Journal of Molecular Liquids 242 (2017) 701–713

Contents lists available at ScienceDirect

Journal of Molecular Liquids

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

A new decision tree based algorithm for prediction of hydrogen sulfide solubility in various ionic liquids

Reza Soleimani [a], Amir Hossein Saeedi Dehaghani [b,*], Alireza Bahadori [c]

[a] Young Researchers and Elite Club, Neyshabur Branch, Islamic Azad University, Neyshabur, Iran
[b] Department of Petroleum Engineering, Faculty of Chemical Engineering, Tarbiat Modares University, Tehran 14115-143, Iran
[c] School of Environment, Science and Engineering, Southern Cross University, Lismore, New South Wales 2480, Australia

ARTICLE INFO

Article history:
Received 5 February 2017
Received in revised form 21 May 2017
Accepted 19 July 2017
Available online 20 July 2017

Keywords:
Ionic liquids
Hydrogen sulfide
Solubility
Prediction
Stochastic gradient boosting

ABSTRACT

The solubility of acidic components at various temperatures and pressures in ionic liquids (ILs) is one of the decisive property needed for the appraisal of ILs as potential substitutes for alkanolamines in industrial natural gas sweetening processes, therefore its modeling encompasses scientific and commercial interest. To that end, in the present work, an advanced machine learning approach called stochastic gradient boosting (*SGB*) tree technique is employed in the calculation of hydrogen sulfide (H_2S) solubility in 11 different ILs within the (303.15 to 363.15) K temperature and (0.0608 to 2.0168) MPa pressure range as a function of critical temperature, critical pressure and acentric factor of ILs accompanied with operational temperature and pressure. A collection of 465 experimental data points were assembled from the literatures. The statistical parameters including correlation coefficient (R) of 0.999543 and mean relative absolute error (MRAE), 0.022198, of the results form dataset values exhibit the high precision of the applied method. Furthermore, the prediction competence of the *SGB* model has been compared to two well-known equation of states (*EOS*) as well as Genetic Expression Programming (*GEP*) and least squares support vector machine (*LSSVM*) models. According to the results of comparative studies, it was found that the *SGB* model is more robust, reliable and efficient than other existing techniques for improved analysis and design of natural gas sweetening process.

1. Introduction

At the wellhead raw natural gas usually contains, aside from methane (CH_4) and heavier hydrocarbon gases, a variety of unwanted components such as acidic gases (e.g. hydrogen sulfide (H_2S) and carbon dioxide (CO_2)). Gas sweetening is one of the vital purification processes which is applied to rid natural gas of such acidic contaminants which, if let remain in the natural gas, would cause problems with corrosion, environmental, health and safety hazards [1]. The industrially method most used to sweeten natural gas are those using aqueous alkanolamine solutions in absorption/regeneration process. However, during the regeneration and recycling operations, these aqueous alkanolamine solutions have issues with corrosion, degradation and loss of amine (due to its volatility and thermal instability) during recycling operations results in environment concerns as well as additional solvent/absorbent requirement/cost [2]. In recent years, ionic liquids (ILs), thanks to their unique features, have emerged as potential alternatives to conventional alkanolamine solutions in removal of acid gases (CO_2 and H_2S) in gas sweetening processes.

ILs are known as molten salts, which are usually formed by a large organic cation and organic or inorganic anions. Their asymmetry (structure of the cation) frustrates them from being solid below 100 °C and this is the reason why these salts have a low melting point and remain liquid at these low temperatures. High thermal stability, high ionic conductivity, negligibly flammable, negligible vapor pressure, ability to dissolve a wide range of compounds, low melting point (below 100 °C), high heat capacity [3,4] are amazing features shared by all ILs.

One key feature that is exhibited by ILs for gas sweetening is their negligible vapor pressure and negligible volume loss due to vaporization when regenerated which is beneficial for both human exposure and replenishing ILs, i.e., less solvent loss and greater recycle to the absorption system [5].

Knowing about the solubility of acidic gases in ILs is a vital factor to explore the potential of ILs for acid gas separation at various temperatures and pressures [6–9]. Experimental data related to H_2S and CO_2 solubility are not accessible for all the existing ILs and potentially novel ILs [10], on the other hand owing to the experimental measurements may be highly problematic, and in some cases, hazardous, and also their time-consuming and costly nature, it is highly desirable to utilize powerful predictive methods for predicting the phase behavior of these kinds of systems. In addition, so as to competently design the gas sweetening processes dealing with mixtures of such acidic gases with ILs, engineers

* Corresponding author.
E-mail address: asaeedi@modares.ac.ir (A.H. Saeedi Dehaghani).

http://dx.doi.org/10.1016/j.molliq.2017.07.075
0167-7322/

Table 1
Temperature, pressure and H_2S solubility range of used ILs in this study.

No.	Ionic liquid	Temperature range (K)	Pressure range (MPa)	H_2S solubility range (mole fraction)	No. of data points	Reference
1	[hmim][PF_6]	303.15–343.15	0.138–1.09	0.05–0.441	34	19
2	[hmim][BF_4]	303.15–343.15	0.111–1.1	0.06–0.499	33	19
3	[emim][$EtSO_4$]	303.15–353.15	0.1137–1.2704	0.012–0.118	36	6
4	[C_8mim][Tf_2N]	303.15–353.15	0.0935–1.9119	0.063–0.7355	47	54
5	[C_6mim][Tf_2N]	303.15–353.15	0.0685–2.0168	0.0368–0.7012	57	54
6	[C_8mim][PF_6]	303.15–353.15	0.0845–1.9584	0.0463–0.6972	48	16
7	[bmim][PF_6]	303.15–343.15	0.123–1.011	0.044–0.405	42	8
8	[bmim][BF_4]	303.15–343.15	0.0608–0.836	0.03–0.354	42	8
9	[bmim][Tf_2N]	303.15–343.15	0.0944–0.916	0.051–0.51	44	8
10	[emim][PF_6]	333.15–363.15	0.1449–1.933	0.032–0.359	40	21
11	[emim][Tf_2N]	303.15–353.15	0.1077–1.686	0.049–0.609	42	21

or designers need to know the operational conditions to satisfy a desired acidic gases solubility in the applied IL solvent [11].

Vega et al. [12] provide a comprehensive overview of the different approaches that have been applied to describe the thermodynamic behavior of ionic liquids and the solubility of selected compounds in them, including CO_2, hydrogen (H_2), water, BF_3 and other compounds. Also, similar reviews have been presented by Haghbakhsh et al. [13], Eslamimanesh et al. [11] and Torrecilla et al. [14] about works on CO_2/IL systems. As review these articles, it can be found that various models for examples cubic equation of states such as the Peng–Robinson (PR) and Soave–Redlich–Kwong (SRK), Monte Carlo methods, Wilson's equation, the Non-Random Two-Liquid (NRTL) model, UNIQUAC, and the group contribution method of UNIFAC, the statistical associating fluid theory (SAFT EoS) and its modifications, such as PC-SAFT, tPC-PSAFT and SAFT-γ, and etc. used for modeling of solubility in ILs by different research groups.

In addition, reviewing the publications reveals, several studies have focused on modeling H_2S solubility in various ILs using different methods including van der Waals equation of state [15], extended Henry's law included Pitzer's virial expansion for the excess Gibbs energy, and the generic Redlich–Kwong cubic equation of state [16], Krichevsky–Kasarnovsky equation [6–8,17–21], modified perturbed hard sphere chain (PHSC) equation of state [22], soft-SAFT equation of state [23], equations of state based on the statistical associating fluid theory for potentials of variable range (SAFT-VR) and the perturbed chain statistical associating fluid theory (PC-SAFT) [24]. However, limited flexibility is a main defect of all above mentioned methods which results in their applicability ranges decrease.

One effective way to handle this problem consists in using powerful soft computing methods. In recent years, soft computing methods have attracted special attention from the scientific community due to their ability to model and analyze vague and complex problems that were previously awkward or impossible to solve. Artificial neural network (ANN), fuzzy logic, evolutionary computing, support vector machines (SVM) and decision tree based models are the most popular constituents of the soft computing area [25]. Artificial neural network have interested researchers to overcome such challenges as prediction of CO_2 solubility in various ILs [11,14], prediction of solubility of different gases in a specific IL [26], predicting Henry's law constants of gases in a particular IL [27] and estimating properties of the mixtures including ILs [28–35]. Interested readers referred to the literature for more application of ANNs and SVMs in the field of ILs [36–46].

Friedman was the first one to introduce a modern and competent decision tree based model called stochastic gradient boosting (*SGB*) tree, which is used for estimation and classification purposes [47]. This technique has emerged in recent years that is one of the most robust schemes for various purposes [48–53].

The benefits of regression trees and boosting (which merge numerous models) are incorporated by *SGB* tree. Small successively created classification or regression trees are entailed by *SGB* rooted in the field of decision trees that is based on 'pseudo'-residuals (i.e., loss function gradient from the previous tree). Assembling a tree from a random data subsample (drawn without replacement) at each iteration results in model improvements. Overfitting is minimized and accuracy of prediction is maximized significantly, using a small portion of the training data and incorporating randomization. In addition to other advantages in theory, SGB minimizes the demands for input variable transformation or for feature selection [47–49] that is probably a beneficial quality when operating upon high-dimensional/hyperspectral datasets [54]. Besides, this method has many other favorable features such as [55]:

- SGB models have been shown to produce more accurate results than competing composite-tree methods such as bagging or boosting using other methods such as AdaBoost.
- SGB models are as easy to create.
- SGB models can handle hundreds or thousands of potential predictor variables.
- SGB models are grown quickly – in some cases up to 100 times as fast as neural networks.
- SGB models are good for classification and regression problems.
- Immune to outliers.
- Irrelevant predictor variables are identified automatically and do not affect the predictive model.
- Guard against over fitting as a consequence of the stochastic (randomization) element in the SGB algorithm – generalizes very well.
- SGB models are often equal to or superior to any other predictive functions including neural networks.

Table 2
The critical properties and acentric factors of ILs used in this study.

Compound	T_c (K)	P_c (MPa)	*w*
[hmim][PF_6]	754.3	1.55	0.8352
[hmim][BF_4]	679.1	1.79	0.9258
[emim][$EtSO_4$]	1061.1	4.04	0.3368
[C_8mim][Tf_2N]	1311.9	2.1	0.4453
[C_6mim][Tf_2N]	1287.3	2.39	0.3539
[C_8mim][PF_6]	800.1	1.4	0.9069
[bmim][PF_6]	708.9	1.73	0.7553
[bmim][BF_4]	632.3	2.04	0.8489
[bmim][Tf_2N]	1265	2.76	0.2656
[emim][PF_6]	663.5	1.95	0.6708
[emim][Tf_2N]	1244.9	3.26	0.1818

Table 3
The *SGB* model option implementation and stopping parameters.

SGB parameters	Value
Learning rate	0.48
Number of additive terms	3000
Random test data proportion	0.2
Subsample proportion	0.7
Minimum n in child node	1
Maximum n of nodes	15