Journal of Applied Geophysics 107 (2014) 93–101

Contents lists available at ScienceDirect

Journal of Applied Geophysics

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

Estimation of Reservoir Porosity and Water Saturation Based on Seismic Attributes Using Support Vector Regression Approach

S.R. Na'imi [a,*], S.R. Shadizadeh [a], M.A. Riahi [b], M. Mirzakhanian [c]

[a] Abadan Facility of Petroleum University of Technology, Petroleum University of Technology, Abadan, Iran
[b] Geophysics Institute, University of Tehran, Tehran, Iran
[c] Geophysics Office, Exploration Directorate of National Iranian Oil Company, Tehran, Iran

ARTICLE INFO

Article history:
Received 11 October 2013
Accepted 14 May 2014
Available online 22 May 2014

Keywords:
Porosity
Water saturation
Seismic Attributes
Support Vector Regression (SVR)

ABSTRACT

Porosity and fluid saturation distributions are crucial properties of hydrocarbon reservoirs and are involved in almost all calculations related to reservoir and production. True measurements of these parameters derived from laboratory measurements, are only available at the isolated localities of a reservoir and also are expensive and time-consuming. Therefore, employing other methodologies which have stiffness, simplicity, and cheapness is needful. Support Vector Regression approach is a moderately novel method for doing functional estimation in regression problems. Contrary to conventional neural networks which minimize the error on the training data by the use of usual Empirical Risk Minimization principle, Support Vector Regression minimizes an upper bound on the anticipated risk by means of the Structural Risk Minimization principle. This difference which is the destination in statistical learning causes greater ability of this approach for generalization tasks. In this study, first, appropriate seismic attributes which have an underlying dependency with reservoir porosity and water saturation are extracted. Subsequently, a non-linear support vector regression algorithm is utilized to obtain quantitative formulation between porosity and water saturation parameters and selected seismic attributes. For an undrilled reservoir, in which there are no sufficient core and log data, it is moderately possible to characterize hydrocarbon bearing formation by means of this method.

1. Introduction

Through the recent years, obvious advances have occurred in the study and application of intelligent systems as strong tools for extracting quantitative formulation between two sets of data (inputs/outputs) that have a basic dependency in the petroleum industry. Seismic data have low resolution and are routinely used only to achieve a structural analysis of the reservoir; but because of their areal coverage researchers have always aimed to use seismic data and their attributes as predictive variables in lithology prediction and reservoir characterization projects. Several researchers have worked on predicting water saturation and porosity from seismic data using statistical methods and intelligent systems (Balch et al., 1999; Nikravesh and Hassibi, 2003; Nikravesh et al., 2001; Russell et al., 2003; Soubotcheva and Stewart, 2006).

Support Vector Regression (SVR)s are supervised machine learning algorithm that have been introduced in the framework of Structural Risk Minimization (SRM) and theory of VC (Vapnik–Chevronenkis bound) and are especially suitable for utilizing with non-linear multi-attributes. For the cases where the populations are small (i.e., only a few well-seismic attribute pairs), statistical significance may be unattainable to achieve and Neural network can be easily over trained and result in "over fitting"; but SVR has good performance on unseen data (Al-Anazi and Gates, 2009, 2010; Gholami and Moradzadeh, 2011; Karimian et al., 2013; Li, 2005; Na'imi et al., 2014; Nazari et al., 2011; Shafiei et al., 2012; Yanzhou et al., 2005; Zhao et al., 2006). In this study, the propounded methodology was applied to small well log data set of one of the Iranian oil field sandstone reservoir prosperously.

2. Support Vector Regression

The basic concept of Support Vector Regression (SVR) for estimation of an unknown continuous valued function in regression problem is to map the input data from the original space into a higher m-dimensional feature space (sometimes even infinite-dimensional space), and compute an optimal linear regression function in this feature space (Cherkassky and Mulier, 1998; Na'imi et al., 2014; Smola and Schölkopf, 1998; Vapnik, 1998, 1999; Vapnik et al., 1997; Wu

* Corresponding author. Tel.: +98 911 336 9144.
E-mail address: srnaimi@gmail.com (S.R. Na'imi).

http://dx.doi.org/10.1016/j.jappgeo.2014.05.011
0926-9851/

et al., 2004). The solution for this linear regression function in feature space produces a nonlinear regression function in original input space that can estimate unseen values accurately.

In mathematical notation, Considering training data set as $\{(x_1,y_1),\ldots,(x_\ell,y_\ell)\}$, where each $x_i \subset R^n$ is the input space and has a corresponding target value $y_i \subset R$ for $i = 1,\ldots, l$ where l refers to the size of the training data (Vapnik, 1998; Kecman, 2001).

The SVR estimating function is (Vapnik, 1998; Wu et al., 2004):

$$f(x) = (w \cdot \Phi(x)) + b \tag{1}$$

where $w \subset R^n$, $b \subset R$ and Φ is a non-linear transformation from R^n to high dimensional space. Now, the goal is to obtain the value of w and b such that values of x can be determined by minimizing the regression risk (Vapnik, 1998; Wu et al., 2004):

$$R_{reg}(f) = C\sum_{i=0}^{\ell} L(f(xi) - yi) + \frac{1}{2}\|w\|^2 \tag{2}$$

where C is a constant that controls penalties to estimation errors and is defined by user, $L(\cdot)$ is a cost function, and vector w can be written in terms of data points as (Vapnik, 1998; Wu et al., 2004):

$$w = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*)\Phi(x_i) \tag{3}$$

where α_i $and\alpha_i^*$ are the Lagrange multipliers. By substituting Eq. (3) into Eq. (1) (Vapnik, 1998; Wu et al., 2004):

$$f(x) = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*)(\Phi(x_i) \cdot \Phi(x)) + b = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{4}$$

In Eq. (4) the dot product can be replaced with function $k(x_i,x)$, known as the kernel function. The kernel function is a function in input space. So the required scalar products in a feature space are calculated directly by computing kernels for given training data vectors in an input space and this property leads to no need of knowing what the actual mapping Φ is. (Kecman, 2001) The most common kernels for regression are Polynomial, Radial Basis Function (RBF) and Sigmoid (S-shaped).

To ignore errors the ε-insensitive loss function, proposed by Vapnik (1998, 1999), is the most prevalent cost function:

$$L(f(x) - y) = \begin{cases} |f(x) - y| - \varepsilon, & for \quad |f(x) - y| \geq \varepsilon \\ 0 & otherwise \end{cases} \tag{6}$$

By solving the quadratic optimization problem in (7), the regression risk in Eq. (2) and the ε-insensitive loss function (6) can be minimized (Vapnik, 1998; Wu et al., 2004):

$$\frac{1}{2}\sum_{i,j=1}^{\ell} (\alpha_i^* - \alpha_i)\left(\alpha_j^* - \alpha_j\right)k\left(x_i, x_j\right) - \sum_{i=1}^{\ell} \alpha_i^*(y_i - \varepsilon) - \alpha_i(y_i + \varepsilon)$$

subject to

$$\sum_{i=1}^{\ell} \alpha_i - \alpha_i^* = 0, \quad \alpha_i, \alpha_i^* \in [0, C] \tag{7}$$

Solving above quadratic problem determines the Lagrange multipliers, α_i $and\alpha_i^*$. The non-zero values of the Lagrange multipliers in Eq. (7) are useful in forecasting the regression line and are known as support vectors. For all points inside the ε-tube, the Lagrange multipliers equal to zero do not contribute to the regression function. Only when the requirement $|f(x) - y| \geq \varepsilon$ is fulfilled, Lagrange multipliers

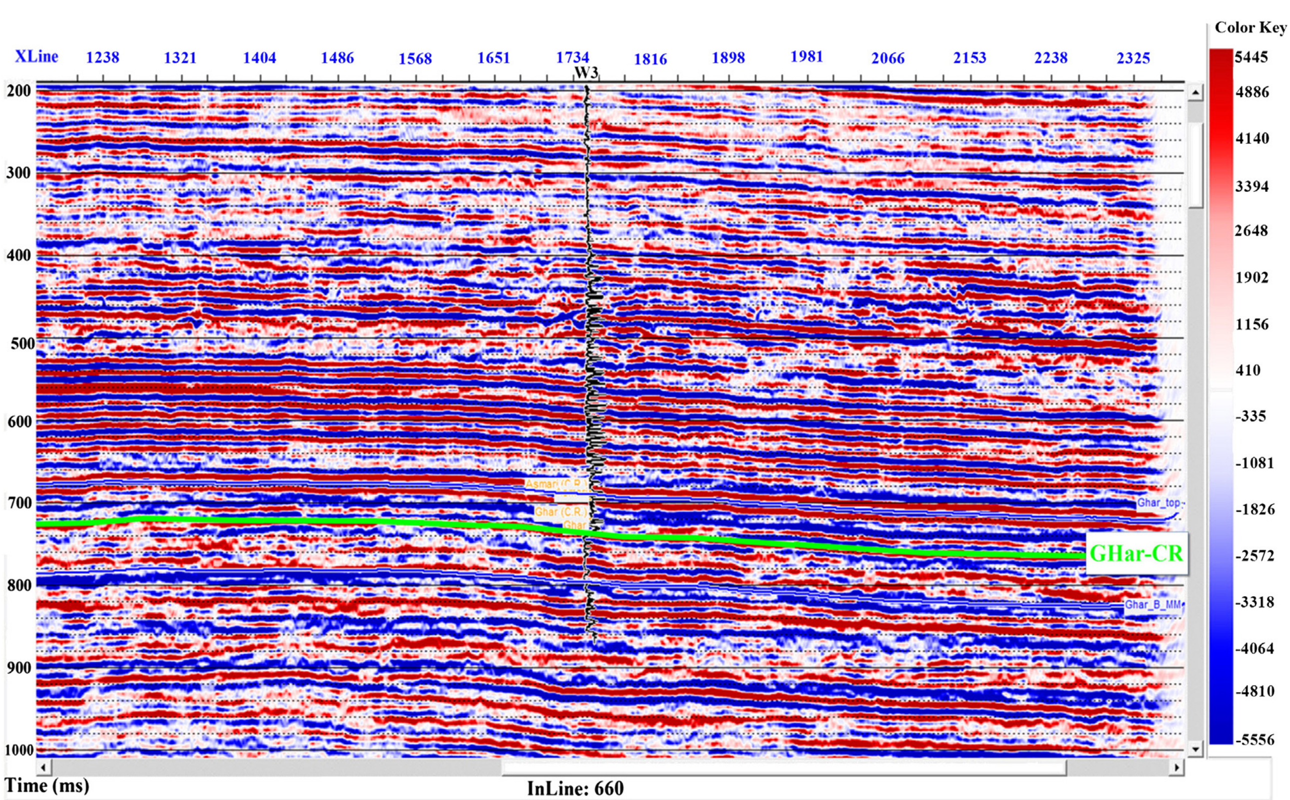

Fig. 1. The seismic display of Inline 660 containing well W3.