they could not handle more than one hypothesis at a time. Ensemble learning method has the capability to combine several hypotheses to evolve a single ensemble solution to a problem. Despite its popular use, especially in petroleum engineering, Artificial Neural Networks (ANN) has posed a number of challenges. One of such is the difficulty in determining the most suitable learning algorithm for optimal model performance. To save the cost, effort and time involved in the use of trial-and-error and evolutionary methods, this paper presents an ensemble model of ANN that combines the diverse performances of seven "weak" learning algorithms to evolve an ensemble solution in the prediction of porosity and permeability of petroleum reservoirs. When compared to the individual ANN, ANN-bagging and RandomForest, the proposed model performed best. This further confirms the great opportunities for ensemble modeling in petroleum reservoir characterization and other petroleum engineering problems.

Categories and Subject Descriptors

I.2.1 [**Applications and Expert Systems**]: Computer Applications - *Earth and atmospheric sciences, Engineering.*

General Terms

Algorithms, Performance, Experimentation.

Keywords

Ensemble, RandomForest, ANN, bagging, hidden neurons, porosity, permeability, learning algorithms.

MLSDA '13, December 3, 2013, Dunedin, New Zealand.

1. INTRODUCTION

The deluge of data in the petroleum industry acquired from day-to-day data acquisition processes such as Sensing While Drilling has made the petroleum reservoir characterization process complicated and difficult to manage. The data acquired from down-hole sensors that are attached to probes at the end of an insulated cable lowered into a borehole is called a well-log. These logs are recorded digitally as a function of depth. They are used for identifying potential reservoir rock, fracture zones, and estimation of reservoir properties such as porosity and permeability. More details about the acquisition of sensor data through the well-logging process can be found in [1].

Artificial Neural Networks (ANN) has been commonly applied in most scientific and engineering fields, including the reservoir characterization process and other petroleum engineering problems [2, 3, 4, 5]. ANN has been able to meet the acceptable level of accuracy required in petroleum engineering problems despite its limitations and deficiencies [6]. Most petroleum engineering problems require a high level of accuracy to ensure successful exploration, production and management of petroleum resources, hence the persistent quest for algorithms with better and increased performance indices [7]. In addition to its well-known limitations and deficiencies, the successful application of ANN has been marred with various challenges such as [8]:

- Which neural network architectures should be used?
- How large should a neural network be?
- Which learning algorithms are most suitable?

There have been some few reported attempts and continuing effort to handle the first two challenges [9, 10]. To our knowledge, no work has been found in literature to address the third, hence the focus of this paper. Researchers conventionally spend considerable amount of time and effort using trial-and-error methods to determine the most appropriate learning algorithm for their ANN applications. Most times, they settle for what is far from the optimal choice as the searches are not exhaustive enough. Evolutionary algorithms have also been used for the optimization of this phenomenon [11 - 16]. However, since these algorithms are population-based exhaustive search techniques, they also come with their own problems such as occasional inability to converge on global optima [11], possibility of overfitting [12], time and computational complexity [13, 17], and occasional inefficiency [18, 19]. To overcome these problems, we propose an ensemble learning model that combines the performances of different ANN learning algorithms to evolve the best solution to our chosen petroleum engineering problem.

The choice of ANN for this work has been due to its successful application in the petroleum industry. It is pertinent to improve on the performance of a technique that is already being successfully applied in the industry. The choice of the petroleum reservoir characterization process to prove the ensemble learning concept is due to its importance in meeting the world energy needs. Hence, even a marginal improvement in the accuracy of predictive models has the potential to yield results for more effective exploration and production of hydrocarbon resources.

The objectives of this paper are: to establish the need for ensemble learning application as a way to overcoming one of the challenges of ANN design; implement an ANN ensemble model with different learning algorithms; and demonstrate the superiority of our proposed ensemble algorithm over conventional techniques.

The rest of this paper reviews the relevant literature in Section 2, presents a detailed research methodology in Section 3, discusses the results in Section 4, and concludes in Section 5.

2. LITEARATURE REVIEW

2.1 Petroleum Reservoir Characterization

Petroleum reservoir characterization is an important process in the petroleum industry in which various properties of petroleum reservoirs are estimated. The process involves measuring various reservoir rock properties such as porosity, permeability, water saturation, diagenesis, etc. from reservoir rock samples obtained from the field using some specialized laboratory equipment. Out of these properties, porosity and permeability are the most important since they are key indicators of reservoir quantity and quality. Most other reservoir properties depend on these two for their estimation. Porosity, measured in percentages, is the amount of pores in a standard sized rock sample. Permeability, in milliDarsy, is the degree of connectivity of the such pores. A good reservoir rock should have large number of pores that are well connected for easy recovery of hydrocarbons [4, 7, 9, 10].

Empirical equations were used to estimate reservoir properties from laboratory core measurements. However, the entire length of a well could not be cored due to the huge cost. Recently, sensor data obtained from the entire profiles of the fields (serving as input variables) are combined with the available core data (serving as target variables) estimated in the laboratory to predict the desired properties of the uncored reservoir zones. This has saved a lot of time and money. However, since the petroleum industry is always in search of improved and better-performing models for increased prediction accuracies [2 - 7], this study is quite imperative, very relevant and highly desirable.

2.2 The Ensemble Learning Methodology

Much has been written about the ensemble learning methodology and its successful applications from the fundamentals [20, 21] through the early bird implementations [22, 23] to the most recent efforts [9, 24, 25]. This methodology has not been well applied in the petroleum industry despite numerous reports of its successful application and superior performance over individual learners outside the petroleum industry [24, 26, 27]. The few successful applications in petroleum engineering [9, 10, 28, 29], despite their limitations, have been a source of motivation to explore more efficient ensemble solutions. Ensemble models are especially suitable for the petroleum industry where sensor data are usually high-dimensional, heterogeneous, highly noisy and sometimes very scarce. The ensemble methodology is justified by its mimic of the human belief that the consensus of a committee of experts is superior to that of individuals, provided each member of the committee has a reasonable level of expertise [30].

One of the ensemble methods for regression tasks is the Bootstrap Aggregate (bagging) [22] and was first implemented in the RandomForest technique [23]. The bagging method trains a set of "weak" learners and combines their outputs by taking an average of the individual learner's output. A weak learner is a model which is only slightly correlated with the true target. This can be contrasted with a strong learner that arbitrarily correlates well with the true target [31].

The aforementioned studies [9, 10, 28, 29] did not address the third challenge. Refs. [9] and [10] addressed the second challenge by employing the ensemble methodology to determine the optimal number of hidden neurons using exhaustive sequential search and randomized assignment respectively. Ref. [28] did not address any of the challenges but proposed a multi-technique ensemble model that combines the results of ANN, Support Vector Machines (SVM) and Adaptive Neuro-Fuzzy Inference System (ANFIS). We argue that not all researchers in a multi-disciplinary endeavor will have the luxury of mastering three Artificial Intelligence (AI) techniques at a time for the purpose of building an ensemble model. Hence, there is the need to focus on a single technique that is widely used in the industry. A similar argument applies to Ref. [29] that used a multi-objective Genetic Algorithm (GA) to design an ensemble model of ANN in the prediction of open-hole triple-combo of oil wells. Further, the negative sides of using evolutionary algorithms such as GA have been discussed in Section 1. This further explains our reasons for the choice of ANN in this proposed ensemble model.

The major contribution of this paper is the successful application of an ensemble model of a single technique that is already familiar in the petroleum industry. This reduces the effort and complexity of dealing with multiple and possibly unfamiliar techniques.

2.3 Overview of Artificial Neural Networks

ANN, modeled after the biological nervous system, is the most commonly applied AI technique in the petroleum industry. It is made up of layers of neurons interconnected by links with weights assigned. The ANN architecture is generalized [32] as:

$$y_i = f\left(\sum_k w_{ik} x_k + \mu_i\right)$$

where x_k are inputs to the input neuron k, w_{ik} are weights assigned to the inputs of the neuron k, μ_i is a bias, $f(\bullet)$ is a transfer function and y_i is the output of a neurons k.

More details on ANN can be found in [32].

3. RESEARCH METHODOLOGY

3.1 Description of Data

A total of six core and well log datasets for porosity and permeability were used for the design, implementation, and validation of our proposed ensemble model. Three of the datasets are for porosity (Data 1) and were obtained from a drilling site in the Northern Marion Plat-form of North America while the other