Fluid Phase Equilibria 402 (2015) 124–132

Contents lists available at ScienceDirect

Fluid Phase Equilibria

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

An improved support vector regression model for estimation of saturation pressure of crude oils

Hamid Reza Ansari*, Amin Gholami

Abadan Faculty of Petroleum Engineering, Petroleum University of Technology, Abadan, Iran

ARTICLE INFO

Article history:
Received 10 April 2015
Received in revised form 19 May 2015
Accepted 23 May 2015
Available online 27 May 2015

Keywords:
Support vector regression (SVR)
Genetic algorithm (GA)
Imperialist competitive algorithm (ICA)
Particle swarm optimization algorithm (PSO)
Cuckoo search algorithm (CS)
Bat-inspired algorithm (BA)

ABSTRACT

Use of intelligence based approach for modeling of crude oil saturation pressure is viable alternative since this parameter plays influential role in the reservoir calculation. The objective of current study is to develop a smart model based on fusing of support vector regression model and optimization technique for learn the relation between the saturation pressure and compositional data viz. temperature, hydrocarbon and non-hydrocarbon compositions of crudes, and heptane-plus specifications. The optimization methods improve performance of the support vector regression (SVR) model through finding the proper value of their free parameters. The optimization methods which embedded in the SVR formulation in this study are genetic algorithm (GA), imperialist competitive algorithm (ICA), particle swarm optimization algorithm (PSO), cuckoo search algorithm (CS), and bat-inspired algorithm (BA). The optimized models were applied to experimental data given in open source literatures and the performance of optimization algorithm was assessed by virtue of statistical criteria. This evaluation resulted clearly show the superiority of BA when integrated with support vector regression for determining the optimal value of its parameters. In addition, the results of aforementioned optimized models were compared with currently available predictive approaches. The comparative results revealed that hybrid of BA and SVR yield robust model which outperform other models in term of higher correlation coefficient and lower mean square error.

1. Introduction

Reservoir fluid properties are overarching in petroleum engineering computations, such as, material balance calculations, well test analysis, reserve estimates, inflow performance calculations, and numerical reservoir simulations [1]. Among reservoir fluid properties, saturation pressure has utmost importance. Saturation pressure of crude oil is assigned to certain pressure that first bubble of gas is appeared in it [2]. Owing to great importance of saturation pressure, accurate calculation of these parameters is essential. Saturation pressure values are determined by laboratory experiments performed on samples of actual reservoir fluids. Requirement the large amount of time and money for experimentally measurement of saturation pressure are main motivation for developing models to estimating of these parameters from easy measured data [1]. Hence, modeling of saturation pressure based on easy obtained data has drawn substantial research attention in the last decade. There are a large number of previous studies relating with saturation pressure prediction in which three main groups of models are introduced. The three models are thermodynamic based models [3–4], models based on production data [1,5–38], and models based on compositional data [4,39–45].

Thermodynamic approach (Peng–Robinson (PR) and Soave–Redlich–Kwong (SRK) equation of state) is a well-known approach for modeling of saturation pressure. Employment of thermodynamic approaches is required to characterization and splitting the heavy fraction of crude oil. Moreover, these models cannot predict the saturation pressure with good precision. This drawback restricts the application of thermodynamic models for modeling of saturation pressure [3–4].

In the last years, a series of empirical correlations as well as soft computing models have been developed for making quantitative formulation between saturation pressure and production data including reservoir temperature, solution gas–oil ratio, oil gravity, and gas relative density [1,5–38]. The disadvantage of this category of models is that they are not globally which mean produce misleading results for crude oil of regions where their data are not used for constructing of model. Therefore these models are not suitable for prediction of saturation pressure.

* Corresponding author. Tel.: +98 936 361 8074.
E-mail addresses: ansari.hamid.r@gmail.com (H.R. Ansari), amingholamiput@yahoo.com (A. Gholami).

http://dx.doi.org/10.1016/j.fluid.2015.05.037
0378-3812/

Hitherto, various researchers are tried to modeling of saturation pressure as a function of compositional data [4,39–45]. Compositional data which is used for modeling of saturation pressure are temperature, hydrocarbon and non-hydrocarbon compositions of crudes, and heptane-plus specifications. Elsharkawy (ELSH) present empirical equation to predicting the saturation pressure from compositional data [4]. Bandyopadhyoy and Sharma (BAND) develop semi-analytical for predicting saturation pressure from compositional data apart from nitrogen fraction, hydrogen sulfide fraction, and specific gravity of the heptane plus fraction [39]. AlQuraishi (ALQ) proposed linear genetic programming to model saturation pressure as function of compositional data [40]. They performed impact analysis to selecting parameters that have the highest impact on the reservoir fluid saturation pressure. An impact analysis of AlQuraishi on the input parameters demonstrating that the three input variables, namely the methane mole fraction, molecular weight of the heptane plus component, and the reservoir temperature are suitable inputs for developing linear genetic programming based model. Farasat et al. proposed model based on least square support vector machine for estimation of saturation pressure as function of compositional data [41]. Kazemi et al. developed neural network based model for predicting of saturation pressure [42]. Ahmadi et al. used gene expression programming finding underlying relation between saturation pressure and compositional data [43]. Bagheripour and Asoodeh combined the results of three models including PR EOS, SRK EOS, and Elsharkawy through committee machine and generate a model with more accuracy [44]. Lately Gholami et al. present novel method based on committee machine (PLCM) for modeling of saturation pressure as a function of compositional data [45]. Firstly, they predict the saturation pressure through two individual models viz. alternating conditional expectation (ACE) model and support vector regression (SVR) model. Then, they fuse the aforementioned models through committee machine. They observed that integration of two models produce approaches with better accuracy for estimation of saturation pressure. In support vector regression model which employed by Gholami et al. hybrid of pattern search and grid search (HPG) is used as an optimization tool for determining the optimal values of SVR parameters. Although this method is powerful, it required large time for optimization implementation. Moreover, this method cannot extract the optimal values of support vector regression parameters. Hence, research for achieving potent optimization method for compute the optimum values of aforementioned parameters is still open issue.

In this study, optimized support vector regression model is proposed for quantitative estimation of saturation pressure from compositional data. Optimization implementation increases the efficiency of support vector regression model through election the optimal value of its parameters. Optimization algorithm in which employed for improving support vector regression efficiency are genetic algorithm (GA), imperialist competitive algorithm (ICA), particle swarm optimization algorithm (PSO), cuckoo search algorithm (CS), and bat inspired algorithm (BA). Finally, a statistical error analysis has been performed on the modeling results to investigate the feasibility and effectiveness of the proposed methods. Also, current developed models have been compared with previous models (thermodynamic based model, empirical equation and intelligence based model).

2. Model description

In this section, first the literature review relevant to the SVR is presented and then, there are some descriptions about the optimization algorithms including: GA, PSO, ICA, CS and BA.

2.1. Support vector regression

Support vector regression (SVR) is a robust approximation technique based on statistical learning theory [46–48]. This method developed by Vapnik et al. and its idea is based on linear regression in an m-dimensional feature space [49–50]. The input $x_i \in R^P$ is first mapped onto a high dimensional feature space using a nonlinear mapping function $\phi(x)$ and then a linear model is constructed in this space with a weight w and bias b term as:

Table 1
Partial dataset used in this study, including saturation pressures values and compositions of crude oils gathered from literature.

Number	N2	CO2	H2S	C1	C2	C3	C4	C5	C6	C7+	GRC7+	MWC7+	Temp(F)	PS (psi)	Reference
1	0.41	0.26	0	6.14	2.38	4.71	6.27	5.64	4.68	69.51	0.86	225	134	346	[72]
2	0.21	0.75	0.51	6.05	2.59	5.83	7.69	6.14	5.42	64.81	0.857	231	148	352	[72]
3	0.33	0.35	0	6.72	2.19	4.04	5.54	5.3	4.47	71.06	0.858	225	138	360	[72]
4	0.31	0.28	0.02	6.8	1.98	4.01	6.62	6.57	6.65	66.76	0.858	237	144	374	[72]
5	0.88	1.34	0	5.63	2.51	4.6	7.31	5.99	4.71	67.03	0.855	224	128	376	[72]
6	0.35	0.56	1.41	9.99	1.45	1.87	3.64	4.47	5.23	71.03	0.872	258	148	506	[72]
7	0.29	0.46	0.49	10.75	1.11	1.58	3.68	4.03	4.75	72.86	0.861	261	145	519	[72]
8	0.21	0.34	0	20.04	7.93	8	6.6	5.87	5.08	45.93	0.861	230	235	900	[73]
9	0.43	3.47	3.68	19.49	8.28	6.85	4.3	4.18	2.42	46.9	0.876	246	230	993	[73]
10	0.24	1.53	0.6	13.16	6.38	7.62	6.77	5.65	6.37	51.68	0.876	275	190	1140	[73]
11	0.77	1.99	1.4	17.38	6.42	7.62	5.62	4.53	5.14	49.13	0.891	267	234	1190	[73]
12	0.25	2.19	1.16	16.33	6.29	7.48	6.09	4.36	3.58	52.27	0.88	249	215	1261	[73]
13	0.17	0.56	1.93	12.59	6.05	6.51	4.26	4.52	1.14	62.18	0.877	230	239	1490	[73]
14	0.32	3.69	0.68	21.55	8.6	7.66	6.4	5.07	2.62	43.41	0.869	243	239	1591	[73]
15	0	0	0	36.15	12.17	8.05	5.81	4.79	5.24	27.79	0.722	191	212	2238	[78]
16	0	0	0	46.78	8.77	7.44	4.01	2.56	4.02	26.4	0.766	158	212	2941	[78]
17	0.65	0.02	0	45.02	12.45	8.93	6.03	3.02	1.44	22.44	0.81	184	140	3002	[80]
18	0.52	6.47	0	39.58	10.68	7.27	5.28	3.65	2.9	23.67	0.858	176	310	3627	[72]
19	0.34	7.1	0	48.43	9.24	5.84	4.39	3.21	2.28	19.17	0.805	183	314	4082	[72]
20	0.38	7.03	0	48.73	8.93	5.48	4.05	3	2.14	20.26	0.805	181	309	4156	[72]
21	0.3	0.9	0	53.47	11.46	8.79	4.56	2.09	1.51	16.92	0.864	143	176	4460	[75]
22	0	0	0	73.36	5.35	4.71	2.62	1	1.62	11.18	0.767	161	212	4742	[78]
23	0.24	0.27	0	66.83	8.28	5.15	3.31	2.04	1.85	12.03	0.8	182	215	4810	[83]
24	1.67	2.18	0	60.51	7.52	4.74	4.12	2.97	0	16.29	0.789	181	246	4823	[84]
25	0	0	0	57.53	10.16	5.83	3.28	2.71	1.4	19.11	0.81	203	212	5065	[78]
26	0.3	0.01	0	7.14	1.54	3.71	7.31	6.65	6.19	67.15	0.86	233	140	374	[72]
27	0.3	0.9	0	53.47	11.46	8.79	4.56	2.09	1.51	16.92	0.836	173	210	4460	[75]
28	0	0	0	74.18	5.32	4.67	2.58	0.97	1.56	10.72	0.766	159	212	4753	[78]