Journal of Chemical and Petroleum Engineering 2019, 53(2): 177-189
DOI: 10.22059/jchpe.2019.266793.1251

RESEARCH PAPER

Bubble Pressure Prediction of Reservoir Fluids using Artificial Neural Network and Support Vector Machine

Afshin Dehghani Kiadehi, Bahman Mehdizadeh, Kamyar Movagharnejad*

Faculty of Chemical Engineering, Babol Noshirvani University of Technology, Babol, Iran

Received: 7 October 2018, Revised: 19 August 2019, Accepted: 2 September 2019

Abstract

Bubble point pressure is an important parameter in equilibrium calculations of reservoir fluids and having other applications in reservoir engineering. In this work, an artificial neural network (ANN) and a least square support vector machine (LS-SVM) have been used to predict the bubble point pressure of reservoir fluids. Also, the accuracy of the models have been compared to two-equation state-based models, i.e. SRK-EOS and PR-EOS and four empirical equations, i.e. Whitson, Standing, Wilson and Ghafoori et al. Compared to the experimental data, the average relative deviations (ARD) of bubble pressure prediction for these equations were obtained to be 14%, 29%, 66%, 30%, 38%, and 11%, respectively. The best semi-empirical equation has an ARD of about 11% while, the ANN and LS-SVM models have an ARD of 8% and 4.68%, respectively. Thus, it can be concluded that generally, these soft computing models appear to be more accurate than the empirical and EOS based methods for prediction of bubble point pressure of reservoir fluids.

Keywords:
Artificial Neural Network,
Bubble Pressure,
Empirical Correlations,
Genetic Algorithm,
Reservoir Fluids,
Support Vector Machine,

Introduction

One of the most important parameters in equilibrium calculations of reservoir fluids is bubble point pressure. Bubble point pressure values are widely used in reservoir engineering calculations such as reservoir simulation, separator condition design, obtaining the optimum production rate, future performance prediction and material balance calculations of reservoir fluids. The accuracy of these calculations strongly depends on the accuracy of bubble point pressure measurements. Bubble point pressure is defined as the pressure at which the first gas bubble comes out of liquid phase at a constant temperature. The most important method used to calculate bubble point pressure is using the K-values predicted based on an empirical relation or equation of states. K-value is defined by the following equation [1]:

$$K_i = \frac{y_i}{x_i} \tag{1}$$

where, y_i and x_i are the mole fractions of component i in the vapor and liquid phases, respectively. Although pressure-volume-temperature (PVT) experiments for K values calculations provide reliable results, they are very time-consuming and expensive. Thus, many researchers try to find fast and accurate ways for forecasting the bubble point pressure such as the development of different empirical relations for K-values prediction [2-4]. Although empirical equations are rapid paths for K-values calculations, the coefficients of these

* Corresponding author
Email: k-movaghar@nit.ac.ir (K. Movagharnejad)

equations usually obtained from the least square method using experimental data and are specific for each component. In the present work, an artificial neural network (ANN) and least square support vector machines (LS-SVMs) has been used for bubble pressure prediction of reservoir fluids. The result showed the average relative deviations (ARD) of these methods are less than classic methods.

Theory

Empirical Correlations

Wilson [5] developed a correlation to estimate the K-value as a function of critical pressure, temperature, and acentric factor as follows:

$$K_i = \frac{P_{ci}}{P} exp\,[5.37(1+\omega_i)(1-T_{Ri}^{-1})] \quad (2)$$

where P_{ci} is the critical pressure of component i, T_{Ri} is the reduced temperature of component i and P is the system pressure.

Standing [6] proposed the following relation for the calculation of K-value:

$$lnK_iP = a + c\left[b_i\left(\frac{1}{1.8}\right)\left(\frac{1}{T_{bi}} - \frac{1}{T}\right)\right] \quad (3)$$

where a and c are functions of pressure and b_i is defined as follows:

$$b_i = log\left(\frac{P_{ci}}{101353.5}\right)/[\left(\frac{1}{1.8}\right)\left(\frac{1}{T_{bi}} - \frac{1}{T_{ci}}\right)] \quad (4)$$

where, P_{ci}, T_{ci}, and T_{bi} are critical pressure, critical temperature and normal boiling point of component i, respectively.

Whitson and Torp [7] presented a new equation using the convergence pressure (P_k):

$$K_i = \left(\frac{P_{ci}}{P_k}\right)^{\alpha-1}\frac{P_{ci}}{P} exp[5.37\alpha(1+\omega_i)(1-T_{Ri}^{-1})] \quad (5)$$

where P_k is the particular pressure that K-values of all components in a reservoir fluid converges to unity at this pressure. Also, α expressed as follows:

$$\alpha = 1 - \left(\frac{P_k}{P}\right)^{0.7} \quad (6)$$

In low pressures, α approaches to unity and Whitson equation (Eq. 5) is reduced to the Wilson equation (Eq. 2). Ghafoori et al. [8] introduced a new equation, which was a function of more common properties such as critical pressure, temperature, and acentric factor as follows:

$$K_i = \left(\frac{1}{P_{ri}}\right) exp[5.37\beta(1+\omega_i)(1-T_{Ri}^{-1})] \quad (7)$$

where P_{ri} is the reduced pressure of component i, and β is obtained from the following equation:

$$\beta = 1 - \left(\frac{P}{P_k}\right)^{T_{Rmix}} \quad (8)$$

where T_{Rmix} is the reduced temperature of the mixture:

$$T_{Rmix} = \frac{T}{T_{cmix}} \tag{9}$$

where T_{Cmix} is the critical temperature of the mixture and obtained from the following mixing rule:

$$T_{Cmix} = \sum Z_i\, T_{Ci} \tag{10}$$

where T_{Ci} and Z_i are the critical temperature and mole fraction of component i.

Least Square Support Vector Machine

Support Vector Machines (SVMs) are a strong kernel based on the statistical learning theory (SLT) and the structural risk minimization (SRM) principle introduced by Vapnik [9]. SVMs recognize patterns, estimate function and also, solve the nonlinear problems via solving the quadratic programming (QP) [10]. LS-SVMs are an alternative formulation of the standard SVMs, which find the solution by solving a set of linear equations instead of a QP problem [11].

The standard LS-SVM algorithm has been described as follows; Given a set of training data like this:

$$\{(x_1, y_1)...(x_k, y_k)\} \subset \mathbb{R}^N \times \mathbb{R} \tag{11}$$

The following regression model is constructed by using a nonlinear mapping function $\varphi(x)$, which maps the input data to a higher dimensional feature space:

$$y = w^T.\varphi(x) + b \quad with\ w \in \mathbb{R}^N,\ b \in \mathbb{R},\ \varphi{:}\ \ \mathbb{R}^N \to\ \ \mathbb{R}^M, M \to \infty \tag{12}$$

where w is the weight vector and b is the bias. When the least squares support vector is used as an approximation function, a new optimization problem is created in the case of SRM. The quadratic loss function is selected in LS-SVM. The optimization problem of LS-SVM is created as:

$$Min\ J\,(w, e) = \frac{1}{2} w^T.w + \frac{1}{2}\gamma \sum_{k=1}^{N} e_k^2 \tag{13}$$

The constraint of these equations is:

$$y = w^T.\varphi(x), + b + e_k \quad k=1....N \tag{14}$$

γ is the regularization parameter that balances the model's complexity and the training error, and e_k is the desired error. In order to solve the constrained optimization problem, a Lagrangian is constructed as:

$$L\,(w, b, e, \propto) = J\,(w, e) - \sum_{k=1}^{N} \propto_k \{w^T.\varphi(x), + b + e_k - y_k\} \tag{15}$$

In this equation $\propto_k$ is Lagrange multipliers and called support value. The solution of the above equation can be obtained by partially differentiating with respect to each variable.

$$\frac{\partial L}{\partial w} = 0 \to w = \sum_{k=1}^{N} \propto_k.\varphi(x_k) \tag{16}$$

$$\frac{\partial L}{\partial b}=0 \rightarrow w=\sum_{k=1}^{N} \propto_k=0 \quad (17)$$

$$\frac{\partial L}{\partial e_k}=0 \rightarrow \propto_k = \gamma . e_k \qquad k=1,\dots, N \quad (18)$$

$$\frac{\partial L}{\partial \propto_k}=0 \rightarrow \ w^T . \varphi(x), + b + e_k - y_k = 0 \qquad k=1,\dots,N \quad (19)$$

When the variable *w* and *e* is removed, the Karush-Kuhn-Tucker (KKT) system is obtained as:

$$\begin{bmatrix} 0_{1\times N} & 1_{1\times N} \\ 1_{N\times 1} & Z+\gamma^{-1}.I \end{bmatrix} \begin{bmatrix} b \\ \propto \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (20)$$

With

$$y= [y_1, \dots, y_N] \quad (21)$$

$$1= [1, \dots,1] \quad (22)$$

$$0= [0, \dots,0] \quad (23)$$

$$\propto=[\propto_1, \dots, \propto_N] \quad (24)$$

$$Z=\{Z_{kj}|k,j = 1, \dots, N\}\ , Z_{kj} = \varphi(x_k)^{\,T}.\varphi(x_j) = K(x_k,x_j)\ j = 1,\dots,N \quad (25)$$

In the above equation $K(x_k, x_j)$ is the kernel function and must follow Mercer's theory [12]. The common examples of kernel function are linear, polynomial, radial basis function (RBF) kernel and multi-layer perceptron (MLP). In the present work, the RBF kernel was selected as the kernel function (Eq. 26).

$$K(x, x_k) = exp\,(-\frac{\| x - x_k \|^2}{\delta^2}) \quad (26)$$

The LS-SVM regression model can be obtained as:

$$y(x) =\sum_{k=1}^{N} \propto_k . K(x, x_k) + b \quad (27)$$

where (b, α) is the solution to Eq.16. The general topology of the LS-SVM model is presented in Fig. 1.

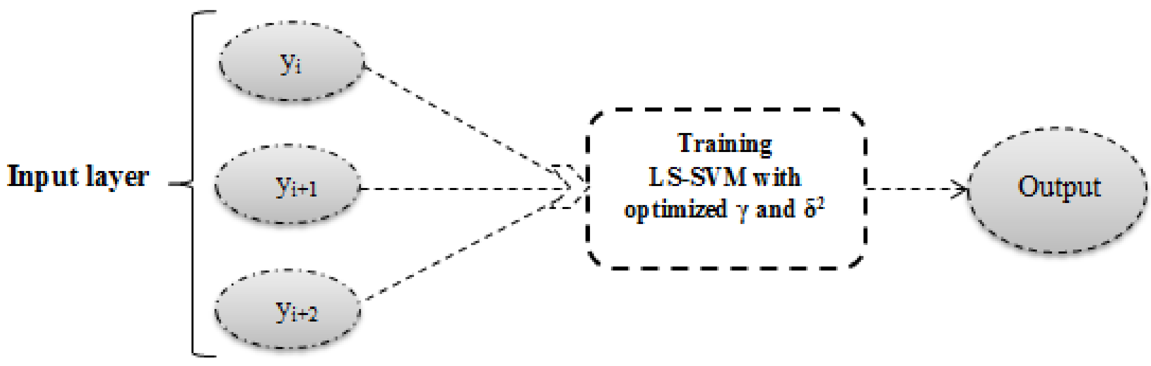

Fig. 1. The general topology of the LS-SVM model

Genetic algorithm

Genetic Algorithms (GA), to obtain a fast search and optimization technique, use the "survival of the fittest" principle of natural evolution with the genetic propagation of characteristics [13]. The most important aspect of a GA is that it determines many possible solutions simultaneously and explores different regions in the desired space chosen by the user [14]. GA, uses a direct analogy to Darwinian natural selection and genetics in biological systems, is a promising alternative to conventional traditional methods. Based on the Darwinian principle of 'survival of the fittest', GA can obtain the optimal solution after a series of iterative computations. The search process is composed of artificial mutation, crossover, and selection [15]. The adjusting processes of GA include three steps:

1. **Chromosome design**: in this step γ and δ^2 coded to form the chromosome. The chromosome X was presented as X= $\{p_1, p_2\}$ where p_1 and p_2 are γ and δ^2 respectively in this work.
2. **Population generation**: in this step randomly initialized a population of possible solutions is generated.
3. **Fitness study**: in this step a fitness function is evaluated. In the present work, ARD of testing data was used as a fitness function. Steps of the GA learning algorithm are detailed in the literature [12].

These three steps generate a new population of possible solutions, which as compared to the previous population; usually lead to better at satisfying the optimization objective. The best-obtained string after repeating the above-described loop forms the solution to the optimization problem [16,17]. Fig. 2 depicts the steps of a GA for tuning the parameters of our proposed LS-SVM methods, which is defined above.

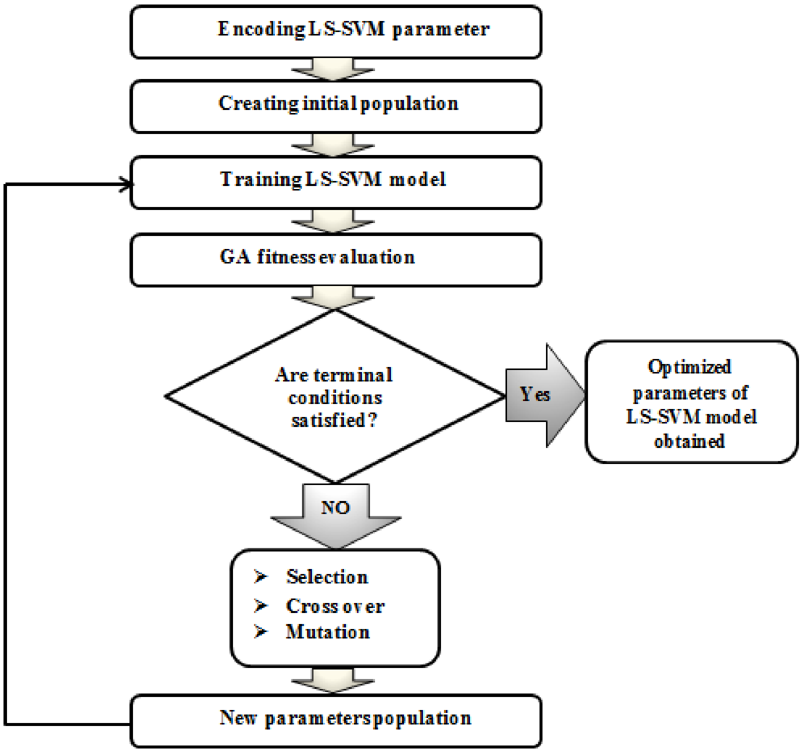

Fig. 2. The overall procedure of tuning the parameters of LS-SVM with GA

Artificial neural network

Neurons are the main building blocks of neural networks. In an ANN a neuron sums the weighted inputs from several connections and then the output of neurons is produced by applying transfer function to the sum. There are many transfer functions but, the common transfer function is sigmoid and we used this transfer function. The sigmoid function can be expressed by the following equation:

$$\theta_j = \frac{1}{1 + e^{-\psi_j}} \tag{28}$$

In Eq. 28 ψ is the sum of weighted inputs to each neuron and θ is the output of each neuron and ψ can be calculated from Eq. 29.

$$\psi_j = \left(\sum_{i=1}^{n} w_{ij} \, . \, \theta_i \right) + b_j \tag{29}$$

In Eq. 29, w_{ij} denotes a connection between node *j* of interlayer *l* to node *i* of interlayer *l*-1, b_j is a bias term and *n* is the number of neurons in each layer. In any interlayer *l* and neuron *j*, input values integrate and generate ψ_j.

In order to minimize the difference between experimental data and the calculation of the neural network, the mentioned process repeats for the total number of training data. After training, validation of the neural network can be done by testing data.

Numerous types of artificial neural networks exist such as MLP, RBF networks and recurrent neural networks (RNN). The type of network used in this work is the MLP network. MLP networks are one of the most popular and successful neural network architectures, which are suited to a wide range of applications such as prediction and process modeling [12,15,18].

Preparation Of Training Dataset

Based on the results of published literature, the bubble pressure of reservoir fluids strongly depends on mole fractions of reservoir fluids components, the molecular weight of C_{7+}, specific gravity of C_{7+} and temperature [9]. In this work, all data was divided into two parts (70% for training and 30% for testing).

To prevent a larger number from overriding a smaller number, all data were normalized. Normalization can be done by several equations. In this work, data was scaled between [0.1-0.9] by means of Eq. 30.

$$(Scaled)_{value} = \frac{(Actual)_{value} - min_{(Actual\ value)}}{max_{(Actual\ value)} - min_{(Actual\ value)}} * 0.8 + 0.1 \tag{30}$$

Table 1 introduces systems that are used in this work in order of bubble pressure prediction of reservoir fluids.

Optimization Of LS-SVM Based On GA

Accurate parameter (γ, δ^2) setting plays a significant role in obtaining a proper LS-SVM regression model with high prediction accuracy. In present work, total available data are divided into two parts: training data (70% of data) and testing data (30 % of data) randomly and *ARD* of testing data is calculated by means of Eq. 31.

$$ARD = \frac{100}{N} \times \sum_{i=1}^{N} \left| \frac{y^{exp} - y^{cal}}{y^{exp}} \right| \tag{31}$$

RBF as a kernel function was used for LS-SVM. The objective is the minimization of *ARD* on the testing dataset.

Table 1. The experimental composition and bubble pressure of reservoir fluid using in this article

No.	N_2	CO_2	H_2S	C_1	C_2	C_3	C_4	C_5	C_6	C_{7+}	γ_{C7+}	M_{WC7+}	T	P	Ref.
1	0.24	0.27	0	66.83	8.28	5.15	3.31	2.04	1.85	12.03	0.8	182	375	33163988	[22]
2	0.3	0.9	0	53.47	11.46	8.79	4.56	2.09	1.51	12.69	0.864	143	353	30750808	[23]
3	0.3	0.9	0	53.47	11.46	8.79	4.56	2.09	1.51	16.92	0.836	173	353	30750808	[23]
4	0.3	0.9	0	53.47	11.46	8.79	4.56	2.09	1.51	16.92	0.8364	173	353	30750808	[23]
5	1.02	0.4	0	54.62	11.47	7.33	4.01	1.96	1.42	17.77	0.808	218	347	26889720	[22]
6	0	0	0	47.47	6.51	4.89	6.61	6.87	7.59	20.06	0.8259	181	394	21994412	[24]
7	0	0	0	60.88	7.38	5.03	2.78	1.96	1.84	20.13	0.836	191	361	34474000	[23]
8	0.38	7.03	0	48.73	8.93	5.48	4.05	3	2.14	20.26	0.805	181	427	28654789	[25]
9	0.35	3.14	0	54.26	8.57	5.72	3.21	1.95	1.53	21.27	0.823	271	388	32453824	[26]
10	0.25	3.6	2.32	47.64	6.5	4.5	4.035	3.3	2.74	25.12	0.818	279	394	27627464	[26]
11	0.9	1.49	0	51.54	6.57	4.83	3.07	2.38	2.17	27.05	0.833	265	364	31481657	[27]
12	0.44	0.38	0	49.1	7.6	6.13	3.84	2.52	2	28	0.836	231	366	25779657	[28]
13	0.06	1.1	0	43.13	10.46	7.24	4.57	2.82	2.34	28.29	0.884	214	389	22187466	[22]
14	0.06	0.94	0	44.44	10.76	6.18	4.04	2.63	2.57	28.38	0.846	195	389	25028124	[19]
15	0.05	0.85	0	41.05	11.07	7.07	4.82	2.89	3.19	29.01	0.848	198	389	22994158	[22]
16	0.03	1.04	0.03	41.88	10.6	7.01	4.29	3.16	2.89	29.07	0.848	200	386	22297783	[22]
17	0.03	0.97	0	41.64	10.82	7.52	4.5	3	2.43	29.08	0.85	208	389	21160141	[22]
18	0.45	1.64	0	45.85	7.15	6.74	3.95	2.68	1.28	30.26	0.826	286	387	25903764	[26]
19	0.55	1.03	0	36.47	9.33	8.85	6	3.78	3.56	30.43	0.837	200	386	18933121	[23]
20	0.55	1.03	0	36.47	9.93	8.85	6	3.78	3.56	30.43	0.837	200	386	18933121	[23]
21	0.04	0.78	0	40.91	10.72	6.64	4.44	2.62	2.51	31.34	0.845	202	389	21511776	[22]
22	0.06	0.85	0	40.7	10.54	6.55	4.41	2.72	2.81	31.36	0.847	195	391	21925464	[22]
23	0.41	0.44	0	40.48	7.74	8.2	5.45	3.64	2.83	31.42	0.845	210	371	19898393	[28]
24	0.34	0.84	0	49.23	6.32	4.46	3.04	2.26	2.06	31.45	0.865	230	367	27448199	[28]
25	0.01	0.89	0.3	39.68	10.37	7.18	4.47	3.09	2.75	31.53	0.932	197	391	21104983	[22]
26	0	9.11	0	45.58	5.11	3.03	2.18	1.92	1.54	31.53	0.894	270	361	26476032	[22]
27	0	0.01	0	31	10.41	11.87	7.32	4.41	2.55	32.43	0.743	199	328	11169576	[22]
28	0.9	0.16	0	47.12	5.97	4.62	3.49	2.55	2.19	33	0.85	217	348	23400951	[28]
29	0.02	0.12	0	1.51	10.64	23.23	15.95	8.76	6.57	33.2	0.8367	251	342	1723700	[29]
30	0.16	0.91	0	36.47	9.67	6.95	5.37	2.85	4.33	33.29	0.852	218	378	18064376	[27]
31	0.02	0.99	0.04	38.79	10.07	6.86	4.18	2.92	2.77	33.38	0.849	210	389	20815401	[22]
32	0.33	3.03	0	41.33	8.89	5.95	4.12	2.66	0	33.69	0.848	200	367	22063360	[22]
33	0.31	0.69	0	47.69	6.64	4.59	2.59	1.16	1.69	34.64	0.869	234	384	26200240	[28]
34	0.11	2.35	0	35.21	6.72	6.24	5.07	5.23	4.1	34.97	0.841	213	394	17561056	[22]
35	0.11	2.35	0	35.21	6.72	6.24	5.07	3.8	4.1	34.97	0.841	213	394	17561056	[22]
36	0.67	2.11	0	34.93	7	7.82	5.48	5.018	3.04	35.15	0.855	230	388	18781435	[28]
37	0.08	1.82	0	32.17	7.627	7.221	6.507	3.82	4.08	35.47	0.812	279	393	15506405	[26]
38	1.64	0.08	0	28.4	7.16	10.48	8.4	3.82	4.05	35.97	0.843	252	328	11679791	[23]
39	1.64	0.08	0	28.4	7.16	10.48	8.4	3.82	4.05	35.97	0.843	252	328	11776318	[6]
40	1.64	0.08	0	28.4	7.16	10.48	8.4	1.6	4.05	35.97	0.843	252	328	11679791	[23]
41	0.56	3.55	0	45.34	5.48	3.7	2.35	1.6	1.33	36.12	0.836	253	366	26786298	[27]

42	0.56	3.55	0	45.34	5.48	3.7	2.35	4.72	1.33	36.12	0.817	255	366	26786298	[28]
43	0	2.13	0	31.28	7.51	6.93	6.26	1.2	4.37	36.8	0.82	270	393	15596038	[26]
44	0	0	0	52	3.81	2.37	1.72	1.2	2.01	36.84	0.841	199	367	26393294	[2]
45	0	0	0	52	3.81	2.37	1.72	1	2.06	36.84	0.841	199	367	26400189	[2]
46	0	0	0	42	8	6	2	2.8	4	37	0.823	255	342	22408100	[6]
47	0.13	1.45	0.75	36.02	8.83	6.13	4.1	3.47	2.49	37.3	0.896	230	348	19808760	[22]
48	0.49	0.47	0	31.55	8.66	7.92	5.48	4.38	4.45	37.51	0.904	289	329	11824582	[22]
49	0.06	5.01	2.67	23.03	8.49	8.26	5.39	0.89	3.89	38.82	0.877	254	366	11934899	[22]
50	0.36	1.06	0	50.5	4.54	0.9	1.15	4.81	1.6	39	0.901	291	342	25441812	[28]
51	0.45	0.08	0	29.35	7.85	7.28	6.03	1.5	4.53	39.62	0.863	227	330	11755634	[22]
52	0.08	1.37	0.69	35.97	8.67	5.94	2.97	3.51	3.1	39.71	0.898	274	349	17271474	[22]
53	0.19	2.51	0	31.89	7.51	6.59	4.77	3.48	3.29	39.74	0.92	324	330	12583010	[22]
54	0.15	0.19	0	31.15	8.92	7.38	4.07	4.03	4.34	40.32	0.848	221	331	11376420	[22]
55	1.67	1.38	0	26.68	9.25	10.18	6.4	4.62	0	40.41	0.855	217	372	13472439	[22]
56	0.41	0.65	0	30.21	6.89	6.6	5.88	2.62	4.27	40.47	0.875	247	331	12962224	[22]
57	0	1.12	0.21	26.95	10.76	9.23	4.97	5.93	3.33	40.81	0.876	249	371	12941540	[22]
58	1.13	0.13	0	25.45	6.99	9.6	7.87	3.76	1.88	41.02	0.851	237	343	9969881	[25]
59	0.45	0.51	0	30.56	7.25	7.28	5.24	2.92	3.85	41.1	0.892	268	331	11962478	[22]
60	0.1	1.16	0	32.93	7.78	6.06	3.73	3.54	3.7	41.62	0.879	279	329	12445114	[22]
61	0.18	2.46	0	30.26	7.3	6.36	4.61	3.42	3.22	42.08	0.914	299	330	12583010	[22]
62	0.29	0.48	0	28.39	8.29	7.38	5.06	3.71	4.41	42.31	0.88	252	329	11252314	[22]
63	0.36	0.17	0	27.23	8.93	8.6	6.07	4.442	2.55	42.38	0.879	271	328	9411402	[22]
64	0.45	2.07	0.38	26.58	7.894	6.73	5.384	2.2	3.35	42.72	0.865	207	394	14775556	[26]
65	0.48	0.41	0	33.25	7.55	6.21	4.11	2.84	2.88	42.91	0.88	236	331	11307472	[22]
66	0.56	0.07	0	29.9	7.84	6.98	6.05	1.62	2.66	43.1	0.877	242	329	11397104	[22]
67	0	1.25	0	33.35	9.24	6.07	2.4	3.37	2.78	43.29	0.851	252	383	14548028	[22]
68	0.2	0.49	0	29.46	8.51	6.63	3.97	5.85	3.79	43.29	0.9	145	331	10342200	[22]
69	0.18	0.82	0	22.92	7.21	7.37	6.79	1	4.84	44	0.847	180	374	9818195	[26]
70	0.21	0.72	0	33.12	7.44	6.07	3.22	1.79	3.05	44.38	0.9	271	333	11928004	[22]
71	0.15	0.48	0	28.51	8.17	6.36	4.15	3.73	3.87	44.59	0.897	141	331	10342200	[22]
72	0	0.17	0	28.56	8.63	7.43	3.81	2.61	4.11	44.68	0.862	249	330	10673150	[22]
73	0.14	0.54	0	29.44	7.96	6.19	3.96	3.25	3.84	44.68	0.901	290	330	10893784	[22]
74	0.19	0.12	0	28.62	8.66	6.52	4.78	2.71	3.62	44.78	0.925	274	331	12176217	[22]
75	0.06	0.12	0	23.71	7.18	8.71	6.85	5.17	2.85	44.81	0.868	238	342	9556193	[25]
76	0.03	0.3	0	27.8	8.26	6.73	3.72	4.08	4.2	44.88	0.88	254	330	11341946	[22]
77	0.2	0.8	0	31.42	7.6	7.23	4.88	1.97	0.99	44.91	0.889	251	329	11983162	[22]
78	0.4	1	0	45.4	4.2	0.89	1.08	0.94	1.01	45.08	0.888	250	344	23897377	[28]
79	0.44	0.83	0	27.75	7.54	6.9	4.61	3.1	3.11	45.72	0.902	272	329	10790362	[22]
80	0.2	1.34	0	23.64	8.56	6.68	5.3	4.45	4.03	45.8	0.872	327	373	11928004	[26]
81	0.33	0.22	0	25.56	6.87	6.39	5.61	4.68	4.33	46.01	0.878	222	329	10997206	[22]
82	0.68	1.02	0.04	25.49	6.37	6.53	5.15	4.88	3.81	46.03	0.902	264	330	10569728	[22]
83	0.53	0.12	0	22.8	6.45	8.51	6.6	4.71	4.24	46.04	0.864	242	342	9135610	[21]
84	0.23	1.06	0.01	24.75	6.78	6.8	6.01	4.35	3.89	46.12	0.905	256	330	10549044	[22]

85	0.35	0.47	0	26.52	7.71	6.05	4.03	4.08	4.22	46.57	0.892	253	331	10342200	[22]
86	0.54	0.5	0	27.79	6.89	6.34	4.45	3.29	3.5	46.7	0.887	214	329	11410894	[22]
87	1.39	0.28	0	21.32	6.21	8.11	6.37	3.84	5.45	47.03	0.87	257	342	8411656	[25]
88	0.26	1.26	0	28.27	6	4.85	3.6	4	4.26	47.5	0.908	274	330	11583264	[22]
89	0.54	0.18	0	21.62	6.03	8.39	7.1	4.75	3.85	47.54	0.863	236	342	8666764	[25]
90	0.06	0.81	0	31.34	7.42	5.68	2.82	1.63	2.64	47.6	0.892	270	329	12072795	[22]
91	0.78	0.1	0	20.64	5.8	7.05	6.9	5.95	5.11	47.67	0.857	237	343	8335813	[25]
92	0.12	0.51	0	28.81	7.91	6.46	3.17	2.07	3.25	47.7	0.876	250	330	11135102	[22]
93	0.79	0.1	0	24.79	6.84	6.5	5.37	4.09	3.67	47.85	0.884	227	330	9645825	[22]
94	0.15	0.39	0	27.55	7.86	6.3	3.59	2.58	3.68	47.9	0.896	152	329	9514824	[22]
95	0.68	0.16	0	22.84	6.28	7.83	6.24	4.63	3.44	47.9	0.864	226	342	9700984	[22]
96	0.57	2.46	0	36.37	3.47	4.05	1.93	1.57	1.62	47.96	0.9594	329	373	20339660	[30]
97	0.45	0.44	0	35.05	4.64	2.46	7.66	1.6	5.46	48.24	0.836	225	356	17374896	[6]
98	0.45	0.44	0	35.05	4.64	2.46	1.66	1.6	5.46	48.24	0.9	225	356	17374896	[20]
99	0.65	0.02	0	45.02	12.45	8.93	6.03	3.02	1.44	22.44	0.81	184	333	20698190	[31]
100	0.47	2.53	0	17.74	5.4	6.38	7.33	5.87	5.43	48.85	0.72	205	397	8604710	[31]
101	0.11	0.28	0	27.53	6.9	5.87	4.42	2.47	3.52	48.9	0.879	245	331	10617992	[22]
102	0.2	1.23	0.01	26.54	6.64	5.54	4.15	3.14	3.34	49.22	0.91	290	329	11010996	[22]
103	0.15	0.65	0	30.48	7.27	5.15	2.04	1.96	2.76	49.54	0.866	239	328	11514316	[22]
104	0.16	0.3	0	24.66	7.51	6.92	4.76	2.65	3.43	49.61	0.875	239	330	9652720	[22]
105	0.21	0.15	0	27.77	7.68	6.19	3.41	1.81	2.84	49.94	0.863	228	330	10962732	[22]
106	0.39	0.14	0	21.4	6.4	7.43	5.06	4.34	4.45	50.39	0.842	245	338	7749755	[25]
107	0.03	0.17	0	31.22	6.5	4.75	2.53	1.71	2.64	50.45	0.889	264	331	11652212	[22]
108	0.21	0.91	1.2	21.4	9.23	5.64	3.96	3.72	3.24	50.47	0.862	230	381	10342200	[29]
109	0	0.15	0	19.5	8.1	8.43	6	2.92	3.61	51.22	0.849	225	329	7067170	[22]
110	0.22	1.37	0	27.92	6.51	5.16	2.52	1.58	2.59	52.13	0.893	255	331	10673150	[22]
111	1.02	0.12	0	19.76	5.52	6.89	4.87	4.49	2.67	54.66	0.85	247	345	7942810	[25]
112	0.33	0.19	0	35.42	3.36	0.9	0.95	0.4	0.72	57.73	0.917	255	344	15906304	[27]
113	0.21	0.75	0.51	6.05	2.59	5.83	7.69	6.14	5.42	64.81	0.857	231	338	2426970	[25]
114	0.31	0.28	0.02	6.8	1.98	4.01	6.62	6.57	6.65	66.76	0.858	237	336	2578655	[25]
115	0.25	0.24	0	40.91	10.398	9.01	5.13	3.28	2.21	28.58	0.8	182	422	20980876	[32]
116	0.88	1.34	0	5.63	2.51	4.6	7.31	5.99	4.71	67.03	0.855	224	327	2592445	[25]
117	0.3	0.01	0	7.14	1.54	3.71	7.31	6.65	6.19	67.15	0.86	233	333	2578655	[25]
118	0.41	0.26	0	6.14	2.38	4.71	6.27	5.64	4.68	69.51	0.86	225	330	2385601	[25]
119	0.35	0.56	1.41	9.99	1.45	1.87	3.64	4.47	5.23	71.03	0.872	258	338	3488769	[25]
120	0.33	0.35	0	6.72	2.19	4.04	5.54	5.3	4.47	71.03	0.858	225	332	2482128	[25]
121	0.25	0.01	0	5.81	1.81	3.89	6.3	6.28	3.75	71.9	0.855	296	317	1482382	[29]
122	0.24	0.39	0	5.82	0.84	0.43	1.14	3.01	4.92	83.2	0.942	304	344	2158072	[22]

GA-LSSVM Modeling

GA-LS-SVM was carried out by LSSVMLab 1.6 free toolbox and the Genetic Algorithm Toolbox of MATLAB R2008 b was used for parameter setting. All programs were run on a Pentium IV(CPU 2.7 GHz and 2GB RAM) personal computer with Windows XP operating system.

Neural modeling

Accuracy of neural network prediction strongly related to the number of neurons in the hidden layer. Fig. 3 illustrates the ARD versus the number of neurons in the hidden layer. This graph clearly shows that 13-15-1 topology is the best topology with a minimum amount of error.

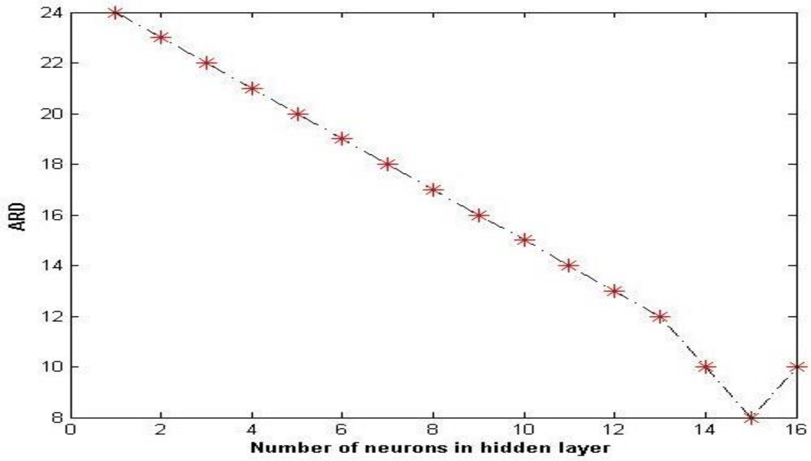

Fig. 3. Average relative deviation versus the number of neurons in the hidden layer

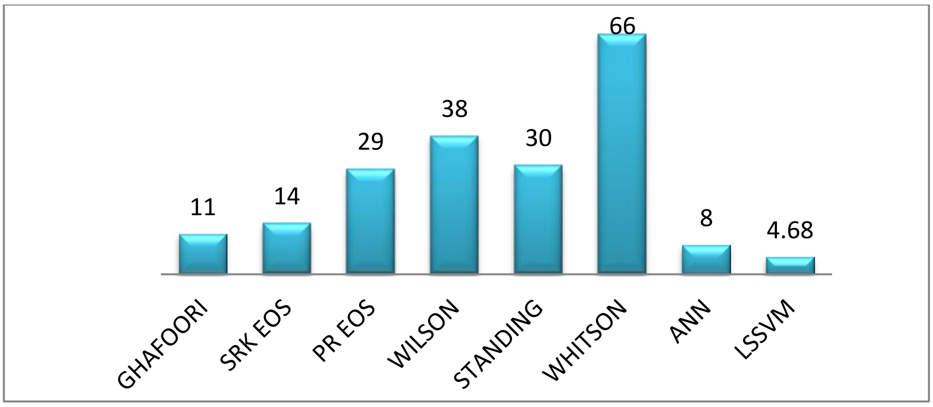

Fig. 4. ARD of different methods

Results and Discussion

The parameter setting step of LS-SVM is the most important step. In this work, we used genetic algorithm optimization method for parameter setting. Data is divided into two parts (70% for training and 30% for testing). After training of data with the training subset, the average relative deviation of testing data was calculated for accurate determination of model. As mentioned before average relative deviation of testing data was used as an objective function in the parameter setting step. The average relative deviation was calculated by means of Eq. 27. Results showed the LS-SVM model with γ =39.82 and δ^2=102.55 presents a minimum ARD. Also, our study showed a neural network with a sigmoid transfer function and 13-15-1 topology is the most accurate neural model.

Results showed GA-LS-SVM is more accurate than neural network and semi-empirical equations. The best semi-empirical equation has an ARD about 11% while our proposed model (with γ =39.82 and δ^2=102.55) has an ARD of about 4.68% and the accuracy of the neural model is 8%. In Fig. 4 ARD of different methods is compared.

Conclusions

In this study, ANN and LS-SVM methods were compared with semi-empirical equations for the prediction of bubble point pressures of 122 crude oil samples. Pressure and temperature of these reservoir fluids were in a wide range of 1482382 and 34474000 Pa, and 317 and 427 K, respectively. Results showed that ARDs of testing data were 4.68% and 8% for LS-SVM and ANN models, respectively while the best semi-empirical equation's ARD was 11%. Also, it was found that a unique LS-SVM with two appropriate adjusting parameters can predict the bubble pressure of reservoir fluids accurately. It can be concluded that LS-SVM and neural modeling can be considered as a suitable substitute for traditional empirical correlations obtained by regression.

Nomenclature

List of symbols

b_j	Bias term
e_k	Desired error
K	Kernel function
K_i	K-value of component i
$M_{W\ C7+}$	Molecular weight of C_{7+}
P	Pressure (Pa)
P_{ci}	Critical pressure of component i (Pa)
P_k	Convergence pressure (Pa)
P_{ri}	Reduced pressure of component i
T	Temperature (K)
T_{bi}	Normal boiling point of component i (K)
T_{Ci}	Critical temperature of component i (K)
T_{Cmix}	Critical temperature of the mixture (K)
T_{Ri}	Reduced temperature of component i
T_{Rmix}	Reduced temperature of the mixture
w	Weight vector
x_i	Mole fractions of component i in the liquid phase
y_i	Mole fractions of component i in the vapor phase
z_i	Mole fraction of component i
(x_k, y_k)	Original values of a sampling point

Greek letters

α_k	Lagrange multipliers
γ, δ	Parameters of LS-SVM model
γ_{C7+}	Specific gravity of C_{7+}
θ	Output of each neuron
φ (x)	Nonlinear mapping function
ψ	Sum of weighted inputs to each neuron
ω_i	Acentric factor of component i

List of abbreviations

ANN	Artificial neural network
ARD	Average relative deviations
GA	Genetic algorithms
KKT	Karush-Kuhn-Trucker
LS-SVMs	Least square support vector machines
MLP	Multi-layer perceptron
QP	Quadratic programming
RBF	Radial basis function
RNN	Recurrent neural networks
SLT	Statistical learning theory
SRM	Structural risk minimization
SVMs	Support vector machines

Acknowledgement

The authors acknowledge the funding support of Babol Noshirvani University of Technology through Grant Program No. BNUT/370675/97.

References

[1] Souders M, Selheimer C, Brown GG. III.-Equilibria between Liquid and Vapor Solutions of Paraffin Hydrocarbons. Industrial & Engineering Chemistry. 1932 May;24(5):517-9.
[2] Hoffman AE, Crump JS, Hocott CR. Equilibrium constants for a gas-condensate system. Journal of Petroleum Technology. 1953 Jan 1;5(1):1-0.
[3] Brinkman FH, Sicking JN. Equilibrium ratios for reservoir studies. Petrole Trans AIME 1960;219:313-9.
[4] Jaubert JN, Mutelet F. VLE predictions with the Peng–Robinson equation of state and temperature dependent kij calculated through a group contribution method. Fluid Phase Equilibria. 2004 Oct 1;224(2):285-304.
[5] Ahmed T. Reservoir engineering handbook. 2nd ed. Houston. Texas:Gulf Publishing; 2001.
[6] Standing MB. A set of equations for computing equilibrium ratios of a crude oil/natural gas system at pressures below 1,000 psia. Journal of Petroleum Technology. 1979 Sep 1;31(09):1-93.
[7] Whitson CH, Torp SB. Evaluating constant volume depletion data. InSPE Annual Technical Conference and Exhibition 1981 Jan 1.
[8] Ghafoori MJ, Aghamiri SF, Talaie MR. A new empirical K-value equation for reservoir fluids. Fuel. 2012 Aug 1;98:236-42.
[9] Vapnik V. Statistical learning theory. New York: John Wiley and Sons Inc.;1998.
[10] Li CH, Zhu XJ, Cao GY, Sui S, Hu MR. Identification of the Hammerstein model of a PEMFC stack based on least squares support vector machines. Journal of Power Sources. 2008 Jan 3;175(1):303-16.
[11] Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural processing letters. 1999 Jun 1;9(3):293-300.
[12] Kelly Jr JD, Davis L. A Hybrid Genetic Algorithm for Classification. InIJCAI 1991 Aug 24 (Vol. 91, pp. 645-650).
[13] Lahiri SK, Ghanta KC. Artificial neural network model with the parameter tuning assisted by a differential evolution technique: The study of the hold up of the slurry flow in a pipeline. Chemical Industry and Chemical Engineering Quarterly/CICEQ. 2009;15(2):103-17.
[14] Cartwright HM. Applications of artificial intelligence in chemistry. Oxford: Oxford University Press; 1993.
[15] Zhang Y. An improved QSPR study of standard formation enthalpies of acyclic alkanes based on artificial neural networks and genetic algorithm. Chemometrics and Intelligent Laboratory Systems. 2009 Oct 15;98(2):162-72.

[16] Mehdizadeh B, Movagharnejad K. A comparative study between LS-SVM method and semi empirical equations for modeling the solubility of different solutes in supercritical carbon dioxide. Chemical Engineering Research and Design. 2011 Nov 1;89(11):2420-7.
[17] Bakhshi H, Dehghani A, Jafaripanah S. Using the Genetic Algorithm based on the Riedel Equation to Predict the Vapor Pressure of Organic Compounds. International Journal of Engineering. 2018 Jun 1;31(6):863-9.
[18] Aghaeinejad-Meybodi A, Ebadi A, Shafiei S, Khataee A, Kiadehi AD. Degradation of Fluoxetine using catalytic ozonation in aqueous media in the presence of nano-γ-alumina catalyst: Experimental, modeling and optimization study. Separation and Purification Technology. 2019 Mar 18;211:551-63.
[19] Gas processors suppliers association engineering data book. 12th ed., FPS version. Oklahoma. 2004.
[20] Standing MB. Volumetric and phase behavior of oil field hydrocarbon systems. Society of petroleum engineers of AIME; 1977.
[21] Privat R, Jaubert JN, Mutelet F. Addition of the nitrogen group to the PPR78 model (predictive 1978, Peng Robinson EOS with temperature-dependent k ij calculated through a group contribution method). Industrial & Engineering Chemistry Research. 2008 Mar 19;47(6):2033-48.
[22] Elsharkawy AM. An empirical model for estimating the saturation pressures of crude oils. Journal of Petroleum Science and Engineering. 2003 May 1;38(1-2):57-77.
[23] Coats KH, Smart GT. Application of a regression-based EOS PVT program to laboratory data. SPE Reservoir Engineering. 1986 May 1;1(03):277-99.
[24] Wiesepape CF, Kennedy HT, Crawford PB. A crude oil-natural gas system vapor-liquid equilibrium ratios (data at 250. degree. F and system containing 20% C7+). Journal of Chemical and Engineering Data. 1977 Jul;22(3):260-1.
[25] Wu R, Rosenegger L. Integrated oil PVT characterization-lessons from four case histories. Journal of Canadian Petroleum Technology. 1999 Dec 1;38(13).
[26] Jaubert JN, Avaullee L, Souvay JF. A crude oil data bank containing more than 5000 PVT and gas injection data. Journal of Petroleum Science and Engineering. 2002 Jun 1;34(1-4):65-107.
[27] Danesh A, Xu DH, Todd AC. A grouping method to optimize oil description for compositional simulation of gas-injection processes. SPE reservoir engineering. 1992 Aug 1;7(03):343-8.
[28] Pedersen KS, Blilie AL, Meisingset KK. PVT calculations on petroleum reservoir fluids using measured and estimated compositional data for the plus fraction. Industrial & engineering Chemistry Research. 1992 May;31(5):1378-84.
[29] Mohammed SA. A simplified method for computing phase behavior of crude oil-carbon-dioxide mixtures. [Ph.D. dissertation]. College Station: Texas A&M University: 1988.
[30] Burke NE, Hobbs RE, Kashou SF. Measurement and Modeling of Asphaltene Precipitation (includes associated paper 23831). Journal of Petroleum Technology. 1990 Nov 1;42(11):1-440.
[31] Almehaideb RA, Al-Khanbashi AS, Abdulkarim M, Ali MA. EOS tuning to model full field crude oil properties using multiple well fluid PVT analysis. Journal of Petroleum Science and Engineering. 2000 May 1;26(1-4):291-300.
[32] Pedersen KS, Fredenslund A, Thomassen P. Properties of oils and natural gases. Gulf Pub Co; 1989.