Application of Genetic Programming to Modeling and Prediction of Activity Coefficient Ratio of Electrolytes in Aqueous Electrolyte Solution Containing Amino Acids

Zaeifi Yamchi, Mahdi; Abdouss, Majid
Department of Chemistry, Amirkabir University of Technology, P.O. Box 15875-4413 Tehran, I.R. IRAN

Modarress, Hamid**+
Department of Chemical Engineering, Amirkabir University of Technology,
P.O. Box 15875-4413 Tehran, I.R. IRAN

ABSTRACT: *Genetic programming (GP) is one of the computer algorithms in the family of evolutionary-computational methods, which have been shown to provide reliable solutions to complex optimization problems. The genetic programming under discussion in this work relies on tree-like building blocks, and thus supports process modeling with varying structure. In this paper the systems containing amino acids + water + one electrolyte (NaCl, KCl, NaBr, KBr) are modeled by GP that can predict the mean ionic activity coefficient ratio of electrolytes in presence and in absence of amino acid in different mixtures better than the common polynomial equations proposed for this kind of predictions. A set of 750 data points was used for model training and the remaining 105 data points were used for model validation. The root mean square deviation (RMSD) of the designed GP model in prediction of the mean ionic activity coefficient ratio of electrolytes is less than 0.0394 and proves the effectiveness of the GP in correlation and prediction of activity coefficients in the studied mixtures.*

KEY WORDS: *Amino acid, Electrolytes, Activity coefficient, Modeling, Genetic programming.*

INTRODUCTION

Much attention has been paid to the development of efficient methods for separation, concentration and purification of valuable bio products in biotechnology [1]. Due to the application of amino acids in various fields such as pharmaceutical, chemical and food industries their physical and thermodynamic properties such as activity coefficients in the presence of electrolytes has been the interest of researchers as a key knowledge in separation process [1, 2]. Ordinary separation of amino acids is done in the presence of an electrolyte and in this regard knowledge about interactions of ions with biological molecules such as proteins and amino biological molecules such as proteins and amino acids is important in understanding their behavior and selecting the method

* *To whom correspondence should be addressed.*
\+ *E-mail: hmodares@aut.ac.ir*
1021-9986/09/3/71 10/$/3.00

the method of separation [2]. The polynomial correlation method, although not a predictive one, is an empirical approach to treat the activity coefficient data. The semi-empirical models usually cannot account for all the factors, which are not clearly understood at present. The universal parameters for the models are also difficult to obtain due to complexity of the factors involved in the activity coefficients. All of these limit the applications of thermodynamic semi-empirical models and polynomial correlation methods in prediction of activity coefficients, especially when the complex processes such as biomolecular separations are involved.

For aqueous electrolyte systems containing amino acids measurement of their thermodynamic properties such as mean ionic activity coefficient ratio $\gamma^{II}_{\pm}$ / $\gamma^{I}_{\pm}$ of electrolyte in presence of amino acid $\gamma^{II}_{\pm}$ and in absence of amino acid $\gamma^{I}_{\pm}$ is very time consuming, and how to model the activity coefficient ratio with accuracy have not been always fairly resolved problem.

Two models which can qualitatively represent the behavior of the water-electrolyte-amino acid systems at low electrolyte and amino acid concentrations have been developed by *Kirkwood* [3, 4]. *Merida et al.* [5] and *Raposo et al.* [6] applied the modified form of the Pitzer model [7] for aqueous solutions of an electrolyte and a nonelectrolyte to model the activity coefficients in water-electrolyte-amino acid systems. *Khoshkbarchi* and *Vera* [8, 9] proposed two models for the activity coefficients of amino acids in aqueous solutions containing an electrolyte.

The first model [8] is a combination of a short-range interaction term represented by the NRTL model [10] or the Wilson model [11] and a long-range interaction term represented by the Bromley model [12] or the K-V model [13]. The second model is based on the perturbation theory with a hard-sphere reference system. Both models were applied to several water-electrolyte-amino acid systems and were shown to be able to correlate the experimental data accurately over a wide range of amino acid and electrolyte concentrations.

Pazuki et al. [14] presented a model which is combination of long-range interactions and short-range interactions. For long range interactions they used *Khoshkbarchi-Vera* [15] model and for short-range interactions they used one of the models, *Wilson* [11], the NRTL [10] and the NRTL-NRF [16,17].

In modeling of mean ionic activity coefficient ratio of electrolytes, experimental data are essential and in fact they are some correlations with much or less adjustable parameters [5, 15]. Therefore, if the experiment data are not available they fail to predict the activity coefficient ratio.

Applications of GP [18,19] to chemical systems have included the generation of non-linear dynamic models of biotechnological batch and fed-batch fermentations [20-22], the identification of complex fluid flow patterns [23], the generation of steady-state input-output models of a range of industrial chemical process systems [24, 25].

One of the important applications of genetic programming is in generating input-output empirical models [26, 27]. The class of empirical models can be divided into two broad categories: (a) models with predefined structure (either linear or nonlinear), and whose parameters are determined to maximize the capacity to predict process data; or (b) black-box models with undetermined structure. An example of the first category would be a linear model relating a dependent variable, *y*, to a set of *n* independent variables, u_i:

$$y = \sum_{i=1}^{n} a_i u_i$$

where the coefficients a_i are determined to maximize the predictive power of the model. *Dehghani* and *Modarress* [28] studied the application of ANN to the systems containing amino acids or peptide +water + one electrolyte (NaCl, KCl, NaBr, KBr) and they modeled this mixture by using different types of neural networks. Their designed artificial neural network (ANN) could predict the mean ionic activity coefficient ratio of electrolytes in presence and in absence of amino acid in different mixtures better than the commonly used polynomial equations.

However ANN can be considered as a black-box model in which the number and identity of the relevant inputs and the number of layers are the only attributes of the structure that are determined by the user. The disadvantage of this model is that the user must specify the structure of the model in advance, which is in general difficult to do. But the main disadvantage of neural network approach in general, is that no formal equation is obtained, and thus, the resulting model is difficult or impossible to analyze. Consequently, great care must be

taken with the ANN approach to prevent over-fitting. Although application of ANN method may result in very accurate root mean square deviation (RMSD), considering the inherent disadvantages of ANN method we preferred the application of GP in this work. Since GP, being an evolutionary method for automatically generating nonlinear input-output models, overcomes both of the disadvantages mentioned above, therefore structured models are obtained, whose complexity is optimized.

When applying GP to automatically generate nonlinear MISO (multiple inputs, single output) models, the probability of a given model surviving into the next generation is proportional to how well it predicts the output data. Components of successful models are continuously recombined with those of others to form new models, similarly to Genetic Algorithms (GAs). The GP optimizes the model structure, with a lower level nonlinear least-squares algorithm harnessed to perform the associated parameter estimation.

In this work GP was utilized to model and predict the mean ionic activity coefficient ratio of the electrolytes in all available (water + amino acid + electrolyte) systems. The results obtained illustrate the possibilities of an alternative and less cumbersome modeling approach relying on the application of GP for correlating and predicting the activity coefficient. The RMSD of the best tree-structure designed for prediction of the activity coefficient of electrolytes is 0.0394.

INTRODUCTION TO GP

GP belongs to a class of probabilistic search procedures known as Evolutionary Algorithms (others include Genetic Algorithm's, Evolution Strategies and Evolutionary Programming). These techniques use computational models of natural evolutionary processes for the development of computer based problem solving systems. All evolutionary algorithms function by simulating the evolution of individual structures via processes of reproductive variation and fitness based selection. The techniques have become extremely popular due to their success at searching complex non-linear spaces and their robustness in practical applications.

Similar to the GA, the GP is based on simple rules that imitate biological evolution. GP is an evolutionary method for automatically generating nonlinear MISO (multiple inputs, single output) models. The probability of a given model surviving into the next generation is proportional to how well it predicts the output data. Combining basis functions, inputs and constants creates an initial model population, whose complexity is controlled by the user. The models are structured in a tree-like fashion, with basis functions linking nodes of inputs, as in the example in Fig. 1, where the tree-structure for the model: y = (a × b + c /d) is shown, where y is the dependent variable, a, b, c and *d* are independent variables. Note that the example in Fig. 1 also illustrates the organization of the tree in terms of its root, the basis function at the highest level, which in this case is the summation function that sum (a×b) with (c/d). It is noted that the basis functions can be those requiring two arguments, such as '+' and '×' as in the example, or those with only one (e.g., exp(·) or √(·)). Each individual model in the population is then fitted to the empirical data using nonlinear regression, and then graded according to how well it matches the data.

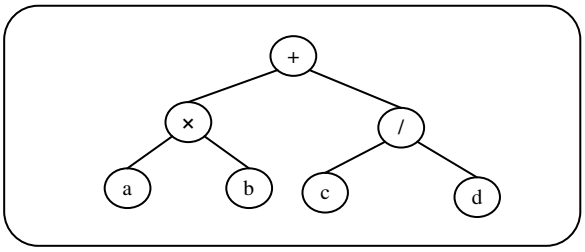

Fig. 1: Tree-structure for the model: (a × b + c /d).

As shown in Fig. 2, which presents the flow diagram for a generic GP, in each generation (iteration) of the algorithm, relatively successful individuals are selected as "parents" for the next generation and form a reproduction pool. A new generation of solutions evolves, using one of three possible operators: crossover, mutation and permutation.

Crossover is applied on an individual by simply switching one of its nodes with another node from another individual in the population (Fig. 2). With a tree-based representation, replacing a node means replacing the whole branch. This adds greater effectiveness to the crossover operator. The expressions resulting from crossover are very much different from their initial parents [18,19].

In mutation, a random change is performed on a selected individual by substitution, Fig. 3, this can be a

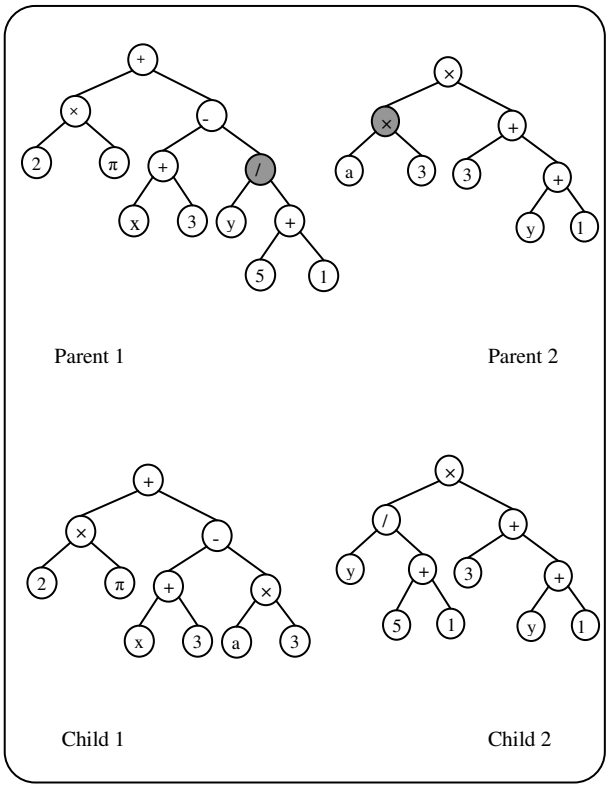

Fig. 2: Generic GP scheme [25].

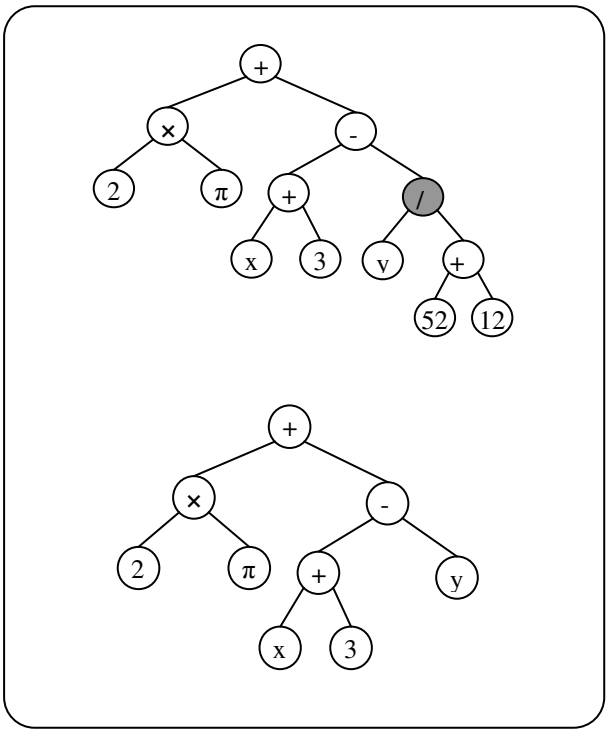

Fig. 3: Parents selected for the crossover operation with crossover nodes indicated by dark circles and resulting descended tree [19].

functional group, an input variable or a constant. Mutation affects an individual in the population. It can replace a whole node in the selected individual, or it can replace just the node's information. To maintain integrity, operations must be fail-safe or the type of information the node holds must be taken into account. For example, mutation must be aware of binary operation node, or the operator must be able to handle missing value [18, 19].

In contrast, branches of a selected individual are randomly switched in permutation. The parameters for each new individual in the new generation are determined by nonlinear regression, the models are then graded by fitness as before, and the procedure is repeated until a stopping criterion is attained. In most cases, as in this application, the population size, n_{Pop}, and the total number of generations, n_{Gen}, are decided in advance. Other tuning parameters that need to be fixed by the user when using GPs are p_c, p_m and p_p, the crossover, mutation and permutation probabilities. For the random population initialization step, used N_{tree}, the maximum number of sub-trees in an initial model, and p_{tree}, the probability of creating a sub-tree.

IMPLEMENTATION OF A GP MODEL

Specifying of the number of inputs and outputs

The first step in the implementation of GP is the definition of a terminal set. In other words, when developing a mathematical model of a process it is necessary to supply the input variables that are thought to be related to the output. In addition, the algorithm should have the ability to generate constants, as it will generally be a combination of input values and numeric constants that produce the required regression model. Our intensive search for experimental data indicated that the only available experimental data on mean ionic activity coefficient ratio ($\gamma^{II}_{\pm}$ / $\gamma^{I}_{\pm}$) of an electrolyte in presence ($\gamma^{II}_{\pm}$) and absence ($\gamma^{I}_{\pm}$) of amino acids in aqueous solutions are those reported in table 1.

The normalized input and format of the scaled output data are presented in table 2. As it is shown in table 2, the input data for ε/k, σ and D of amino acid have been normalized to shrink their variation range. This normalization is necessary due to their large variations, which can affect the performance of the model training. However for molalities as the input data, the normalization was not necessary due to availability of sufficient data in

the studied molality range. Also as table 2 shows the cation and anion diameters have limited variation and therefore their normalization did not affect the performance of the model training.

SPECIFYING O F THE OPERATORS

It is also necessary to define a functional set. This is the set of arithmetic operations and mathematical functions that the algorithm may use when constructing a potential solution to a given problem.

Typically, the functional set will include arithmetic operators such as addition and multiplication, and common mathematical functions (such as the square root, logarithm or exponential functions). It is important to ensure that each functional has the property of closure. That is, it must be able to accept and return a numeric value when presented with an input. The square root function does not have this property if presented with negative numbers. Thus, in order to ensure closure in this case, the absolute value of the input should be taken before evaluation. The operators that we use in this implementation reported in table 3.

The search space of the GP contains all possible models that can be constructed from elements in the functional and terminal sets. Regardless of the specific pplication, an algorithm searches the problem space using simple operations involving the copying, recombining and random alteration of mathematical expressions.

Creating a database of input-otput variables

A set of 750 data points was used for model training and the remaining 105 data points were used for model validation. The training set was used for identification of the tree structure of the model in training the GP and for the establishment of the candidate set of the most expressive input variables leading to smallest error. The second set which is the test set, is the 30 % remaining fraction and was used to test how well the GP model was trained.

The fitness function

Each individual in nature has some in-born skills that help it survive and reproduce. Following Darwinian philosophy, mainly the fittest individuals reproduce, while individuals that are less successful in coping with the challenges of their environment have fewer chances to survive and reproduce.

Table 1: Physical property of amino acids and ions [7].

Amino acid	σ (A°)	ε/k	Dipole moment
Glycine	4.76	65.40	11.85
Dl valine	4.22	8.30	10.68
Dl alanine	4.09	86.40	9.53
Amino butyric	4.73	83.00	10.87
Glycylglycine	6.50	170.20	23.09
Serine	4.83	189.33	10.34
Electrolyte properties			
Na^+	1.90		
K^+	2.66		
Cl^-	3.62		
Br^-	3.90		

Table 2: Format of input and output data.

Inputs
I_1 (electrolyte molality) = m_s
I_2 (amino acid molality) = m_A
I_3(normalized dipole moment of amino acid) = (23.09−D)/ (23.09−9.53)
I_4 (normalized ε/k of amino acid) = (189.33−ε/k)/ (189.33−8.3)
I_5 (normalized size parameter of amino acid) = (6.5−σ)/ (6.5−4.09)
I_6 (cation diameter) = σ_+
I_7 (anion diameter) = σ_-
Output
$O_1 = \gamma^{II}_{\pm} / \gamma^{I}_{\pm}$

Table 3: The operators that were used for the basic GP model.

0	1	2	3	4	5
+	-	×	/	^	$\sqrt{}$

Table 4: The parameters that were used for the basic GP test.

Parameter	Value
n_{Gen}	1000
n_{Pop}	200
p_c	0.75
p_m	0.15
p_p	0.10
Initial population parameters	
N_{tree}	5
p_{tree}	0.3

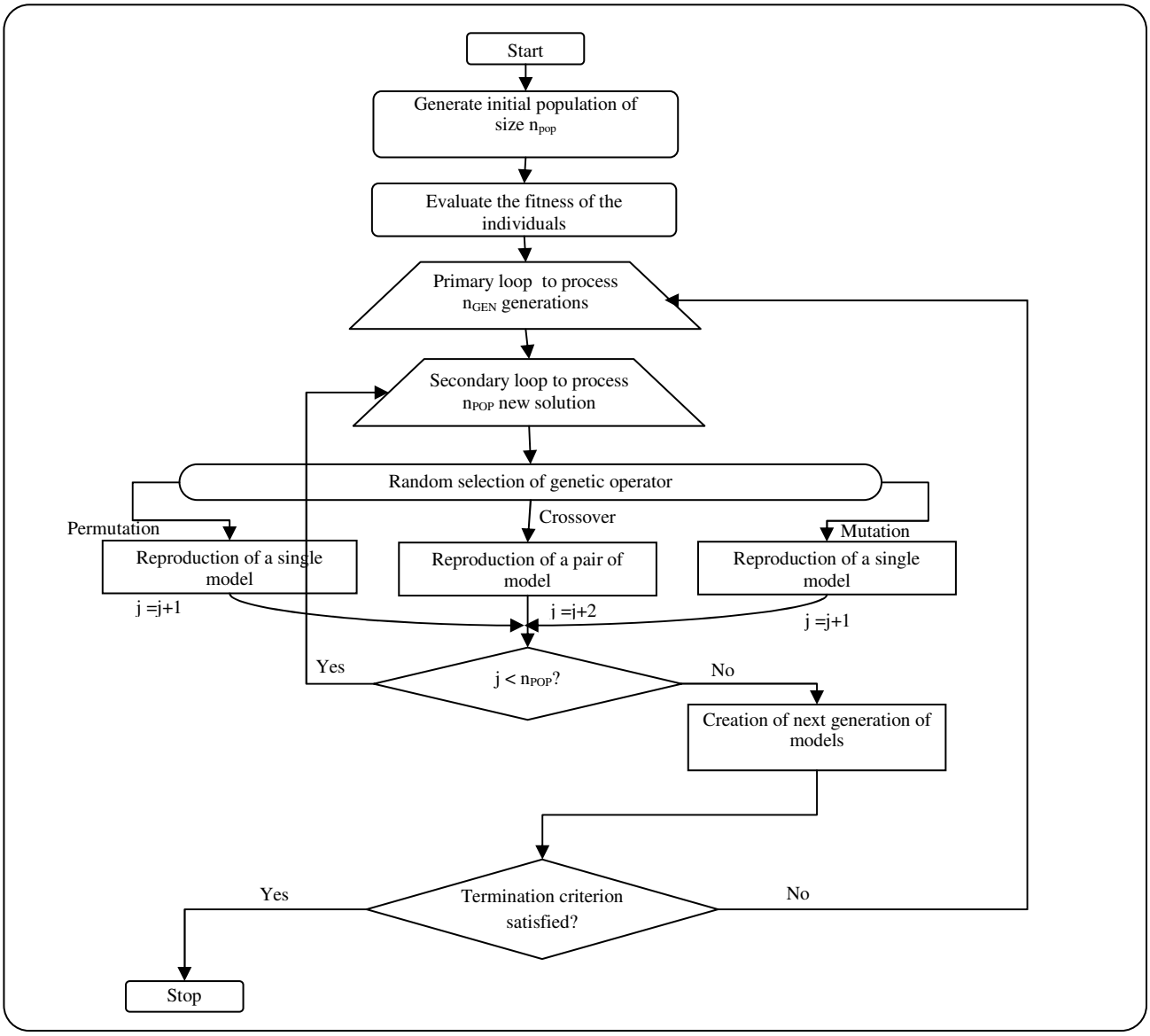

Fig. 4: Mutation operation [19].

In this application, the "environment" of the models is the output data-points to be predicted, with the fittest models being those that predict them the best.

In this work accuracy and precision of model were evaluated based on RMSD (root mean square error) according to Eq. (1):

$$RMSD = \left(\sum_{i=1}^{n} \left(y_i^{exp} - y_i^{calc} \right)^2 \Big/ n \right)^{1/2} \quad (1)$$

where y_i^{exp} is the experimental data, y_i^{calc} the calculated value by GP model, and n is the number of data sets used to training.

Determination of model str cture

On the basis of our experiments, we found that with the parameters given in table 4 the GP is able to find good solutions for our problem**.**

Since the results obtained using GPs are probabilistic in nature, it is important to study a problem by carrying out multiple runs of the algorithm, and analyze the results using descriptive statistics before arriving at conclusions. Thus, 50 distinct GP runs were done.

CALCULATION STEPS FOR GP ALGORITHM

The GP based synthesis algorithm includes the following steps:

1- The initial population is generated randomly.

2- Translate every node of a GP code into its corresponding actual connection pattern. It is easy to be concluded that the smaller is for the fitness; the better is for the GP code.

3- The operations of reproduction, crossover and mutation are performed according to the method defined in 2 where the reproduction rate, the crossover rate and the mutation rate is set to according to table 4. When the operations of reproduction is performed, the competition selection method of GP is adopted [19] and there are 10 % individuals of nowadays generation are selected and are sent into the next generation. To ensure the randomization of GP algorithm, when the operations of crossover and mutation are performed, the stochastic selection is adopted in this paper to determine the types of the evolution operation (i.e. the operation of node or the operation of node property).

4- The step 2 and 3 are repeated until the fitness does not change in the last two generations or the number of the maximal generation is reached that means the algorithm should be terminated.

RESULTS

We use GP for modeling of the activity coefficients ratio of electrolyte in presence and absence of amino acid or peptide for {water + amino acid or peptide + (KCl, NaCl, NaBr or KBr)} mixtures. The RMSD of model with acceptable error was recorded and the results were compared with experimental data, the results are presented in table 5. The mean absolute deviations on the learning set (7 0% of the database) and on the test set (remaining 30 %) were very close and thus confirming the high learning ability of the GP. Also in table 5 the RMSD obtained for the same systems by application of ANN method (Dehghani and Modarress [28]) are reported. It is seen that RMSD by GP method are very close those of ANN method. However considering the disadvantages of ANN, the results obtained by GP method in this work are more reliable and meaningful in interpretation.

In order to check the predict ability of GP model, we exclude some experimental data of table 6 in some ranges of molality from training set data, after training and testing of GP, it was validated with experimental data of systems and fortunately the trained GP could predict the

Table 5: Database and errors.

Systems	GP Results RMSD	ANN Results RMSD [28]	Reference
Glycine + NaCl	0.0717	0.0036	[33]
Glycine + NaBr	0.0992	0.0029	[33]
Glycine + KCl	0.0699	0.0032	[29]
Amino butyric + NaCl	0.0584	0.0091	[31]
Amino butyric + KCl	0.0146	0.0075	[31]
Amino butyric + NaBr	0.0115	0.0036	[31]
Amino butyric + KBr	0.0731	0.0048	[31]
Dl valine + NaCl	0.0373	0.0042	[38]
Dl valine + KCl	0.0139	0.0021	[38]
Dl alanine + NaCl	0.0272	0.0032	[37]
Dl alanine + KCl	0.0178	0.0028	[37]
Glycylglycine + NaCl	0.0669	0.0036	[30]
Glycylglycine + NaBr	0.0138	0.0027	[30]
Glycylglycine + KCl	0.0491	0.0023	[30]
Glycylglycine + KBr	0.0160	0.006	[30]
Dl serine + NaCl	0.0176	0.003	[8]
Dl serine + KCl	0.0131	0.002	[8]

$$\mathrm{RMSD}=\left(\sum_{i=1}^{n}\left(\left(\gamma_{\pm}^{\mathrm{III}}/\gamma_{\pm}^{\mathrm{I}}\right)_i^{\exp}-\left(\gamma_{\pm}^{\mathrm{III}}/\gamma_{\pm}^{\mathrm{I}}\right)_i^{\mathrm{calc}}\right)\Big/\mathrm{n}\right)^{1/2}$$

Table 6: Excluded data from training set.

System	Excluded range
α-Amino butyric + NaCl	m_{KCl} = 0.2, 0.5, 1.0 mol kg^{-1}
Glycine + NaBr	m_{NaBr}= 0.1, 0.2, 1.8 mol kg^{-1}
Glycylglycine + KBr	m_{NaBr}= 0.3, 0.7, 1.3 mol kg^{-1}
DL valine + KCl	m_{NaBr}= 0.2, 0.4 mol kg^{-1}

systems with an acceptable error of RMSD= 0.0394. Fig. 3 shows the parity plot of experimental and predicted data.

In the second step in order to check the predict ability of the GP we exclude all experimental data of (water + Glycylglycine + NaCl) system in all ranges of molality from training set data, after training and testing of GP, it was validated with experimental data of (water + Glycylglycine + NaCl) system and fortunately the trained GP could predict the new system of (water + Glycylglycine + NaCl) with an acceptable error of 0.08.

Table 7: Experimental [30] data and predicted result for activity coefficient ratio of Glycylglycine in different molality of NaCl and Glycylglycine.

$m_{Glycylglycine}$ (mol/kg)	m_{NaCl} (mol/kg)	Experimental $\gamma_{\pm}^{II}/\gamma_{\pm}^{I}$	Predicted $\gamma_{\pm}^{II}/\gamma_{\pm}^{I}$
0.1	0.1	0.9708	0.9825
0.2	0.1	0.9472	0.9630
0.3	0.1	0.9266	0.9356
0.5	0.1	0.8937	0.9100
0.7	0.1	0.8675	0.8814
1.0	0.1	0.8377	0.8487
1.3	0.1	0.8145	0.8211
1.5	0.1	0.8014	0.8045
0.1	0.3	0.9853	0.9987
0.2	0.3	0.9734	0.9870
0.3	0.3	0.9601	0.9742
0.5	0.3	0.9367	0.9459
0.7	0.3	0.9190	0.9247
1.0	0.3	0.8934	0.9008
1.3	0.3	0.8754	0.8825
1.5	0.3	0.8663	0.8721
0.1	0.5	0.9925	1.0054
0.2	0.5	0.9796	0.9879
0.3	0.5	0.9706	0.9754
0.5	0.5	0.9520	0.9612
0.7	0.5	0.9365	0.9403
1.0	0.5	0.9173	0.9214
1.3	0.5	0.9024	0.9034
1.5	0.5	0.8942	0.8950
0.1	0.7	0.9935	1.0029
0.2	0.7	0.9853	0.9951
0.3	0.7	0.9760	0.9828
0.5	0.7	0.9595	0.9626
0.7	0.7	0.9467	0.9511
1.0	0.7	0.9282	0.9300
1.3	0.7	0.9146	0.9168
1.5	0.7	0.9071	0.9062
0.1	1.0	0.9980	1.0102
0.2	1.0	0.9895	0.9954
0.3	1.0	0.9837	0.9889
0.5	1.0	0.9699	0.9737
0.7	1.0	0.9598	0.9621
1.0	1.0	0.9445	0.9506
1.3	1.0	0.9337	0.9328
1.5	1.0	0.9267	0.9259
RMSD			0.0844

Experimental data and predicted data for this system are presented in table 7.

CONCLUSIONS

The application of GP to the development of prediction models of activity coefficient of amino acids in electrolyte solutions has been considered.

The results revealed that the GP algorithm could generate an accurate input-output model based solely on observed data. It is also shown that the GP model can predict activity coefficient ratio of an electrolyte without using the adjustable parameters which restrict the thermodynamics models only to the systems with available experimental data. The GP method has the added advantage over ANN method that all the effective parameters on behavior of the studied mixture can be

included in the input data and therefore the method has the capability of predicting the behavior of systems where the experimental data are not available or there are limitations on experimental measurements.

A distinct advantage of this method is that no *a-priori* assumptions have to be made about the actual model form: the structure and complexity of the model evolve as part of the problem solution. Moreover, it would appear that GP has potential in the area of data analysis. One possible perceived disadvantage of the input-output models obtained using GP is that the identified structure does not provide detailed phenomenological information about the system being modelled (a characteristic that this approach shares with other black-box modelling techniques, including neural networks). However, in each case the final model form clearly indicates the relative contribution of each input to the output.

Based on the properly selected and the training procedure of GP, the mean ionic activity coefficient ratio ($\gamma^{II}_{\pm}$ / $\gamma^{I}_{\pm}$) was described fairly. It was demonstrated that then GPs are a promising strategy for solving the thermodynamic properties in complex systems, such as bio-macromolecular systems. Application of this method avoids the limitations of conventional thermodynamic methods that are tedious and requires determination of "parameters", which are arbitrary in many ways.

Received : 11th May 2008 ; Accepted : 26th August 2008

REFERENCES

[1] Cohn, E.J., Edsall, J.T., "Proteins Amino Acids and Peptides as Ions and Dipolar Ions", Hanfer, New York (1965).

[2] Marcozzi, G., Correa, N., Luisi, P.L., Caselli, M., *Biotechnol. Bioeng*., **38**, 1239 (1991).

[3] Kirkwood, J.G., *Chem. Phys.,* **2**, 351 (1934).

[4] Kirkwood, J.G., *Chem. Rev.,* **24**, 233 (1939).

[5] Merida, L.F., Raposo, R.R., Garcia, G.E.G., Esteso, M.A., *J. Electroanal. Chem.*, **379**, 63 (1994).

[6] Raposo, R., Merida, L., Esteso, M.A., *J. Chem. Thermodyn*., **26**, 1121 (1994).

[7] Pitzer, K.S., "Activity Coefficients in Electrolyte Solutions", 2nd Ed., CRC Press, Boca Raton, FL, (1991).

[8] Khoshkbarchi, M.K., Vera, J.H., *Ind. Eng. Chem. Res.,* **35**, 2735 (1996).

[9] Khoshkbarchi, M.K., Vera, J.H., *AIChE J.,* **42**, 2354 (1996).

[10] Renon, H., Prausnitz, J.M., *AIChE J*., **14**, 135 (1968).

[11] Wilson, G.M., *J. Am. Chem. Soc.*, **86**, 124 (1964).

[12] Bromley, L.A., *AIChE J.* **19**, 313 (1973).

[13] Khoshkbarchi, M.K., Vera, J.H., *AIChE J.,* **42**, 249 (1996).

[14] Pazuki, G.R., Rohani, A.A., Dashtizadeh, A., *Fluid Phase Equilib.*, **231**, 171 (2005).

[15] Khoshkbarchi, M.K., Vera, J.H., *Ind. Eng. Chem. Res.,* **35**, 4755 (1996).

[16] Haghtalab, A., Vera, J.H., *AIChE J.,* **34**, 803 (1988).

[17] Haghtalab, A., Sarkisian, E., *Sci. Iran.*, **5**, 67 (1998).

[18] Koza, J., "Genetic Programming: On the Programming of Computers by Means of Natural Selection", The MIT Press, USA, (1992).

[19] Koza, J., "Genetic Programming II: Automatic Discovery of Reusable Programs", The MIT Press, USA, (1994).

[20] Bettenhausen, K.D., Marenbach, P., Freyer, S., Renenmaier, H., Nieken, U., Proc. lEE Conf. on Genetic Algorithms in Engng. Systems: Inovations and Applications - GALESIA'95, Sheffield, UK, No 414, **481,** (1995).

[21] McKay, B., Elsey, J. Willis, M.J., Barton, G.W., Prec. IFAC, 13th World Congress '96, San Francisco, USA, 277 (1996).

[22] Watson, A.H., Parmee, I.C., Prec. Of the Second Online Workshop on Evolutionary Computation - WEC2, 45 (1996).

[23] McKay, B., Lennox, B., Willis, M.J., Montague, G., Barton, G.W., Prec. of the UKACC Int. Conf. on CONTROL-96, Exeter, UK, **2**, 734 (1996).

[24] McKay, B., Willis, M.J., Barton, G.W., *Computers and Chemical Engineering*, **21**, 981 (1997).

[25] Grosman, B., Lewin, D.R., *Computers and Chemical Engineering*, **26**, 631 (2002).

[26] Grosman, B., Lewin, D.R., *Computers and Chemical Engineering*, **28**, 2779 (2004).

[27] Dehghani, M.R., Modarress, H., Bakhshi, A., *Fluid Phase Equilibria* **244**, 153 (2006).

[28] Ardakani, M.K., Modarress, H., Taghikhani V., Khoshkbarchi, M.K., *J. Chem. Thermodyn.,* **33**, 821 (2001).

[29] Chung, Y.M., Vera, J.H., *BioPhys. Chem.,* **92**, 77 (2001).

[30] Chung, Y.M., Vera, J.H., *Fluid Phase Equilib.,* **203**, 99 (2002).
[31] Cohn, E.J., Edsall, J.T., "Proteins Amino Acids and Peptides as Ions and Dipolar Ions", Hanfer, New York (1965).
[32] Khavaninzadeh, A., Modarress, H., Taghikhani V., Khoshkbarchi, M.K., *J. Chem. Thermodyn.,* **35**, 1553 (2003).
[33] Khavaninzadeh, A., Modarress, H., Taghikhani V., Khoshkbarchi, M.K., *J. Chem. Thermodyn.,* **34**, 1297 (2002).
[34] Khoshkbarchi, M. K., "Ph.D. Thesis", McGill University, (1996).
[35] Marenbach, P., Bettenhausen, K.D., Freyer, S., Prec. 1st Annual Conf. on Genetic Programming, Stanford University, USA, **327** (1996).
[36] Soto-Campos, A.M., Koshkbarchi, M.K., Vera, J.H., *Fluid Phase Equilib.*, **142**, 193 (1998).
[37] Soto-Campos, A.M., Koshkbarchi, M.K., Vera, J.H., *Fluid Phase Equilib.*, **158**, 893 (1999).