https://doi.org/10.36375/prepare_u.iiche.a366

Generalized NRTL model for predicting vapor-liquid equilibrium data from activity coefficient of binary component systems: using molecular descriptors

ANNISHH BEHHARA[†1], DANUSH SAI RUDRAPATTI BADRINARAYANAN[†2], IMRAN RAHMAN[2*], KARTHIKEYAN M[2]

[1]*Department of Chemical Engineering, Birla Institute of Technology and Science, Pilani, K K Birla Goa Campus, Goa, India*

[2]*Chemical Engineering and Process Development, CSIR – National Chemical Laboratory, Pune, Maharashtra, India*

**Corresponding author's mail id: imranrahmanncl@gmail.com*

[†]Equally contributed

Abstract:

Vapor-Liquid Equilibrium data is crucial for separation processes like distillation, extraction and manufacturing. Obtaining this data experimentally for desired conditions and systems is time-consuming and expensive. Therefore there is a necessity for an *a priori* generalized model which predicts this data based on the molecular descriptor information (of the desired system) given as an input to the model. This model is based on the Non-Random Two-Liquid (NRTL) model to predict binary interaction (NRTL) parameters. These predicted parameters are, in turn, used to calculate the activity coefficient, which is used to calculate the vapor-phase composition from the liquid-phase composition of the system. In this study, the molecular descriptors for individual components of the 28 binary systems were generated. The arithmetic mean of the molecular descriptors of the corresponding components was used as the molecular descriptor set for that binary system. The molecular descriptors based on properties relevant to vapor-liquid equilibrium were selected and used as independent variables to build the model using an Artificial Neural Network (ANN) in python. Better predictions were obtained with the coefficient of determination greater than 0.85 for each NRTL parameter. Once the liquid-phase composition is known, the model can predict the vapor-phase composition at the desired pressure and temperature.

Keywords: Non-Random Two Liquid (NRTL), Vapor-Liquid Equilibrium, Artificial Neural Networks, Molecular Descriptors, Binary Component System

Introduction:

Information about thermodynamic properties and phase equilibria of mixtures are essential for designing industrial processes like distillation, extraction, separation, designing new cost-effective processes, synthesis of new materials, and technological advancements. Experimenting to determine properties takes time, money, energy and effort[1]. As a result, theoretical and empirical thermodynamic property models are required, with experimental data used to predict properties. However, these models only succeed when insufficient or no experimental data is available, so developing a generalized property model to predict properties is critical.

The fugacity and activity coefficient must be considered since, in reality, most mixtures behave non-ideally. However, the method for calculating phase equilibrium in systems that are non-ideal in the liquid phase only is based on activity coefficient models such as Margules, Van Laar, Wilson, NRTL, UNIQUAC, and UNIFAC. In contrast, non-ideality in the vapor phase is described using equations of state models such as the Van der Waals Equation, Redlich Kwong Equation, Redlich Kwong Soave Equation, and Peng Robinson Equation at higher pressures[2].

Ravindranath[3] developed a structure-based generalized model for predicting pure-fluid vapor pressures, saturated phase densities, and binary mixture vapor-liquid equilibrium. Linear and non-linear (back-propagation network) NRTL and UNIQUAC models were used. Gebreyohannes[4] worked on a generalized predictive model for Non-Random Two-Liquid binary interaction parameters. The molecular descriptor values of the binary systems were taken as the absolute difference between the corresponding molecular descriptor values of the component molecules in the systems. Furthermore, in this study, a modified NRTL model was taken, in which only one parameter was used, which was used to prevent different prediction

https://doi.org/10.36375/prepare_u.iiche.a366

models for different binary interaction parameters (for example, different models for a_{ij} and a_{ji}), as that could give different values for the same parameter depending on the order of components taken. In the present work, the molecular descriptor values of the binary systems were taken as the arithmetic mean of the corresponding molecular descriptor values of the component molecules in the systems. A standard NRTL model was used in the present work. Different predicting models were made for different binary interaction parameters by taking a particular order of components based on their boiling points. The "i" component was considered to have the lower boiling point, and the "j" component was considered to have the highest boiling point between the system's two components.

The local composition method parameters are temperature-dependent, which makes them applicable to any temperature. In contrast, the parameters apply only to a given temperature range in simple empirical methods. The local composition method's multi-component behaviour can be obtained from binary data. In contrast, simple empirical methods are restricted to binary systems. The local composition methods consider that local compositions are different from bulk composition, which is the case when the interaction among similar molecules is different from between molecules of one component and the other. In contrast, the simple empirical methods do not differentiate between local and bulk compositions[5]. The present work obtained the activity coefficient using a non-random two-liquid (NRTL) local composition model.

$$\ln \gamma_i = \frac{\sum_j x_j \tau_{ji} G_{ji}}{\sum_k x_k G_{ki}} + \sum_j \frac{x_j G_{ij}}{\sum_k x_k G_{kj}} \left(\tau_{ij} - \frac{\sum_m x_m \tau_{mj} G_{mj}}{\sum_k x_k G_{kj}} \right)$$

$$\text{Where, } \tau_{ij} = a_{ij} + \frac{b_{ij}}{T(K)} \text{ and } G_{ij} = e^{-c_{ij}\tau_{ij}}$$

The activity coefficient of component 'i' is represented by γ_i. The binary interaction parameters for the local domain are a_{ij}, b_{ij}, and c_{ij}, with the molecule of component 'j' as the center surrounded by molecules of component 'i'. G_{ij} and τ_{ij} represent the adjustable parameters for the local domain, with molecules of component 'j' in the center and molecules of component 'i' surrounding it. The component's mole fraction in the liquid phase is represented by 'x.'

The prediction model for the parameters a_{ij}, b_{ij} and c_{ij} correspond to the local domain where the "j" component molecule is at the center, surrounded by the "i" component molecules and the prediction model for the parameter a_{ji}, b_{ji} and c_{ji} corresponds to the local domain where the "i" component molecule is at the center, surrounded by the "j" component molecules. Because of this reason, a_{ij} and a_{ji} are not equal to each other. Similarly, b_{ij}, c_{ij} and b_{ji}, c_{ji} and are not equal.

Materials and methods:

A set of 28 binary mixtures containing Acetone, Benzene, Chloroform, Ethanol, Methanol, Toluene, Water and p-Xylene are considered. The SMILES format files of molecular properties for each component in the systems were obtained from the ChEMBL online database[6]. The SMILES format files were used as input for the paDEL software[7], which was used to generate the molecular descriptors data for each component in the systems. 1444 molecular descriptors were obtained in total. The arithmetic mean of the corresponding molecular descriptor values of the components of the binary systems was taken as their molecular descriptor values. This was done to make the model applicable to systems of the same component.

A set of relevant molecular descriptors (those representing chemical bonds and functional groups) were shortlisted for this work based on their properties. These molecular descriptors' data were preprocessed by discarding those descriptors whose value was zero (0) for all the binary systems and those that had 'NaN' values. A final set of 162 molecular descriptors was obtained, and these descriptors were the independent variables, and the NRTL parameters (a_{ij}, b_{ij}, c_{ij}, a_{ji}, b_{ji} and c_{ji}) were the dependent or target variables in the respective models.

While making the 28 binary systems (from Acetone, Benzene, Chloroform, Ethanol, Methanol, Toluene, Water and p-Xylene), "i" component was considered to have the lower boiling point and "j" component was considers to have the highest boiling point between the two components of a system. Other ways might be to take the reverse of it or use any other property (like the molecular weight of the molecule, etc.) to assign "i" and "j". The purpose of classifying "i" and "j" based on a specific property, and not randomly chosen, is to have uniformity in the model and to get more ordered results. The binary interaction parameters database for the 28 binary compound systems was taken from Appendix B[8]. An Artificial Neural Network (ANN) was trained in Python (Jupyter Notebook - Anaconda platform) for the NRTL parameters a_{ij}, b_{ij}, a_{ji}, b_{ji}, c_{ij} and c_{ji} to determine the relationship between these parameters and the appropriate molecular descriptor set.

Results and Discussions:

The Train dataset consists of 16 binary systems. The Validation dataset consists of 6 binary systems, and the Test dataset consists of 6 binary systems.

Initially, when the first few random states (random numbers that vary the composition of train and test datasets) were used (starting from 0), negative values of coefficient of determination (R^2, a metric to measure the accuracy of the prediction, ranging from 0 to 1, where 1 stands for highest accuracy) were obtained, which occurred due to insufficient or incomplete learning from the train dataset used in that random state. Then, iterations were run up to 10,000 random states for ANN. The Train and Test dataset combination that gave the highest coefficient of determination value was taken to train the model.

Table 1. Parameters and their Corresponding Coefficient of Determination and Random States

Binary Interaction Parameter	Coefficient of Determination	Random State
a_{ij}	0.989	1670
a_{ji}	0.880	3074
b_{ij}	0.955	9053
b_{ji}	0.873	4566
c_{ij}	0.904	8243
c_{ji}	0.886	2158

Conclusions:

The coefficient of determination obtained with ANN was greater than 0.85 for all NRTL parameters. Thus, these models can be used to get reliable predictions of the NRTL parameters of binary systems when their molecular descriptors data is given as an input. When the liquid-phase composition and temperature are known, the predicted NRTL parameters can be used to calculate the activity coefficients of the binary systems using the NRTL equation, which can then be used to calculate the vapor-phase composition using the modified Lewis-Randall rule.

References:

(1) Del-Mazo-Alvarado, O., & Bonilla-Petriciolet, A. (2022).Assessment of the simultaneous regression of liquid-liquid and vapor-liquid equilibria data of binary systems using NRTL and artificial neural networks. Fluid Phase Equilibria, 561, 113537. https://doi.org/10.1016/J.FLUID.2022.113537

(2) Gebreyohannes, S., Robinson, R. L., Ramsey, J. D., & Hagan, M. (2014). Quantitative structure-property relationship generalized activity coefficient models. https://shareok.org/handle/11244/14847

(3) Gow, A. S. (1993). Calculation of Vapor-Liquid Equilibria from Infinite-Dilution Excess Enthalpy Data Using the Wilson or NRTL Equation. Industrial and Engineering Chemistry Research, 32(12), 3150–3161. https://doi.org/10.1021/IE00024A028/ASSET/IE00024 028.FP.PNG_V03

(4) Ravindranath, D. (2005). Structure-based Generalized Models for Pure-fluid Saturation Properties and Activity Coefficients. https://shareok.org/handle/11244/9662

(5) Vetere, A. (2004). The NRTL equation as a predictive tool for vapor–liquid equilibria. Fluid Phase Equilibria, 218(1), 33–39. https://doi.org/10.1016/J.FLUID.2003.10.013

(6) Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, Magariños MP, Mosquera JF, Mutowo P, Nowotka M, Gordillo-Marañón M, Hunter F, Junco L, Mugumbate G, Rodriguez-Lopez M, Atkinson F, Bosc N, Radoux CJ, Segura-Cabrera A, Hersey A, Leach AR, *Nucleic Acids Res. 2019; 47(D1): D930-D940.* *https://doi.org/*10.1093/nar/gky1075

(7) Yap CW (2011). PaDEL-Descriptor: An open-source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry.* **32** (7): 1466-1474

(8) Beneke, Glasser, Hildebrandt, Peters. (2013) Appendix B In: Understanding Distillation Using Column Profile Maps. John Wiley & Sons, Inc., pp. 345-347.