Journal of the Taiwan Institute of Chemical Engineers 50 (2015) 123–130

Contents lists available at ScienceDirect

Journal of the Taiwan Institute of Chemical Engineers

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

Prediction of the properties of brines using least squares support vector machine (LS-SVM) computational strategy

Milad Arabloo [a,*], Hossein Ziaee [a], Moonyong Lee [b], Alireza Bahadori [c,**]

[a] Young Researchers and Elites Club, North Tehran Branch, Islamic Azad University, Tehran, Iran
[b] School of Chemical Engineering, Yeungnam University, Gyeungsan, Republic of Korea
[c] School of Environment, Science & Engineering, Southern Cross University, Lismore, NSW, Australia

ARTICLE INFO

Article history:
Received 24 July 2014
Revised 30 November 2014
Accepted 7 December 2014
Available online 23 January 2015

Keywords:
Oil production
Salt water
Vapor pressure
Density
Enthalpy
Support vector machine

ABSTRACT

Natural brines occur underground or in salt lakes are commercially main sources of common salt and other salts, such as sulfates and chlorides of potassium and magnesium. This paper reports the implementation of a novel least square support vector machine (LS-SVM) algorithm for the development of improved models capable of predicting the properties of reservoir brine properties *i.e.*, liquid saturation vapor pressure, density and enthalpy. The validity of the presented models was evaluated by using several statistical parameters. The predictions of the developed models for determining the liquid saturation vapor pressure, density and enthalpy were in excellent agreement with the reported data with an average absolute relative deviation (AARD) of %0.069, %0.033, %0.072, respectively and coefficient of determination values (R^2) 0.999. According to the results of comparative studies, the developed models are more robust, reliable and efficient for calculating properties of oil field formation water during crude oil production than other techniques.

1. Introduction

Almost all hydrocarbon reservoirs are bounded by and in communication with water-bearing rock called aquifers [1]. Aquifer drive or natural water drive is one of the most effective oil production driving mechanisms, either as an edge water-drive or bottom water-drive [1]. Brine production increases with decreasing reservoir pressure during the production life [2]. The associated cost of handling this water production is forecasted to exceed several billion dollars per year [2]. In the USA, the water produced in oil production comprised 98% of all waste produced by the exploration and production (E&P) industry.

In other words, an average of ten barrels of salt water is produced for every barrel of oil [3]. Even with the best field management methods, brine production may eventually increase to a point that will represent more than 90% of the liquid volume brought to the surface [4]. The production of wet crude in many oil fields has been a growing field concern and the production of wet crude adversely affects the quality of the oil produced. A number of wells have had to be shut in due to a lack of adequate treatment facilities [5]. Understanding and dealing with these costs needs precise knowledge of the PVT properties of formation water. The nature, and physico-chemical properties of the produced formation water have a direct influence on the well productivity, degree of depletion and oil recovery efficiency and highlight the need to understand the characteristics of water associated with oil [6]. In addition, PVT properties of oilfield brines are used, either directly or indirectly, in many petroleum engineering calculations. Thus, the errors in estimation of PVT properties will propagate throughout estimates of other computations. Hence, it is absolutely necessary that the predictions for PVT properties be as accurate as possible. Over the last four decades, many experimental studies of the behavior of systems comprising salts in water have been published in the physical chemistry literature. Generally, a number of technical papers have been reported to address the various technical and practical points on formation water issues [7–9]. PVT properties of the formation water can be obtained either by conducting a laboratory study on reservoir fluid samples or by the use of predictive models. However, experimental determination of these properties is relatively expensive and time consuming [10]. Therefore, in the absence of experimental facilities, the characteristics are estimated from the correlations and soft computing techniques [11–16]. Based on the success of applying support vector machine (SVM) algorithm to solve a range of engineering problems, we are pursuing continued development and application of SVM for PVT modeling. On the other hand, to the best of our knowledge, there are no reports of

* Corresponding author. Tel.: +61 2 6626 9743.
** Corresponding author. Tel.: +98 917 140 5706.
E-mail addresses: milad.arabloo@gmail.com (M. Arabloo), alireza.bahadori@scu.edu.au (A. Bahadori).

http://dx.doi.org/10.1016/j.jtice.2014.12.005
1876-1070/

Nomenclature

A^T	transpose of matrix A
AARD	average absolute relative deviation
ANN	artificial neural network
ARD	average relative deviation
b	bias term
H	enthalpy (kJ/kg)
I_N	$N \times N$ identity matrix
$K(x_i, x_j)$	kernel function
L	Lagrangian
LS-SVM	least square support vector machine
P	vapor pressure (kPa)
RMSE	root mean squared error
T	temperature (K)
w	weight vector
α_i	Lagrange multipliers
Φ	map from input space into feature space
γ	regularization constant
Ω	kernel matrix
ρ	density (kg/m^3)
σ	width of the RBF kernel
ψ	salt concentration (mass fraction)

modeling the formation water PVT properties using the SVM approach. Therefore, the aim of this study was to propose computer based models for accurate calculations of the reservoir brine properties *i.e.*, enthalpy, density, and liquid vapor pressure. Highlighting the contribution of the paper, our developed LS-SVM model covers a wide range of input PVT data. Moreover, comparative studies are conducted between the developed models in this study and the existing correlations/models. The importance level of the input variables was also determined by using parametric sensitivity analysis technique. It is important to note that much higher accuracy in predicting brine PVT properties is attained while employing the developed intelligent model compared to the predictive models which would be an asset for engineering and research activities in this area. The reminder of this paper is organized as follows; Section 2 reports the background and computational procedure of the LS-SVM. Section 3, evaluates the adequacy and superiority of the proposed model by statistical and graphical error analysis. Section 4 summarizes the impacts of the factors on the reservoir brine PVT properties by applying a sensitivity analysis. Section 5 reports the conclusions of the work.

2. Methodology

2.1. Backgrounds of LS-SVM modeling

The support vector machine is a supervised learning technique from the field of machine learning applicable to both classification and regression analysis [17–23]. On the other hand, one of the major drawbacks of the SVM is the necessity to solve a large-scale quadratic programming problem [24]. This disadvantage has been overcome by a modification to the traditional SVM called least-squares SVM (LS-SVM), which solves linear equations (linear programming), instead of quadratic programming problems to reduce the complexity of optimization process [25–27]. Considering the problem of approximating a given dataset $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$ with a nonlinear function:

$$f(x) = \langle w, \Phi(x) \rangle + b \tag{1}$$

where $\langle ., . \rangle$ represents dot product; $\Phi(x)$ represents the nonlinear function that performs regression; b and w are bias terms and weight vector, respectively. In LS-SVM for function estimation, the optimization problem is formulated as [24,28]:

$$\min_{w,b,e} J(w, e) = \frac{1}{2}\|w\|^2 + \frac{1}{2}\gamma \sum_{k=1}^{N} e_k^2 \tag{2}$$

$$\text{s.t. } y_k = e_k + \langle w, \Phi(x_k) \rangle + b \quad k = 1, \ldots, N \tag{3}$$

where, $e_k \in R$ are error variables; and $\gamma \geq 0$ is a regularization constant. To solve this optimization problem, Lagrange function is constructed as [24,28]:

$$L_{\text{LS-SVM}} = \frac{1}{2}\|w\|^2 + \frac{1}{2}\gamma \sum_{k=1}^{N} e_k^2 - \sum_{k=1}^{N} \alpha_k \{e_k + \langle w, \Phi(x_k) \rangle + b - y_k\} \tag{4}$$

where, $\alpha_k \in R$ are Lagrange multipliers. The solution of Eq. (4) can be determined by partially differentiating with respect to w, b, e_k

Table 1
Operating ranges of gathered database for PVT properties of oilfield brine.

Variable	Minimum	Maximum	Mean
Temperature (K)	311.515	589.15	449.9
Salt concentration (mass fraction)	0.05	0.25	0.15
Vapor pressure (kPa)	111.0	12461.0	3530.3
Enthalpy (kJ/kg)	151.85	1375.06	674.1
Density (kg/m^3)	750.0	1199.0	1000.5

Table 2
Optimized parameters of the developed LS-SVM based models for determination of reservoir brine PVT properties.

	γ	σ^2
Saturated liquid enthalpy	7.617E+6	1.776
Saturated liquid density	7.432E+9	3.968
Saturated vapour pressure	2.5076E+6	0.3259

Table 3
The statistical parameters of the developed model for prediction of reservoir brine saturated liquid enthalpy.

Parameter	
Training set	
R^2	0.9999
Average relative deviation (%)[a]	−7.4E−4
Average absolute relative deviation (%)[b]	0.052
Root mean square error[c]	0.418
Number of data samples	33
Validation set	
R^2	0.9999
Average relative deviation (%)	5.9E−4
Average absolute relative deviation (%)	0.103
Root mean square error	0.742
Number of data samples	11
Test set	
R^2	0.9999
Average relative deviation (%)	0.012
Average absolute relative deviation (%)	0.111
Root mean square error	0.709
Number of data samples	11
Total	
R^2	0.9999
Average relative deviation (%)	2.55E−3
Average absolute relative deviation (%)	0.074
Root mean square error	0.561
Number of data samples	55

[a] Average relative deviation: $\text{ARD\%} = \frac{100}{N}\sum_{i=1}^{N}\left(\frac{Y_i^{\text{exp}} - Y_i^{\text{pred}}}{Y_i^{\text{exp}}}\right)$.

[b] Average absolute relative error: $\text{AARD\%} = \frac{100}{N}\sum_{i=1}^{N}\left(\left|\frac{Y_i^{\text{exp}} - Y_i^{\text{pred}}}{Y_i^{\text{exp}}}\right|\right)$.

[c] Root mean square error (RMSE): $\text{RMSE} = \left(\frac{\sum_{i=1}^{N}(Y_i^{\text{exp}} - Y_i^{\text{pred}})^2}{N}\right)^{\frac{1}{2}}$.