Journal of Natural Gas Science and Engineering xxx (2015) 1–12

Contents lists available at ScienceDirect

Journal of Natural Gas Science and Engineering

journal homepage: www.elsevier.com/locate/jngse

Ensemble model of non-linear feature selection-based Extreme Learning Machine for improved natural gas reservoir characterization

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

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

ARTICLE INFO

Article history:
Received 3 September 2014
Received in revised form
3 February 2015
Accepted 12 February 2015
Available online xxx

Keywords:
Reservoir characterization
Porosity and permeability
Feature selection
Ensemble machine learning
Extreme Learning Machine

ABSTRACT

The deluge of multi-dimensional data acquired from advanced data acquisition tools requires sophisticated algorithms to extract useful knowledge from such data. Traditionally, petroleum and natural gas engineers rely on "rules-of-thumb" in the selection of optimal features with much disregard to the hidden patterns in operational data. The traditional multivariate method of feature selection has become grossly inadequate as it is incapable of handling the non-linearity embedded in such natural phenomena. With the application of computational intelligence and its hybrid techniques in the petroleum industry, much improvement has been made. However, they are still incapable of handling more than one hypothesis at a time. Ensemble learning offers robust methodologies to handle the uncertainties in most complex industrial problems. This learning paradigm has not been well embraced in petroleum reservoir characterization despite the persistent quest for increased prediction accuracy. This paper proposes a novel ensemble model of Extreme Learning Machine (ELM) in the prediction of reservoir properties while utilizing the non-linear approximation capability of Functional Networks to select the optimal input features. Different instances of ELM were fed with features selected from different bootstrap samplings of the real-life field datasets. When benchmarked against existing techniques, our proposed ensemble model outperformed the multivariate regression-based feature selection, the conventional bagging and the Random Forest methods with higher correlation coefficient and lower prediction errors. This work confirms the huge potential in the capability of the new ensemble modeling paradigm to improve the prediction of reservoir properties.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Petroleum and natural gas reservoir exploration and production data have become complicated and heterogeneous due to the advancement in data acquisition technologies. Huge data with sometimes high dimensionality are now acquired during exploration and production activities using various advanced data acquisition technologies such as Logging While Drilling (LWD), Measurement While Drilling (MWD) and Sensing While Drilling (SWD). There is the need to discover and extract useful knowledge from these heterogeneous data for effective decision making and successful characterization of various reservoir properties.

The petroleum engineering field has immensely benefited from the application of Computational Intelligence (CI) techniques (Ali, 1994; Anifowose and Abdulraheem, 2011a). A number of prediction and classification problems have been solved using the CI technology. Most of the problems in this field are complex and marred with uncertainties. This is partly due to the high non-linearity involved and partly due to the diversity of expert opinions involved in the solution to the problems. The problem of non-linearity has been partly solved using the CI technology with various techniques such as Artificial Neural Network (Monedero et al., 2012), Type-2 Fuzzy Logic (Olatunji et al., 2011), and Functional Networks (El-Sebakhy et al., 2012).

More recently, hybrid CI technologies have been used to complement the weaknesses of some of the individual CI techniques with the strength of others (Helmy et al., 2010; Wang et al., 2010; Anifowose and Abdulraheem, 2010; Anifowose et al., 2011a, 2013a). However, the selection of optimal features, from the huge data acquired, using the individual CI techniques as well as their hybrid methodologies remains a challenge especially in the face of

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

http://dx.doi.org/10.1016/j.jngse.2015.02.012
1875-5100/© 2015 Elsevier B.V. All rights reserved.

Please cite this article in press as: Anifowose, F.A., et al., Ensemble model of non-linear feature selection-based Extreme Learning Machine for improved natural gas reservoir characterization, Journal of Natural Gas Science and Engineering (2015), http://dx.doi.org/10.1016/j.jngse.2015.02.012

uncertainties. Petroleum and natural gas engineers still sometimes use those features that were selected based on decades-old linearly-derived empirical equations with the CI techniques. This has resulted in different researchers suggesting different sets of features for the prediction of various reservoir properties without considering the non-linear nature of the data. Hence, the optimality of such feature selection methods remains local rather than global.

This feature selection optimality problem can be solved adequately with the ensemble machine learning paradigm. This paradigm has the capability of integrating the result of different base learners (also called "weak" learners) fed with different locally optimal features to evolve a single ensemble hypothesis with a global flavor. We have chosen Extreme Learning Machine (ELM) for the implementation of this learning technique since it is a recent extension of Artificial Neural Networks (ANN) that has been popularly and widely used in the reservoir characterization field (Huang et al., 2004) despite the deficiencies of the latter (Petrus et al., 1995). Functional Network (FN) has been widely published in the literature for its excellent functional approximation capability and efficient backward elimination and forward selection based non-linear feature selection algorithm (Castillo, 1998; El-Sebakhy et al., 2012).

The choice of input features is very important to ensure the most accurate predictions from machine learning based models and avoid the pollution of model performance with irrelevant features. Using suboptimal input features usually lead to underfitting, a situation where a set of input features are not representative enough to generalize on the target feature resulting in large training and test errors (Bramer, 2002; Alpaydin, 2010; Aalst et al., 2010). The feature selection process has become an important pre-processing stage to avoid the curse of dimensionality, a learning problem that causes the prediction error to increase with increasing number of features (Trunk, 1979).

1.1. The basis for ensemble modeling in reservoir characterization

Various studies in an area of application with different features have been presented in literature as being optimal. An interesting example in the petroleum and natural gas engineering knowledge domain is the existence of diverse "expert opinions" on the features that are necessary for the estimation of reservoir rock properties. While Elphick and Moore (1999) used nine features in the estimation of porosity and permeability, calling them "all available logs", Xu et al. (2005) used only three features for the same purpose calling them "conventional logs". In a similar but different development, while Xie et al. (2005) used only three features for permeability estimation, calling it "input logs"; Rafiei et al. (2009) used only one input variable (log porosity) as the "only available" input feature. Most interestingly, Abdulraheem et al. (2007) acquired a log data containing 11 to 13 features for the estimation of reservoir permeability. However, based on a linear correlation analysis, they used only 6 out of them for their prediction task. These six features, apart from being the most linearly correlated to permeability, were also described as the "most commonly used" features. This shows clearly that there is no standard set of features that are available for the prediction of porosity and permeability in the petroleum and natural gas industry.

If we treat all these "optimal" feature suggestions as different expert opinions, ensemble learning is the ideal learning paradigm to incorporate all these opinions to evolve the best solution to a prediction problem. We argue that all the claims should be accepted as valid "expert" opinions based on individual research efforts and available data, while a predictive model should be allowed to determine the optimal values of the phenomena by itself based on the nature of the data and the non-linear relationship among its attributes. This is the concept on which the ensemble learning paradigm is based (Polikar, 2006, 2009).

In this study, we employed this new methodology to implement a novel ELM-based ensemble learning modeling system that incorporates 10 expert opinions on the best set of features for the prediction of porosity and permeability of petroleum reservoirs using datasets that represent different geological formations in the world. The final results of the proposed ensemble model was determined through a set of algebraic combination rules and benchmarked with the results produced by a version of ELM with linearly-selected features, another ELM implemented with the conventional Bootstrap Aggregate (bagging) method and the Random Forest technique, a Decision Tree ensemble model that is based on the conventional bagging method.

It would be unfair to compare the proposed FN-ELM ensemble model with other conventional techniques such as ANN, FN, SVM, and ANFIS without implementing the bagging version of these techniques. Implementing the bagging version of the conventional techniques will be too cumbersome for this paper alone. Because of this, the comparison of this work has been limited to only the conventional ELM technique, its bagging version and the conventional Random Forest ensemble benchmark.

1.2. Contribution to knowledge

The main contributions of this study are:

- The novel combination of the functional approximation capability of FN and the excellent predictive capability of ELM, two advanced versions of the traditional ANN.
- The combination of the conventional bootstrap sampling of the conventional bagging method with non-linear feature selection in a single ELM technique.
- The successful application of this non-linear feature selection based ensemble learning model of ELM for improved prediction of petroleum and natural gas reservoir properties.
- With this proposed ensemble model, different expert opinions on the optimal features for reservoir porosity and permeability predictions could be used to evolve the best solution to a problem without being concerned about the expert opinions themselves.

1.3. Consequences of study

This study is expected to spur more interest for further studies in the application of ensemble modeling in petroleum reservoir characterization. Even a marginal improvement in the prediction accuracy of reservoir properties produced by this proposed model will have a multiplier effect on the performance of full-field reservoir simulation models for more effective exploration and production of energy.

2. Literature review

2.1. Related work

The application of the ensemble learning paradigm has been well established in the machine learning literature. This includes the early fundamental studies of Breiman (1996, 2001) through the later refinements proposed by Polikar (2006, 2009) to the most recent successful application efforts of Sun and Li (2012), Anifowose et al. (2013a) and Wang and Yao (2013). This methodology has not been adequately applied in the reservoir characterization field. The few successful trials (Chen et al., 2004; Anifowose

Please cite this article in press as: Anifowose, F.A., et al., Ensemble model of non-linear feature selection-based Extreme Learning Machine for improved natural gas reservoir characterization, Journal of Natural Gas Science and Engineering (2015), http://dx.doi.org/10.1016/j.jngse.2015.02.012