CrossMark

ORIGINAL PAPER - PRODUCTION ENGINEERING

Support vector regression between PVT data and bubble point pressure

Parisa Bagheripour · Amin Gholami · Mojtaba Asoodeh

Abstract Accurate determination of oil bubble point pressure (Pb) from laboratory experiments is time, cost and labor intensive. Therefore, the quest for an accurate, fast, and cheap method of determining Pb is inevitable. Since support vector based regression satisfies all components of such a quest through a supervised learning algorithm plant based on statistical learning theory, it was employed to formulate available PVT data into Pb. Open-sources literature data were used for SVR model construction and Iranian Oils data were employed for model evaluation. A comparison among SVR, neural network and three well-known empirical correlations demonstrated superiority of SVR model.

Keywords Oil bubble point pressure · Support vector regression · Neural network · Empirical correlations · Petroleum chemistry

Introduction

Bubble point pressure is a crucial characteristic of reservoir fluids which is involved in most petroleum calculations such as material balance, reserve estimation, well testing, reservoir simulation, and production planning (Asoodeh and Bagheripour 2012). Laboratory experiments can accurately determine bubble point pressure (Danesh 1998). However, these measurements are highly time, cost, and labor intensive. Thus, researchers have tried to find a precise method for prediction of bubble point pressure which reduces costs and time consumption. Several empirical correlations have been published as outcome of their works (Standing 1947; Lasater 1958; Glaso 1980; Al-Marhoun 1988; McCain 1991). In last decade, some researchers utilized neural networks in solving their problems. In case of bubble point pressure estimation from PVT data, including solution gas-oil-ratio (Rs), gas specific gravity (Yg), temperature (T), and stock-tank oil gravity (γ_o); several studies showed superiority of neural network to empirical correlations (Al-Marhoun and Osman 2002; Kh.A. El-M Shokir and Sayyouh 2003; Obanijesu and Araromi 2008; Dutta and Gupta 2010). In recent years, a growing tendency is observed among researchers to utilize support vector in their regression problems. Traditional learning algorithms such as neural networks use empirical risk minimization (ERM) principle, while support vector regression exploits both structural risk minimization (SRM) and ERM principles. This feature serves more generalization capability for SVR algorithm. In this study, a support vector based regression between bubble point pressure and PVT data, including solution gas-oil-ratio (Rs), gas specific gravity (Yg), temperature (T), and stock-tank oil gravity (γ_o) was established and performance of SVR model was compared with neural network and empirical correlations. SVR modeling was performed on worldwide oil samples and was checked by Iranian Oils. Results indicated superiority of SVR model compared with other methods.

P. Bagheripour (✉)
Department of Petroleum Engineering, Gachsaran Branch, Islamic Azad University, Gachsaran, Iran
e-mail: p.bagheripour@gmail.com

A. Gholami
Abadan Faculty of Petroleum Engineering, Petroleum University of Technology, Abadan, Iran

M. Asoodeh
Islamic Azad University, Birjand Branch, Birjand, Iran
e-mail: asoodeh.mojtaba@gmail.com

Springer

Support vector regression

Support vector regression based on statistical learning theory was developed by Vapnik (1995). This method as an alternative and a powerful technique achieves worldwide growing applications to solve various nonlinear regression estimation problems in the upstream and downstream oil industry (Al-Anazi and Gates 2010; Fei and Sun 2008; A.A. Shujath Ali and Muhammadain 2012). This method has found widespread popularity among intelligence system due to its attractive feature and good generalization performance (Üstün et al. 2005). Here a brief review of SVR method is given in the subsequent paragraph. The main aim of SVR method is to find a linear function $f(x)$ which estimates the functional dependence of the dependent (output) variable $y \in R$ on an n-dimensional input variable x: (Kecman 2005).

$$f(x, w) = w^T x + b, \tag{1}$$

where w denotes the weighting vector of the linear function and b denotes the bias term. The empirical risk is given by (Kecman 2005)

$$R_{\text{emp}}^{\varepsilon}(w, b) = \frac{1}{l} \sum_{i=1}^{l} |y_i - w^T x_i - b|_{\varepsilon} \tag{2}$$

with Vapnik's linear loss function with ε-insensitivity zone defined as

$$|y - f(x, w)|_{\varepsilon} = \begin{cases} 0 & \text{if} \quad |y - f(x, w)| \leq \varepsilon \\ |y - f(x, w)| - \varepsilon & \text{Otherwise} \end{cases}. \tag{3}$$

In fact Vapnik's ε-insensitivity loss function defines a tube with radius ε around the hypothetical regression function. For value of data point located within this tube the loss function equals 0 while for other value the loss function is equal to difference between data point and tube radius (Kecman 2005).

This problem can be reformulated in a dual space by

$$\text{Maximize } L_p(\alpha_i, \alpha_i^*) = -\sum_{i,j=1}^{l} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) x_i^T x_j$$
$$- \varepsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) y_i \tag{4}$$

$$\text{Subject to } \begin{cases} \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C, i = 1, \ldots, l \\ 0 \leq \alpha_i^* \leq C, i = 1, \ldots, l \end{cases} \tag{5}$$

where $\alpha_i, \alpha_i^* \geq 0$ are positive Lagrange multipliers. C is penalty parameter defined by user. The selection of

accurate value for C is carried out by cross validation technique (Kecman 2005).

Having determined the optimal Lagrange multipliers, the linear hypersurface regression is calculated as

$$f(x, w) = w_o^T x + b = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i^T x + b \tag{6}$$

With

$$w_o = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i \tag{7}$$

Training point with Lagrange multiples between 0 and C can be used for computing bias term value in which this point is called unbounded or free support vector data point. It is much better to compute the bias term b by averaging over all the free support vector data points (Kecman 2005). Generalization to nonlinear regression estimation can be achieved through nonlinear mapping function (i.e. kernel function). The nonlinear hyperplane regression model is given by Kecman (2005).

$$f(x, w) = G(\alpha - \alpha^*) + b \tag{8}$$

Again, the bias term b is determined by averaging over all the free support vectors (Kecman 2005).

Results and discussion

In this study, ε-SVR method was used for developing a model meant to predict bubble point pressure (Pb) from PVT data, including solution gas-oil-ratio (Rs), gas specific gravity (Yg), temperature (T), and stock-tank oil gravity (γ_o). This should be borne in mind in all SVR modeling, input and output data must be normalized in range of $[-1-1]$ due to better execution of kernel functions in this range. A group of 361 worldwide oil samples from open-sources literature was gathered for model construction. This model was checked by 223 samples from Iranian Oils. Owing to fewer parameters meant to be tuned and low computational cost (Vapnik and Lerner 1963; Keerthi and Lin 2003), radial basis function (RBF) was used as kernel function in SVR modeling. Involved parameters in SVR and in embedded kernel functions (i.e., C, Gamma, and Epsilpn) determine performance of modeling. Therefore, a thorough survey was done through combination of grid search and pattern search techniques to find optimal values of "C", "Gamma", and "Epsilon". The specified search ranges for "C", "Gamma", and "Epsilon" were [0.1500000], [0.00000120], and [0.0001100], respectively, while the extracted optimal points from these ranges are 178,366.43619, 0.228349, and 0.020427, correspondingly. The number of support
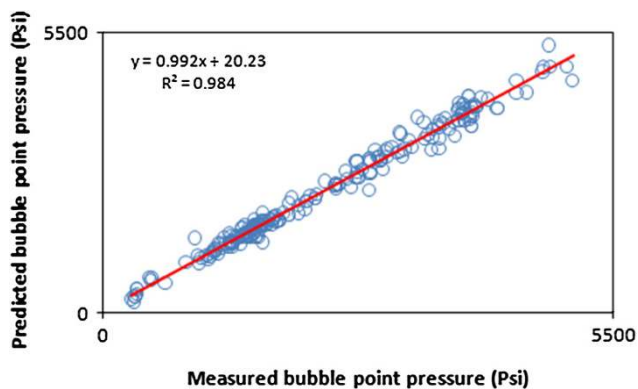
Fig. 1 Cross plot showing correlation coefficient between measured and SVR predicted bubble point pressure. High value of correlation coefficient, i.e. 0.992 (R−square $= 0.984$) indicates success of SVR approach in estimation of bubble point pressure
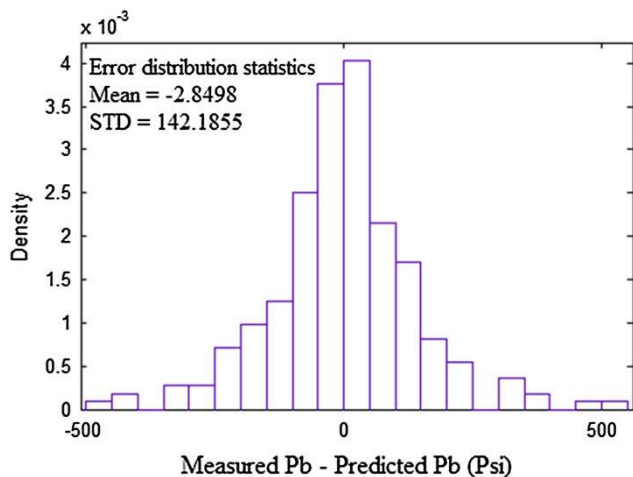

Fig. 2 Error distribution statistics for SVR model meant to predict bubble point pressure. Small values of mean and standard deviation (STD) reveal high performance of SVR modeling. Error distribution indicates 68 % of predicted values have errors in range of -2.8498 ± 142.1855

vectors used by this model is 70. After model construction, a group of 223 data points originated from Iranian Oils was used to assess performance of SVR model. Figure 1 indicates the cross correlation between measured and predicted bubble point pressure using the concept of correlation coefficient. High value of correlation coefficient ($R = 0.992$) confirms success of SVR modeling. For a deeper analysis of SVR performance, Fig. 2 provides the opportunity for visually and statistically observing the precision of prediction and effect of overestimation/underestimation. Error distribution of bubble point prediction has mean and standard deviation of -2.8498 and 142.1855, respectively. These small values once again verify robustness of SVR modeling. Figure 3 indicates relative error of prediction for different values of bubble
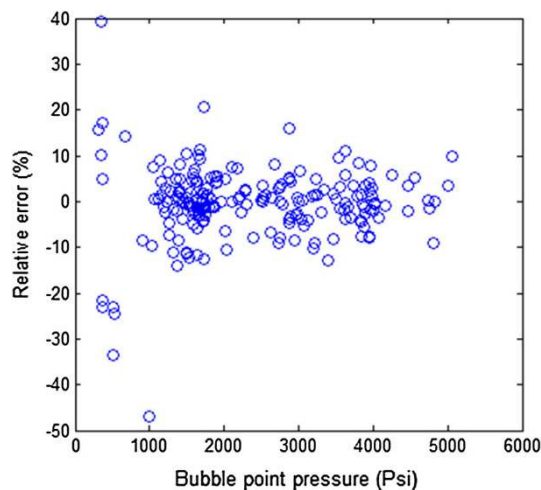

Fig. 3 Graph showing relative error of SVR predictions versus bubble point pressure values. Relative error for small values of bubble point pressure (i.e. <1,000 Psi) is higher
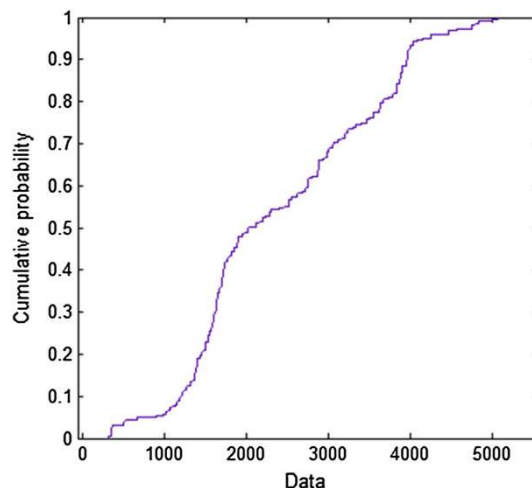

Fig. 4 Graph showing cumulative probability of bubble point pressure data used in training set. This figure illustrates just small portion of data points are correspondence to bubble point pressure values less than 1,000 Psi. Therefore, SVR modeling was not successful in this region and relative error increases in small bubble point pressure values

Table 1 Comprising SVR model with neural network and three well-known empirical correlations based on correlation coefficient (R), average relative error (ARE), absolute average relative error (AARE), and root mean square error (RMSE)

Method	R	ARE	AARE	RMSE
SVR	0.992	−2.8498	101.2773	141.8949
NN	0.97148	8.0889	271.4315	359.1266
Al-Shammasi	0.93655	−636.9246	664.2424	898.3148
McCain	0.93701	30.5056	382.8586	591.3628
Velarde et al.	0.92942	209.3286	452.4348	602.5413

Fig. 5 A comparison between
measured values of bubble point
pressure and SVR predicted
values. This figure confirms
there is a good agreement
between SVR predicted values
and reality
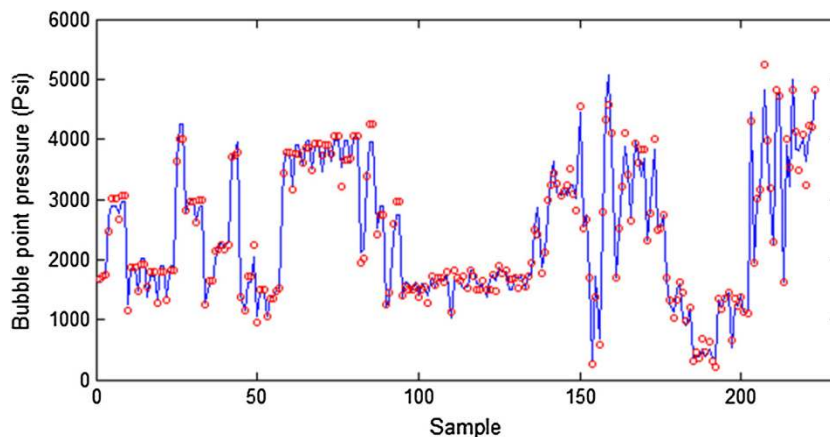
point pressure. This figure indicates for small values, SVR method performed weaker, while in large values the accuracy of prediction is satisfying. To explain this weakness, cumulative probability of bubble point pressure is helpful (Fig. 4). Performance of SVR decreases for bubble point pressures less than 1,000 Psi (Fig. 3). A careful insight into Fig. 4 reveals only small portion of data points belongs to region in which bubble point pressure is less than 1,000 Psi. Insufficient number of data points in this region causes the mentioned weakness. It should be borne in mind number of training data, which represents the underlying dependency between input/output data space is a key parameter in constructing intelligent models. At the last stage of this study, SVR model was compared with neural network and three well-known empirical correlations, including Al-Shammasi (1999), McCain (1991), and Velarde et al. (1999). Comparison results using concepts of correlation coefficient (R), average relative error (ARE), absolute average relative error (AARE), and root mean square error (RMSE) are shown in Table 1. Results indicate SVR model outperformed other methods making a judgment based on R, ARE, AARE, and RMSE criteria. Finally, a comparison between measured values of bubble point pressure and SVR predicted values is illustrated in Fig. 5 which shows there is a good agreement between SVR predicted values and reality.

Conclusions

Owing to significance of bubble point pressure (Pb), different empirical correlations and neural network models have been presented up to now. However, the competition is run for finding more accurate methods. This study proposed an improved method for estimating Pb from PVT

data using support vector regression algorithm and following conclusions were deduced from its results:

(a) Support vector regression provides a fast, cheap, and accurate way of determining bubble point pressure from PVT data.
(b) SVR model was compared with neural network and empirical correlations. Results of comparison revealed superiority of SVR model to others.
(c) Comparison between neural network and empirical correlations indicates neural network outperformed empirical correlations. Therefore, intelligent systems produce more reliable results.
(d) Structural risk minimization in conjugation with empirical risk minimization fortifies SVR model as robust tool for extracting the quantitative formulation between a set of input/output data which have an underlying dependency.
(e) Implementation of proposed strategy eliminates the need for tedious laboratory experiments and saves time and money.

References

Al-Marhoun MA, Nizamuddin S, Abdul Raheem AA, Shujath Ali S, Muhammadain AA (2012) Prediction of crude oil viscosity curve using artificial intelligence techniques. J Petrol Sci Eng 86:111–117
Al-Anazi AF, Gates ID (2010) Support vector regression to predict porosity and permeability: effect of sample size. Comput Geosci 36:1494–1503
Al-Marhoun MA (1988) PVT correlations for Middle East crude oils. J Petrol Technol 40(5):650–666

Al-Marhoun MA, Osman EA (2002) Using artificial neural networks to develop new PVT correlations for saudi crude oils. Abu Dhabi international petroleum exhibition and conference, Abu Dhabi, United Arab Emirates, 13−16 October

Al-Shammasi AA (1999) Bubble point pressure and oil formation volume factor correlations. Middle east oil show and conference, Bahrain, 20−23 February

Asoodeh M, Bagheripour P (2012) Estimation of bubble point pressure from *PVT* data using a power-law committee with intelligent systems. J Petrol Sci Eng 90–91:1–11

Danesh A (1998) PVT and phase behavior of petroleum reservoir fluids. Elsevier, Amsterdam, Netherlands

Dutta S, Gupta JP (2010) PVT correlations for Indian crude using artificial neural networks. J Petrol Sci Eng 72:93–109

Fei SW, Sun Y (2008) Forecasting dissolved gases content in power transformer oil based on support vector machine with genetic algorithm. Electr Power Syst Res 78:507–514

Glaso O (1980) Generalized pressure-volume-temperature correlations. J Petrol Technol 32(5):785–795

Goda HM, Fattah Kh A, El-M Shokir EM, Sayyouh MH (2003) Prediction of the PVT data using neural network computing theory. Presented at the Nigeria annual international conference and exhibition, Abuja, Nigeria, 4−6 August

Kecman V (2005) Support vector machines: an introduction. In: Wang L (ed) Support vector machines: theory and applications, vol 1. Springer, Berlin Heidelberg, pp 1–47

Keerthi SS, Lin CJ (2003) Asymptotic behaviors of support vector machines with Gaussian kernel. Neural Comput 15:1667–1689

Lasater JA (1958) Bubble point pressure correlation. Trans AIME 213:379–381

McCain WD (1991) Reservoir·fluid property correlations-state of the art. SPE Reserv Eng 6(2):266–272

Obanijesu EO, Araromi DO (2008) Predicting bubble-point pressure and formation-volume factor of Nigerian crude oil system for environmental sustainability. J Pet Sci Technol 26(17):1993–2008

Standing MB (1947) A Pressure–volume-temperature correlation for mixture of California oils and gases. American petroleum institute

Üstön B, Melssen WJ, Oudenhuijzen M, Buydens LMC (2005) Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. Anal Chim Acta 544:292–305

Vapnik V (1995) The nature of statistical learning theory. Springer, New York

Vapnik V, Lerner A (1963) Pattern recognition using generalized portrait method. Automation and Remote Control 24:774–780

Velarde J, Blasingame TA, McCain WD (1999) Correlation of black oil properties at pressures below bubble point pressure-A new approach. J Can Pet Technol 38(13):62–68