Fuel 232 (2018) 600–609

Contents lists available at ScienceDirect

Fuel

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

Full Length Article

Dew point pressure prediction based on mixed-kernels-function support vector machine in gas-condensate reservoir

Zhi Zhong[a,b,*], Siyan Liu[b], Mohammad Kazemi[b], Timothy R. Carr[a]

[a] Department of Geology and Geography, West Virginia University, Morgantown, WV 26506, USA
[b] Department of Chemical and Petroleum Engineering, University of Kansas, Lawrence, KS 66047, USA

ARTICLE INFO

Keywords:
Dew point pressure
Support vector machine
Mixed kernel function
Particle swarm optimization

ABSTRACT

Dew point pressure, at which the first condensate liquid comes out of solution in gas condensate reservoir, is a vital parameter for fluid characterization, field development, reservoir management and facility construction. Fast and accurate measurement of dew point pressure is always a challenge. Laboratory measurement can give accurate dew point pressure, but are expensive and time consuming. Equation of state is an alternative way, but can not converge in light oil and gas condensate reservoirs. Different empirical correlations have been built up between reservoir properties, fluid composition and dew point pressure. However, those correlations do not accurately reflect complex, non-linear relationships between them. With the development and improvement of artificial neural networks, different neural networks; such as multilayer perceptron neural network, radial basis function neural network, and gene expression programming can be used to describe complex relationships. Recently, one popular machine learning algorithm-(support vector machine) attracts attention due to its strong generalization ability. In this paper, we introduce a mixed kernel function based support vector machine (MKF-SVM), which has both strong interpolation and extrapolation abilities. This support vector machine model was trained and tested using 564 measurements of dew point pressure.The performance of this model is compared against four well known empirical correlations for dew point pressure calculation. The result, high $R^2 = 0.9150$, low root mean square error RMSE = 476.392 and low average absolute percent relative error (AAPE = 7.01%) indicates good performance of mixed kernel function based support vector machine (MKF-SVM).

1. Introduction

Gas condensate reservoirs are valuable hydrocarbon resources in terms of providing huge amounts of energy. Commonly, well deliverability decreases significantly when well bottom-hole pressure (BHP) in a gas condensate reservoir begins to drop below a certain pressure because relative permeability of the gas phase decreases due to separation of liquid and gas phase. This specific pressure is called the dew point pressure (DPP) or saturation pressure, at which the first condensate liquid comes out of solution in a gas condensate [1,2]. Dew point pressure is a critical parameter for rock-fluid characterization, gas condensate reservoir performance evaluation, pipelines and other facilities construction [3]. Research shows that gas condensation has negative impact on gas productivity and recovery [4–7]. Normally, dew point pressure in a gas condensate reservoir is equal to the initial reservoir pressure as the reservoir is saturated at initial condition. During reservoir production, condensate gas begins to separate out as reservoir pressure drops below DPP, and well-bore blockage begins. Therefore, fast and accurate determination of hydrocarbon dew point pressure is critical issue in the area of reservoir development and management in gas condensate field [8].

Two laboratory methods including constant composition expansion (CCE) and constant volume depletion (CVD) are used to measure the dew point pressure [1]. Those methods are accurate and precise but very expensive, time consuming and often not available. Equation of state (EoS) is an alternative option to calculate the DPP, but it may not converge in light oil and gas condensate reservoirs [9]. According to Olds et al. [10,11], temperature, gas-to-oil ratio, oil API gravity and intermediate molecular weight have impacts on the dew point pressure. Eilerts and Smith [12] add boiling point pressure as one more factor to predict the dew point pressure. Reamer and Sage [13], Organick and Golding [14] indicate that the hydrocarbon composition has huge impacts on the DPP because of the complexity of properties of each component. Nemeth and Kennedy [8] also add characteristics of the C_{7+} fraction and H_2S as control parameter to predict the dew point pressure.

However, those empirical relationships can not handle the highly complex non-linear relations between reservoir and fluid properties and

* Corresponding author at: Department of Geology and Geography, West Virginia University, Morgantown, WV 26506, USA.
E-mail addresses: zizhong1990@gmail.com (Z. Zhong), liusiyan@ku.edu (S. Liu), kazemiul@gmail.com (M. Kazemi), Tim.Carr@mail.wvu.edu (T.R. Carr).

https://doi.org/10.1016/j.fuel.2018.05.168
Received 9 February 2018; Received in revised form 29 May 2018; Accepted 30 May 2018

dew point pressure. Due to the rapid development and great improvement of artificial intelligence, advanced methodology, such as multilayer perceptron neural network (MLP-NN) [15], radial basis function neural network (RBF-NN) [9], neuron-fuzzy neural network (NF-NN) [16], expert system [17], gene expression programming (GEP) [18], genetic programming with orthogonal least squares algorithm (GP-OLS) [1] have been used to generate a dew point pressure prediction model. A machine learning algorithm, named support vector machine (SVM), which is based on the statistical learning theory, and its variant -least square support vector machine (LS-SVM) is used to generate non-linear relationship between different reservoir, fluid properties and dew point pressure [19,20]. However, the kernel function previously used in SVM and LS-SVM is either the global or the local kernel function, and can not have interpolation or extrapolation abilities at same time. We propose a mixed kernel function, which has good interpolation and extrapolation ability.

This paper introduces the principle of radial basis function neural network, support vector machine, kernel and mixed kernel function in Section 2. Section 3 discusses data preparation, model development and different evaluation parameters in order to evaluate the performance of various dew point pressure empirical models. The results in Section 4 indicate that mixed kernel function based support vector machine performs well when this model is applied in blind test dataset. The result from four most well-known empirical dew point pressure correlations are compared with mixed kernel function based support vector machine model to indicate the strong generalization ability of this methodology.

2. Methodology

2.1. Radial basis function neural network (RBF-NN)

Artificial neural networks are a branch of artificial intelligence, which mimics human brain information processing. Normally, it consists of three layers, which are input layer, output layer and hidden layer, as shown in Fig. 1a. The input layer is the first layer, and the number of neurons in this layer is problem dependent. The output layer is the last layer, which exports the final result calculated by hidden layer(s) (Fig. 1a). The hidden layer(s) sits between input and output layers. The number of hidden layers and hidden layer's neurons vary depending on the complexity of problem and training dataset' quality and size [21]. Generally, the number of neurons in first hidden layer should be greater than the input parameter's number to avoid information loss. Too many hidden layers will cause over-fitting, which results in the trained neural network remembering the training information and losing its generalization ability [22,23]. To avoid over-fitting and increase the model's generalization ability, a radial basis function neural network (RBF-NN) was introduced, with only three layers: input, hidden and output layer [24]. The structure of the network is shown in Fig. 1b. The number of neurons in the hidden layer determines the accuracy of RBF-NN and also computation cost [24]. The activation function of hidden layer is the radial basis function (ϕ_n), as formulated in Eq. (1).

$$\phi_n(\|x-c_i\|) = exp\left(-\frac{\|x-c_i\|^2}{2\Sigma_i^n}\right), \tag{1}$$

where, X_i is the input, b is biased terms, c_i is prototype of center of the i^{th} hidden neuron, σ_i is the bandwidth of i^{th} kernel node. $\|x-c_i\|$ denotes the Euclidean norm. This activation function calculates the closeness between input and stored prototype in that neuron.

2.2. Support vector machine (SVM)

Support vector machine (SVM), a supervised non-parametric statistical learning technique, was first introduced in 1960s [25–27]. Originally, this method is used for pattern recognition and classification problems. The basic support vector machine uses an adaptive margin-based loss function called kernel function – $K(x_i, x_j)$ to map the original linear or non-linear data from original space into higher dimensional feature space. Support vector regression (SVR) is variant of support vector machine, which is mainly used for linear/nonlinear regression and time series prediction problems [28–33]. The objective of support vector regression is to find a function $f(x)$ by which the deviations between estimated values of output and actual training output data equal to or less than a tolerance (ϵ)(Figure S_1). The SVM for regression using a kernel function and the ϵ-insensitive loss function is formulated as:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i \tag{2}$$

$$s.\ t.\ \begin{cases} y_i-\langle w\cdot\phi(x_i)\rangle-b \leqslant \epsilon + \xi_i, & i = 1, 2, \cdots, m \\ \langle w\cdot\phi(x_i)\rangle + b-y_i \geqslant \epsilon + \xi_i^*, & i = 1, 2, \cdots, m \\ \xi_i, \xi_i^* \geqslant 0, & i = 1, 2, \cdots, N, \end{cases} \tag{3}$$

The first term of Eq. (2) is the Vapnik-Chervonenkis (VC) confidence interval, whereas the second one is the empirical risk [34]. ξ and ξ^* are slack variables, which are asymmetric bounds to satisfy constraints on the loss function instead of the 'hard margin' ϵ loss function.

2.2.1. Mixed kernel function based support vector machine (MKF-SVM)

The projection function $\Phi(x)$ used in support vector machine is kernel function ($K(x_i, x_j)$), which is defined as the inner product $\langle\phi(x_i), \phi(x_j)\rangle$. Kernel function maps the original linearly or non-linearly learning data into high dimensional feature space, where all of the data can be presented linearly [35]. A kernel function must meet Mercer's conditions One corollary derived from Mercer's conditions is concluded following. Assuming $K_1(\vec{x}, \vec{x}^*)$, $K_2(\vec{x}, \vec{x}^*)$ are admissible support

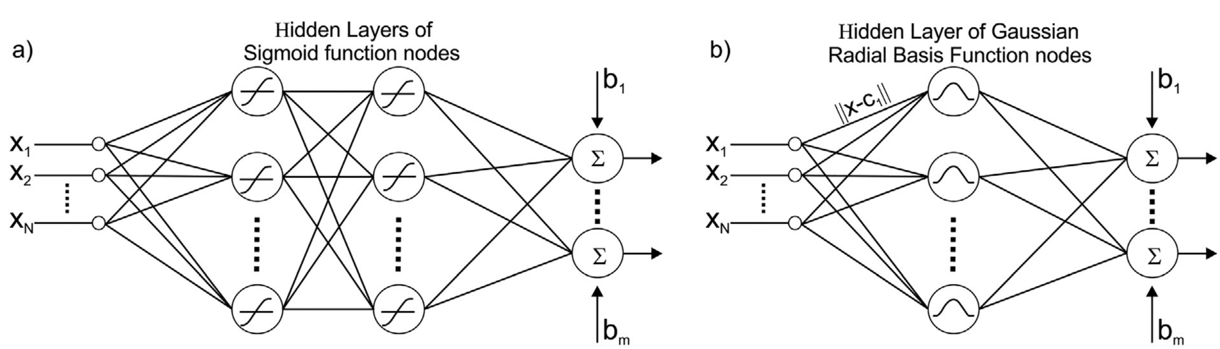

Fig. 1. (a) Is the structure of fully connected multilayer perceptron neural network with two hidden layers, the active function of each hidden layer is sigmoid function ($\phi(x_i) = 1/(1 + e^{-x_i})$); (b) is the structure of fully connected radial basis function neural network, the activation function of each hidden neurons is Gaussian radial basis function (Eq. (1)) [24].