Journal of Petroleum Science and Engineering 134 (2015) 143–149

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

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

Lithology prediction by support vector classifiers using inverted seismic attributes data and petrophysical logs as a new approach and investigation of training data set size effect on its performance in a heterogeneous carbonate reservoir

Mohammad Ali Sebtosheikh [a,*], Ali Salehi [b]

[a] Sebt Petro Company, Tehran, Iran
[b] National Iranian South Oil Company, Ahwaz, Iran

ARTICLE INFO

Article history:
Received 15 November 2013
Received in revised form
11 April 2015
Accepted 4 August 2015
Available online 5 August 2015

Keywords:
Lithology prediction
Support vector machines
Training data set size
Heterogeneous carbonate reservoirs
Seismic attributes data.

ABSTRACT

Lithology prediction is one the most affective requirements in all of the petroleum engineering embranchments. Petrophysical analysis, geophysical modeling, statistical methods and artificial intelligent approaches have been used to lithology prediction. Support vector machines (SVMs) based on statistical learning theory (SLT) and the principles of structural risk minimization (SRM) and empirical risk minimization (ERM) use an analytical approach to classification and regression. In this research, SVM classification method is used to lithology prediction from inverted seismic attributes data and petrophysical logs based on petrographic studies of cores lithology in a heterogeneous carbonate reservoir in Iran. Also, because of high impact of the data set size on most of machine learning techniques, effect of training data set size on different SVMs was deliberated by training and testing SVMs by six different partitioned cases according to the learning ratio of each case. Data preparation including normalization, attribute selection, kernel parameters optimization by grid search technique and data partitioning to construct training and testing data sets were performed on the data. The results showed that the SVM performs well in lithology prediction using inverted seismic attributes data and petrophysical logs, and by training data set size reduction, SVM performance has not affected too much, which it is an advantage for SVM as a machine learning method. Also, in order to predict lithology by SVMs using small training data sets, it is recommended to use normalized polynomial kernel function. Kernel functions and generally SVMs are not affected by the training data set size when the learning ratio varies in normal learning ratios. Using the kernels with their associated optimum values of the parameters obtained from grid search technique, it is possible to predict lithology in the investigated reservoir.

1. Introduction

Lithology prediction using petrophysical logs and seismic data is one of the most important issues in all fields of the petroleum engineering such as reservoir characterization, formation evaluation, geological studies, and reservoir modeling, enhanced oil recovery processes and well planning including drilling and well completion management, special in heterogeneous reservoirs. In order to decision making in petroleum engineering lithology should be known by seismic data and petrophysical logs. Lithology can be known from drilling cuttings that usually it is not accurate; also it can be obtained from core analysis that because of the coring costs, it is too expensive and may not be economic. Lithology also can be predicted by geophysical modeling and geophysical inversion method. Accordingly, petrophysical logs and seismic data are used to lithology identification as a accurate, more efficient and cheaper approach than drilling cutting and core analysis (Rider, 2002). Furthermore, the seismic cube data can be used to lithology recognition of reservoir. Raw seismic data may not be applicable in lithology determination. Time, amplitude and frequency are the basic information in seismic data, so there are many fewer measurements than there are unknowns. As filters, Seismic attributes may be helpful to interpret seismic data (Brown, 1999). So these filters, known as seismic attributes, such as coherence, structural smoothing, chaos, graphic equalizer, etc., can be applied to identify complications and analyze and interpret them. There are some traditional methods that have been

* Corresponding author.
E-mail address: masebt@yahoo.com (M.A. Sebtosheikh).

http://dx.doi.org/10.1016/j.petrol.2015.08.001
0920-4105/

developed to quick evaluation of lithology identification from petrophysical logs by combining them such as cross plots (Ellis and Singer, 2008). Traditional cross plotting methods in large data sets of heterogeneous reservoir have lost their efficiency. Several approaches have been introduced to lithology classification and prediction such as support vector machines using conventional wireline well logs (Sebtosheikh et al., 2015), cross plots interpretation and statistical analysis (Delfiner et al., 1987), statistical analysis based on histogram plotting (Busch et al., 1987), associating analysis by fuzzy logic, neural network and multivariable statistical methodologies (Carrasquilla et al., 2008), artificial intelligent approach and multivariate statistical analysis (Lim et al., 1999), hybrid neural network methods (Chikhi and Batouche, 2007), Self-Organized Map (SOM) as a probabilistic unsupervised neural method (Chang et al., 2002; Chikhi and Batouche, 2005) fuzzy logic technique (Cuddy, 2000), artificial neural network methodologies (Katz et al., 1999; Tang, 2009; Raeesi et al., 2012), multi-agent collaborative learning architecture approach (Gifford and Agah, 2010), multivariate statistical method (Tang and White, 2008), aggregation of principal component, clustering and discriminate analysis (Teh et al., 2012), statistical characterization, discrimination and stratigraphic correction methodologies (Borsaru et al., 2006).

Performances of artificial neural networks and fuzzy logic methods are better than statistical analysis (Busch et al., 1987; Katz et al., 1999; Chang et al., 2002; Carrasquilla et al., 2008; Tang and White, 2008; Tang, 2009; Raeesi et al., 2012). Self-Organized Map (SOM) method shows a better performance in lithology classification rather than other used machine learning methods (Chikhi and Batouche, 2005). Probabilistic neural network is slower than other kinds of neural networks because it involves more computational steps (Tang, 2009).

SVM based on the statistical learning theory (SLT) and the principles of SRM and ERM was invented (Boser et al., 1992). SVM has shown good performance in classification tasks. This is distinguished by this fact that SVMs minimize an upper bound of the generalization error while maximize the margin between the separating hyper planes (Amari and Wu, 1999). Recently, SVMs have been applied successfully to a number of applications such as drug design (Burbidge et al., 2001), lithology prediction using conventional wireline well logs (Sebtosheikh et al., 2015), fault diagnosis in power transmission systems (Ravikumar et al., 2008), microarray data classification (Huerta et al., 2006), protein structure prediction (Hua and Sun, 2001), text detection in digital videos (Shin et al., 2000), microarray data and satellite radiance data classification (Lee et al., 2004), speaker identification (Mezghani et al., 2010), document classification (Wang and Sun, 2011) and hyper spectral images classification (Ding, 2011).

In this research, SVM classification method is used to lithology prediction from petrophysical logs and acoustic impedance cube data attained from seismic attributes data based on petrographic studies of cores lithology in a heterogeneous carbonate reservoir in Iran.

The size of training data set effects on some of machine learning methods. In other words, with a small training data set, they construct low accurate decision functions that makes generalization error and finally results in increasing the misclassification error rate. Thus effect of training data set size on different SVMs was investigated by training and testing data sets construction in six different partitioning cases according to the learning ratio of each case. A decision function should be constructed using training instances of each case to predict lithology from petrophysical logs and seismic attributes data using SVMs, and also its parameters should be optimized. To prevent machine learning perplexity by irrelevant or uncouth attributes to the input data set and also to choose the most effective attributes on the SVM performance, an attribute selection approach was used. In order to find the best SVM parameters, a grid search technique has been performed to optimize SVM parameters. The effect of different kernel types and data set size impact on the SVM performance was examined.

2. Methodology

2.1. Support vector machines

SVMs represent a machine learning algorithm based on SLT for both classification and regression tasks (Hamel, 2009; Alpaydin, 2010). SVMs, as classifiers, are dual maximum margin classifiers that be used to classify linear and nonlinear separable data sets (Schölkopf and Smola, 2002; Steinwart and Christmann, 2008; Hamel, 2009; Hastie et al., 2009). Because of SVMs ability to model complex nonlinear decision boundaries, they are highly accurate decision function makers (Han and Kamber, 2006). By allowing the underlying maximum-margin classifiers to make mistakes on the training data set, SVMs can be nominated to soft-margin classifiers. This is fulfilled through the introduction of slack variables (Steinwart and Christmann, 2008).

2.1.1. Dual maximum margin optimization

First the margin between the classes' boundaries is formulated in the direct space and then is transformed into the dual space by means of the Lagrangian in order to find an optimal decision function (Boser et al., 1992; Hamel, 2009). SVMs are shown as the dual of maximum margin classifier that derivation of this dual is performed by applying the technique of the Lagrangian dual to maximum margin classifiers. Constitution of the corresponding Lagrangian of the optimization of a dual maximum margin optimization eventuates Eq. (1) (Hamel, 2009)

$$\begin{aligned} L\left(\vec{\alpha}, \vec{w}, b\right) &= \phi\left(\vec{w}, b\right) - \sum_{i=1}^{L} \alpha_i g_i\left(\vec{w}, b\right) \\ &= \frac{1}{2}\vec{w}\,.\,\vec{w} - \sum_{i=1}^{l} \alpha_i\left(y_i\left(\vec{w}\,.\,\vec{x}_i - b\right) - 1\right) \\ &= \frac{1}{2}\vec{w}.\,\vec{w} - \sum_{i=1}^{l} \alpha_i y_i \vec{w}\,.\,x_i + b\sum_{i=1}^{l} \alpha_i y_i + \sum_{i=1}^{l} \alpha_i \end{aligned} \quad (1)$$

The $g_i(\vec{w}, b)$ is the constraint of the primal optimization problem. The values $\alpha_1, \ldots, \alpha_l$ called the Lagrangian multipliers in the vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_l)$. As $\vec{w}$ and b are called the primal variables and $\vec{\alpha}$ is the dual variable.

The Karush–Kuhn–Tucker (KKT) conditions should be satisfied by the solutions $\vec{w}^*$, $\vec{\alpha}^*$ and b^* (Steinwart and Christmann, 2008)

$$\frac{\partial L}{\partial \vec{w}}\left(\vec{\alpha}^*, \vec{w}^*, b^*\right) = \vec{0}, \quad (2)$$

$$\frac{\partial L}{\partial b}\left(\vec{\alpha}^*, \vec{w}^*, b^*\right) = 0, \quad (3)$$

$$\alpha_i^*\left(y_i\left(\vec{w}^*.\ \vec{x}_i - b^*\right) - 1\right) = 0, \quad (4)$$

$$y_i\left(\vec{w}^*.\ \vec{x}_i - b^*\right) - 1 \geq 0, \quad (5)$$

The optimum vector of $\vec{w}^*$ lies on the saddle point of Lagrangian dual of Eq. (1) and satisfies Eq. (2), that is