Artificial Neural Networks for the Solution of the Phase Stability Problem

Jones E. Schmitza, Roger J. Zempa, Mário J. Mendesab*

a Chemical Engineering Department, State University of Campinas, Campinas, Brazil

Abstract

Heterogeneous azeotropic distillation is a very important process in chemical and petrochemical industry. The design of such distillation processes makes intensive use of the thermodynamic properties of the systems involved, particularly those of the phase-equilibrium (liquid-liquid, liquid-vapor and liquid-liquid-vapor). The prediction of the thermodynamic properties of multiphase systems is complex, because, besides the equilibrium calculations, it involves the determination of the number and nature of the phases present in the system (phase stability tests). Hence, the problem is to develop methods that, for example, can tell whether, for a given overall composition, a system lies inside or outside the binodal surface (two liquid phases in equilibrium or a single stable liquid phase). From the numerous methods proposed for the construction of the phase diagram of heterogeneous azeotropic systems we chose that proposed by Pham and Doherty [Chem. Eng. Sci., 45, 7, 1823-1836, 1990]. This method is useful for systems presenting a two liquid phase region that is limited by an upper critical solution temperature (but it can be reformulated to accept systems limited by a lower critical solution temperature). These systems are the most commonly encountered in industrial heterogeneous distillation. The method is basically concerned with the determination of the maximum temperature for which, for a given overall composition, two liquid phases still coexist. The Pham and Doherty’s method was implemented and conveniently modified to generate data for the prediction of the kind of equilibrium possible under certain conditions. The data obtained was used to train an Artificial Neural Network. The resulting network was able to predict the correct kind of equilibrium with precision better than 99.9%.

Keywords: Thermodynamic Properties, Phase Stability Analysis, Equilibrium Phase Diagrams, Artificial Neural Networks

1. Introduction

The knowledge of the thermodynamic behavior is very important in the design, simulation and operation of separation processes. Consider, for example, the dehydration of ethanol, produced by the fermentation of sugar cane. The system water-ethanol has a homogeneous azeotropic point at about 90 mol% of ethanol, and thus cannot be separated by simple distillation. To produce anhydrous alcohol for use as a fuel, an entrainer is added to the mixture; the entrainer builds a binary heterogeneous azeotrope with water and a ternary heterogeneous azeotrope with ethanol-water, thus leading to a heterogeneous azeotropic distillation process.

In this process, a heterogeneous stream is produced at the top of the column that is then split in a decanter into an entrainer-lean distillate and an entrainer-rich reflux stream. Therefore the knowledge of the thermodynamic behavior of this system can help us in the choice of the better separation method.

Two main problems are met in analyzing such a multiphase distillation process: (a) to determine if at the upper plates of the column there occurs indeed the formation of the two ternary liquid phases; (b) to determine the amounts and compositions of each phase present in those plates. These two types of problems always appear in the analysis of the equilibrium for a multicomponent multiphase system. The first type of problem is a phase stability problem and the second one is a phase split problem.
Basically, multiphase equilibrium problems can be solved by two different approaches [1]. In the first approach, methods are based on the conservation of mass and the equal-fugacity criterion [2]. To obtain the correct solution using these methods one needs to have access to good initial guesses for the composition of those phases. Such methods, however, do not guarantee the attainment of the minimum Gibbs energy of the system, leading frequently to the prediction of trivial solutions or of an incorrect phase distribution. The techniques following the second approach are based in the minimization of the Gibbs energy subjected to the mass conservation and the equal-fugacity criterion [1, 3]. These methods can be considered superior, but they are much more computationally demanding. The precision of both the types of methods can be limited by the accuracy of the liquid phase thermodynamic model (EOS, NRTL, etc…).

Here we show the application of artificial neural networks for the solution of the phase stability problem.

2. Phase Stability Analysis

The aim of the phase stability analysis is to verify if, at specified conditions, a mixture will split into multiple phases. An alternative definition of this problem is “to decide if, at specified conditions, a thermodynamically feasible phase of a mixture is stable or unstable”. The solution of the phase stability problem is frequently formulated in terms of the tangent-plane criterion [4]. The tangent-plane criterion establishes that for a system for a specified pressure, temperature and overall composition, \( z \), a phase is stable if the plane tangent to the surface at composition \( y \) always falls below the molar Gibbs energy of mixing of the system.

Mathematically, the necessary and sufficient condition for the stability of the phase with overall composition \( z \) is [5]

\[
F(y) = \sum_{i=1}^{nc} y_i \left( \mu_i(y) - \mu_i(z) \right) \geq 0
\]

(1)

where \( \mu \) is the chemical potential, \( y \) is the mole fraction of the trial phase, \( nc \) is the number of components in the mixture and \( F \) is the distance function, which represents the distance from the tangent plane to the Gibbs energy surface at composition \( y \).

To apply condition (1) it is, theoretically, necessary to verify if it is satisfied at all compositions \( y \); it was however shown [5] that it is enough to verify if this condition is satisfied at all the stationary points of the surface \( F(y) \). Furthermore, the points of tangency \( (F(y)=0) \), are good estimates for the composition of the phases present in equilibrium.

The difficulty in finding all the solutions of Eq. 1 with complete certainty led to the development of numerous techniques for the solution of this equation [6, 7, 8]. The aim of this work was, among others, to develop a simple yet accurate method for the solution of both the phase stability and the phase split problems, for the particular case of a ternary system with a heterogeneous azeotropic point.

2.1. Pham and Doherty’s method

Pham and Doherty [9] proposed a simple method for the stability analysis of heterogeneous azeotropic ternary systems that have a upper critical solution temperature (UCST), by far the most common in azeotropic distillation. In these systems, three different situations concerning the phase equilibrium can be met, Liquid-Liquid (LLE), Liquid-Liquid-Vapor (LLVE) and Liquid-Vapor (LVE).

![Fig. 1. The phase diagram of a ternary system maintaining a wide miscibility gap.](image-url)
Figure 1 qualitatively represents the phase diagram of a ternary system, where the two components A and B exhibit a wide miscibility gap. For this system, depending on the overall composition, two types of phase equilibrium are found, LLE and a single Liquid-Phase. The stability analysis problem would then consist in determining if, for a given temperature and overall composition the system in equilibrium would present two or only one liquid phases.

Suppose a system of an overall composition \((x_1, x_2)\) and temperature \(T_0\), as represented by point \(P\) (Figure 1). The question is if “at these conditions the equilibrium state will have one or two liquid phases”. Instead of applying the tangent-plane criterion to the Gibbs energy surface of the system, Pham and Doherty propose the use of a maximum temperature criterion.

As shown in Figure 1, if the point \(P\) falls in the binodal plane of the system at temperature \(T_0\), then the line drawn vertically upward from \(P\) into the \(T-x\) will always intersect the binodal envelope at a point \(P_1\) corresponding to the same overall composition and to a temperature \(T_{\text{max}}\). This temperature \(T_{\text{max}}\) is thus the maximum temperature for which, at the overall composition \((x_1, x_2)\), two liquid phases coexist. At temperature \(T_{\text{max}}\) there is a second liquid phase, represented by point \(P_1\), that is connected to \(P_1\) by a tie-line. The two liquid phases \(P_1\) and \(P_1\) must satisfy the equilibrium condition:

\[
x_i \gamma_i(T_{\text{max}}, x, P) = x_i' \gamma_i(T_{\text{max}}, x', P) = 0 \quad (i = 1, 2, 3)
\]  

subject to the conditions

\[
\sum_{i=1}^{3} x_i - 1 = 0 \quad (3a)
\]

\[
\sum_{i=1}^{3} x_i' - 1 = 0 \quad (3b)
\]

where \(\gamma\) is the activity coefficient, \(x\) and \(x'\) are the mole fractions of the liquid phases in equilibrium, \(P\) is the pressure.

Equations (2,3a-b) build a system of 5 equations in 5 variables. Thus, given two values of the \(x_i\) (for example, \(x_1, x_2\)) and \(P\), the values of \(T_{\text{max}}, x_3, x_1, x_2, x_1\) can be calculated.

Pham and Doherty [9] originally used the Eqs. (2, 3a-b) as a set of ordinary differential equations and solved it using the Gear’s method. However, the results obtained in the steady state correspond to that solving the algebraic set of equations using the Newton-Raphson’s method. The procedure is described in the Figure (2).

This method depends on the initial guesses for \(x_1\) and \(x_2\). Normally, each of the liquid phases is rich in one of the components. Thus, if the point \((x_1, x_2)\) lies in the region rich in the component 1 the initial guesses for \(x_1\) and \(x_2\) must be in the region dominated by the component 2. The program developed can decide the “side” of the point \(P\) (in the figure 2) and choose convenient initial guesses.

When the specified pressure and \((x_1, x_2)\) are such that the point lies outside the binodal, no solution for \(T_{\text{max}}\) will be found. In this cases, the search for \(T_{\text{max}}\) must be made in a restricted
range of temperature, like \((T_L \leq T_{\max} \leq T_H)\). If one of these limit temperatures has been reached in the calculations, then phase splitting is not possible and the mixture must be considered homogeneous. The values of \(T_L\) and \(T_H\) can be chosen between the physical properties of the components of the mixture or can be specified conveniently. If the composition \((x_1, x_2)\), for given pressure, lies in the region in which VLLE occurs, the value of \(T_{\max}\) found will be virtual, because the actual equilibrium condition requires that the two liquid phases \(P_1\) and \(P_1'\) (in Figure 1) must be in equilibrium with a vapor phase and the binodal envelope will not exist under this conditions.

The Pham and Doherty’s method was presented for a mixture that exhibit a upper critical solution temperature, but it can be used for systems with a lower critical solution temperature (LCST). In this case the equations \((2, 3a-b)\) are used to found \(T_{\min}\). However, for both the cases the mixture must not split in more than two liquid phases.

2.2. Use of the \(T_{\max}\) criterion for phase equilibrium determination

The determination of \(T_{\max}\) is the key to solve the problems which arise in the LLE and VLLE calculations. Once the value for \(T_{\max}\) is known the equilibrium calculations becomes relatively easy. A computer program was made to obtain the phase characteristics for given \(T, P\) and \(x\) conditions. The program can distinguish between the various possible regions: liquid homogeneous, vapor homogeneous, VLE, LLE and VLLE; and, for the last three possibilities, to calculate the compositions of the phases in equilibrium. The execution of this program can be summarized by the flow chart in the Figure (3).

The calculations of \(T_{\text{BUB}}\) and \(T_{\text{DEW}}\), were made based in the procediments presented by Smith et al.[11]. However, the implementation of a robust algorithm to generate the initial guesses was necessary. The calculation of \(T_{\text{AZEO}}\) was made using the method presented by Pham and Doherty [9]. These calculations allows also to obtain the compositions in the vapour-liquid and vapour-liquid-liquid equilibria. If LLE was detected, the calculations were made using the procedure proposed by Walas [12].

The mixture of ethanol-ethyl acetate-water was used in the following demonstrations. This mixture have a wide miscibility gap and belongs to the UCST class. The NRTL equation were used for the calculation of the activity coefficients with the binary interaction parameters taken from Soares et al.[10]. The vapour pressures were calculated using the Antoine equation whose constants comes from the same reference. For the mixture of ethanol-ethyl acetate-water, at 760 mmHg, convenient values of \(T_H\) and \(T_L\) are 493 K and 293 K, respectively. These values are good choices because: the binodal envelope for this mixture, at high pressure, extends until temperatures like 473 K; and 293 K is a temperature low enough to guarantee that the whole range of temperature, possible in azeotropic distillation, lies between these limits. If the trivial solution is found, \(T_{\max}\) is reduced by 10 K. This process is carried out until the proper solution will be found or until a limit will be reached.

Some results of this method were compared with experimental data and presented in the Figure (4). The correct phase was predicted in this cases, but, obviously in the border regions some misclassified points can occur. This is due the parameters of the NRTL model. The calculated compositions, included the heterogenous azeotrope, was in good agreement with the experimental data.

3 – Artificial Neural Networks (ANNs)

Artificial neural networks were developed in attempts to imitate, mathematically, the characteristics of the biological neurons. They are composed by interconnected artificial neurons responsible by the processing of input-output relationship. This relationship is learned by training the ANN with a set of input-output patterns.

The ANNs can be used for different purposes. Approximation of functions and
classification are examples of its possible applications. The most common types of ANNs used for classification are the feedforward neural networks (FNNs) and the radial basis function (RBF) networks.

Probabilistic neural networks (PNNs) are a kind of RBFs that uses a Bayesian decision strategy [13]. In PNNs each input have its distance from the input vector calculated in the first layer. This process results a vector whose elements indicate how close the input is in relation of the training input. The second layer produce a vector of probabilities that will be used in the determination of the input class.

Feedforward neural networks are the most used for engineering purposes. They are designed with one input layer, one output layer and hidden layers. The number of neurons in the input and output layers equals the number of inputs and outputs, respectively. The great problem for FNNs is the determination of the ideal number of neurons in the hidden layer(s), few neurons produce a network with low precision and a higher number leads to overfitting, i.e., the network memorizes the patterns present in the training and the results for new inputs are poor. The use of techniques such as Bayesian regularization can overcome this problems, but a excessive number of neurons can still cause problems. Bayesian regularization techniques are commonly used together with the Levenberg-Marquardt algorithm, that is appears to be one of the fastest methods, but requires big computational storage.

The design of a PNN is faster than their feedforward counterparts and their generalization capabilities are very good. However, for PNNs the number of neurons depends on the size of the input set. Therefore, the PNNs are bigger than the FNNs, but no optimization of the number of neurons is necessary.

![Flow chart for the determination of the kind of phases (or equilibrium) present and their characteristics.](image)

Fig. 3. Flow chart for the determination of the kind of phases (or equilibrium) present and their characteristics.
4 – Results and Discussion

The Pham and Doherty’s method was used to obtain a dataset with 58,828 input-output patterns covering the whole range of compositions and temperatures between 335.2 K until 373.6 K for the mixture of ethanol(1)-ethyl acetate(2)-water(3) at 760 mmHg. This patterns are composed by the overall compositions of the ethanol and ethyl acetate, the temperature and the phase equilibrium possible under these conditions. These patterns were used in the training of ANNs for the classification of the kind of phase equilibrium that exists under certain conditions. This ANN is illustrated in Figure 5.

The two types of ANN’s, FNN’s and PNN’s, were used in this classification problem and their results compared. The radial basis transfer function used for the PNN’s was:

\[ f(o) = e^{-o^2} \]  \tag{4}

and two types of transfer functions were used for the FNN’s, the logarithmic sigmoid (logsig):

\[ f(o) = \frac{1}{1 + e^{-a}} \]  \tag{5}

and the hyperbolic tangent sigmoid (tansig):

\[ f(o) = \frac{2}{1+e^{-2o}} - 1, \]  \tag{6}

were used.

Initially, the PNN’s were trained with different number of input-output patterns. The results were compared in terms of the percentage of misclassifications for the testing dataset. The results are showed in the Table 1. The largest training dataset have the better results. This is a consequence of the use of a RBF, these functions can only respond to small regions of the input space and the result is that more neurons are required. As stated above the number of neurons of a PNN’s depends upon the number of training patterns, so the better results (in Table 1) were obviously for the larger training datasets.

FNN’s with one hidden layer were trained, but even very large networks do not were able to classify satisfactorily the kind of equilibrium. A FNN with 80 neurons in the hidden layer, trained using the Levenberg-Marquardt algorithm and Bayesian regularization gets 2.89 % of wrong classifications, using logsig transfer function, and 2.98 % for the tansig transfer function. The training with more than 80 neurons requires very much computational memory.
The choice of one multi-hidden layers network was made based varying the number of parameters (the number of synaptic weights plus bias) of the network, the type of transfer function and the size of the training dataset for networks with two and three hidden layers, as demonstrated in the Table 2.

The 2.(3³) possible combinations of this variables were used for the evaluation of FNN’s with 2 and 3 hidden layers. These networks were trained for 1 000 epochs (one epoch corresponds to the presentation to the network of all the training patterns). For the networks with 3 hidden layers the number of parameters used was 2 units lower, because no one integer number of neurons was able to match that (217) used for the 2 hidden layers. The best results are presented in the Table 3.

The best results for FNN’s with 3 hidden layers were better than their 2 hidden layers counterparts. In both the cases the best results were obtained with the higher number of parameters and using more than one type of transfer function. The size of the training dataset was important only for the FNN’s with 2 hidden. In the other hand, for the 3 hidden layers FNN’s, the use of mixed transfer functions was important.

The PNN’s were able to classify correctly almost all the testing patterns for large training datasets, but their performances is weak for small ones. For FNN’s the most important aspect is the number of synaptic weights and bias, the more neurons the best results will be achieved. Big FNN’s can to overcome the precision of a PNN, but the training of a PNN is faster than that of a FNN. The better PNN was trained in 2 seconds and the better FNN taken almost 8 000 seconds (approximately 2 hours and 13 minutes).

Table 1
Percentage of misclassifications for different sizes of the training dataset.

<table>
<thead>
<tr>
<th>Number of patterns in the testing dataset</th>
<th>Number of patterns in the training dataset</th>
<th>Number of patterns misclassified</th>
<th>% of patterns misclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 928</td>
<td>53 900</td>
<td>1</td>
<td>0.0203</td>
</tr>
<tr>
<td>9 779</td>
<td>49 049</td>
<td>16</td>
<td>0.1640</td>
</tr>
<tr>
<td>14 707</td>
<td>44 121</td>
<td>59</td>
<td>0.4011</td>
</tr>
<tr>
<td>25 179</td>
<td>33 649</td>
<td>261</td>
<td>1.0400</td>
</tr>
</tbody>
</table>

Table 2
Variables for the choice of the best multi-hidden layer network.

<table>
<thead>
<tr>
<th>Number of parameters</th>
<th>Transfer functions</th>
<th>Size of the testing dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>217 or 215 (for 3 hidden layers)</td>
<td>logsig-tansig or logsig-tansig-logsig (for 3 hidden layers)</td>
<td>25 179</td>
</tr>
<tr>
<td>136</td>
<td>only tansig</td>
<td>14 707</td>
</tr>
<tr>
<td>73</td>
<td>only logsig</td>
<td>9 779</td>
</tr>
</tbody>
</table>

5 – Conclusions

In this work the use of artificial neural networks was tested for the phase stability problem. Two types of ANN’s were compared, the probabilistic neural networks and the feedforward neural networks. The first kind is better when a big collection of data is available,
Table 3
Configurations which reach the largest right classifications.

<table>
<thead>
<tr>
<th>Number of pattern misclassified</th>
<th>Percentage of pattern misclassified</th>
<th>Number of parameters</th>
<th>Type of transfer functions</th>
<th>Size of the testing dataset</th>
</tr>
</thead>
</table>
| 2 hidden layers
| 56                              | 0.5726                              | 217                  | logsig-tansig             | 9 779                       |
| 60                              | 0.6135                              | 217                  | tansig                    | 9 779                       |
| 67                              | 0.6851                              | 217                  | logsig                    | 9 779                       |
| 3 hidden layers
| 31                              | 0.2107                              | 215                  | logsig-tansig-logsig      | 14 707                      |
| 81                              | 0.3217                              | 215                  | logsig-tansig-logsig      | 25 179                      |
| 32                              | 0.3272                              | 215                  | logsig-tansig-logsig      | 9 779                       |

a priori, or when the training must be rapid. In the other hand, FNN’s works well for lesser datasets and when very much time is available. For large training dataset, 99.98 percent of classifications were correct, using a PNN. If a large experimental dataset be available, a ANN can be used without used without the restrictions of precision caused by the thermodynamic model. Thus, the results of the ANN can be more precise than that obtained by the methods based in the minimization of the Gibbs (or Helmholtz) energy.

List of symbols

- \( f \) transfer function
- \( F \) distance function
- \( o \) input signal
- \( x \) composition of the liquid phase
- \( y \) composition of the trial phase
- \( z \) overall composition

Greek letters

- \( \gamma \) activity coefficient
- \( \mu \) chemical potential

References