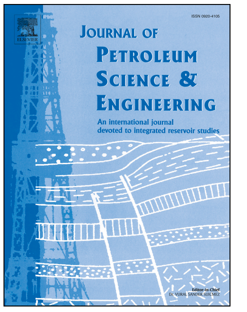
Accepted Manuscript



Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances

Yunxin Xie, Chenyang Zhu, Wen Zhou, Zhongdong Li, Xuan Liu, Mei Tu

PII: S0920-4105(17)30809-4

DOI: 10.1016/j.petrol.2017.10.028

Reference: PETROL 4351

To appear in: *Journal of Petroleum Science and Engineering*

Received Date: 17 March 2017

Revised Date: 1 October 2017

Accepted Date: 10 October 2017

Please cite this article as: Xie, Y., Zhu, C., Zhou, W., Li, Z., Liu, X., Tu, M., Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances, *Journal of Petroleum Science and Engineering* (2017), doi: 10.1016/j.petrol.2017.10.028.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Evaluation of machine learning methods for formation lithology identification: a comparison of tuning processes and model performances

Abstract
Identification of underground formation lithology from well log data is an important task in petroleum exploration and engineering. Recently, several computational algorithms have been used for lithology identification to improve the prediction accuracy. In this paper, we evaluate five typical machine learning methods, namely the Naïve Bayes, Support Vector Machine, Artificial Neural Network, Random Forest and Gradient Tree Boosting, for formation lithology identification using data from the Daniudui gas field and the Hangjinqi gas field. The input to each model consists of features selected from different well log data samples. To determine the best model to classify the lithology type, this study used validation curve to determine the parameter search range and adopted the hyper-parameter optimization method to obtain the best parameter set for each model. The performance of each classifier is also evaluated using 5-fold cross validation. The results suggest that ensemble methods are good algorithm choices for supervised classification of lithology using well log data. The Gradient Tree Boosting classifier is robust to overfitting because it grows trees sequentially by adjusting the weight of the training data distribution to minimize a loss function. The random forest classifier is also a suitable option. An evaluation matrix showed that the Gradient Tree Boosting and Random Forest classifiers have lower prediction errors compared with the other three models. Although all the models have difficulties in distinguishing sandstone classes, the Gradient Tree Boosting performs well on this task compared with the other four methods. Moreover, the classification accuracy is remarkably similar across the lithology classes for both the Random Forest and Gradient Tree Boosting models.

Key Words: lithology identification; supervised learning; gradient boosting; tuning parameter

1. Introduction

Geophysical well log data have advantageous characteristics such as high vertical resolution, good continuity and convenient data acquisition. Therefore, they are an important material resource for underground rock information. Lithology classification based on well log data is the basis of reservoir parameter calculations, and provide the foundation for geological research studies in such fields as sedimentary facies and the environment. Apart from the significance in formation evaluation and geological analysis, lithology interpretation also has practical value in reserve calculation at exploration stage and detailed reservoir description at development stage. Two ways to determine the lithologies and lithofacies are to make inferences from the cuttings obtained during drilling operations and through observation and analysis of the core samples taken from the underground formations (Salehi and Honarvar, 2014). However, these two approaches are not always reliable because different geologists may provide different interpretations (Akinyokun, et al.,2009). Given the constraints on the sample data, the trend has been toward the use of well log data, which can serve not only to predict general petrophysical parameters but also as a tool for sedimentologists and reservoir engineers (Serra and Abbott, 1982). However, the well log data could be highly sampled and numerous, which can burden the geologist who must integrate the data with their workflow and interpret the lithology within certain time constraints (Horrocks, et al., 2015).

Since the introduction of well logs, many mathematical methods have been used to predict lithology based on well log data (Delfiner, et al.,1987). In recent years, using computer technology to automatically predict lithology is becoming an important aspect in well logging and drilling technologies. These computer technologies assist the geologists to avoid the unnecessary data analysis work and improve the lithology identification accuracy. Given the approaches that can identify different grain size of clastic rock with better accuracy, geologists can build better quantitative reservoir evaluation models of different grain size, which can also improve the precision of reservoir evaluation.

Several machine learning techniques have been introduced to lithology classification and identification, including the support vector machine, neural network and random forest classifiers. The support vector machine classification formulation is achieved using features selected based on fuzzy logic from well logs. This approach performs better than do probabilistic neural networks (Al-Anazi, et al.,2010). Adopting the radial basis function kernel also improve the classification accuracy because it has been found to yield the minimum misclassification rate error (Sebtosheikh, et al.,2015). Sebtosheikh also concluded that it is beneficial to implement the normalized polynomial kernel function by using the optimum values obtained

Download English Version:

https://daneshyari.com/en/article/8125490

Download Persian Version:

https://daneshyari.com/article/8125490

Daneshyari.com