Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines

Fatai Anifowose [a,*], Jane Labadin [a], Abdulazeez Abdulraheem [b]

[a] Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
[b] Department of Petroleum Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

ABSTRACT

The ensemble learning paradigm has proved to be relevant to solving most challenging industrial problems. Despite its successful application especially in the Bioinformatics, the petroleum industry has not benefited enough from the promises of this machine learning technology. The petroleum industry, with its persistent quest for high-performance predictive models, is in great need of this new learning methodology. A marginal improvement in the prediction indices of petroleum reservoir properties could have huge positive impact on the success of exploration, drilling and the overall reservoir management portfolio. Support vector machines (SVM) is one of the promising machine learning tools that have performed excellently well in most prediction problems. However, its performance is a function of the prudent choice of its tuning parameters most especially the regularization parameter, C. Reports have shown that this parameter has significant impact on the performance of SVM. Understandably, no specific value has been recommended for it. This paper proposes a stacked generalization ensemble model of SVM that incorporates different expert opinions on the optimal values of this parameter in the prediction of porosity and permeability of petroleum reservoirs using datasets from diverse geological formations. The performance of the proposed SVM ensemble was compared to that of conventional SVM technique, another SVM implemented with the bagging method, and Random Forest technique. The results showed that the proposed ensemble model, in most cases, outperformed the others with the highest correlation coefficient, and the lowest mean and absolute errors. The study indicated that there is a great potential for ensemble learning in petroleum reservoir characterization to improve the accuracy of reservoir properties predictions for more successful explorations and increased production of petroleum resources. The results also confirmed that ensemble models perform better than the conventional SVM implementation.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The ensemble learning paradigm is the most recent Computational Intelligence (CI) tool for combining a "mixture of experts". It has proved to be relevant in solving most challenging industrial problems. Its superior performance over the conventional method of learning individual techniques has been confirmed when applied on classification and regression problems. Its superior performance over the conventional method of learning individual techniques has been confirmed when applied on classification and regression problems. The ensemble learning paradigm is an advancement in the supervised machine learning technology. While the latter searches for the best hypothesis among all possible hypotheses that describe the solution to a problem, the former combines the best hypothesis of different instances of the base learner and its associated hypotheses. The ensemble learning paradigm has gained much ground with classification problems in many fields. However, it is still a new technology whose great benefit is still waiting to be tapped in the petroleum industry. The ensemble learning methodology is a close emulation of the human socio-cultural behavior of seeking several people's opinions before making any important decision [1]. With the reports of the successful application of ensemble modeling over their individual base learners in other areas [2–6], the petroleum industry is in dire need of this new modeling approach in the petroleum reservoir characterization business.

A lot of data is being generated and acquired in the petroleum industry due to the proliferation of various sensor-based logging tools such as Wireline, Open-Hole, Logging-While-Drilling, Measurement-While-Drilling, and seismic measurements of increasing dimensions. Due to the high dimensionality that may be involved in the data acquired through these systems, the ensemble methodology is most ideal for extracting useful knowledge out of them without compromising expert opinions and model performance. For those outside the facilities that may not have access to these voluminous data, the ensemble methodology is still the ideal technique to manage the little data that may be available to them. The ensemble learning methodology is ideal for handling both cases of too much data and too little data [7]. Ensemble models have the capability to combine different architectures of their

* Corresponding author. Present address: The Research Institute, Center for Petroleum & Minerals, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia. Tel.: +966 138604383; mobile: +966 532649740.

E-mail addresses: fanifowose@gmail.com (F. Anifowose), ljane@fit.unimas.my (J. Labadin), aazeez@kfupm.edu.sa (A. Abdulraheem).