Application of Genetic Algorithm Based Support Vector Machine Model in Second Virial Coefficient Prediction of Pure Compounds

Soleimani Lashkenari, Mohammad*[+]

Faculty of Engineering Modern Technologies, Amol University of Special Modern Technologies, 4616849767 Amol, I.R. IRAN

Mehdizadeh, Bahman

National Iranian South Oil Company, Ahwaz, I.R. IRAN

Movagharnejad, Kamyar

Faculty of Chemical Engineering, Babol University of Technology, Babol, I.R. IRAN

ABSTRACT: *In this work, a Genetic Algorithm boosted Least Square Support Vector Machine model by a set of linear equations instead of a quadratic program, which is improved version of Support Vector Machine model, was used for estimation of 98 pure compounds second virial coefficient. Compounds were classified to the different groups. Finest parameters were obtained by Genetic Algorithm method for training data. The accuracy of the Genetic Algorithm boosted Least Square Support Vector Machine was compared with four empirical equations that are well-known and are claimed can predict all compounds second virial coefficients (Pitzer, Tesonopolos, Gasanov RK and Long Meng). Results showed that in all classes of compounds, the Genetic Algorithm boosted Least Square Support Vector Machine method was more accurate than these empirical correlations. The Average Relative Deviation percentage of overall data set was 2.53 for the Genetic Algorithm boosted Least Square Support Vector Machine model while the best Average Relative Deviation percentage for empirical models (Tesonopolos) was 15.38. When the molecules become more complex, the difference in accuracy becomes sharper for empirical models where the proposed Genetic Algorithm boosted Least Square Support Vector Machine model have predicted good results for classes of compounds that empirical correlations usually fail to give good estimates.*

KEYWORDS: *Second virial coefficient; Prediction; Support vector machine; Genetic algorithm; Optimization.*

INTRODUCTION

Virial expantion, simplified with two terms, is the most useful equation of state for predicting the thermodynamic

***** *To whom correspondence should be addressed.*

+ E-mail: m.soleimani@ausmt.ac.ir

1021-9986/2018/5/189-208 10/$/6.00

properties of real gases for reduced densities up to 0.5 [1]. Coefficients in the virial equation of state are related directly to the interactions between molecules and they can be used for prediction of different properties. The second coefficient of virial expantion is considered to be the most important because it can be used as a tool for calculating the non-ideality of compounds. Virial coefficients are obtained from different methods such as; PVT measurements, relative permittivity measurements, sound's speed measurements, refractive index, Joule–Thomson measurements, vapor pressure, and enthalpy of vaporization measurements [2, 3]. Virial coefficients are strong functions of temperature and vary sharply between different compounds, thus, it is not possible to rely exclusively on experimental data. Because of the importance of second virial coefficient a series of researches have been done to introduce new methods for prediction and application of it. The second virial coefficient has been used for equilibrium calculations [4, 5]. In addition, second virial coefficients can be used for solubility determination [6, 7].

In 1957, *Pitzer* and *coworkers* [8] proposed a proper precision correlation for calculation of the second virial coefficients of non-polar gases. Their work was the basis for several later studies. Then several correlations have been developed to calculate the second virial coefficients. A simplified version of their correlation may be shown as follows:

$$\frac{BP_c}{RT_c} = f^{(0)}\left(\frac{T}{T_c}\right) + \omega f^{(1)}\left(\frac{T}{T_c}\right) \quad (1)$$

In Eq. (1), $f^{(0)}$ presents the main section of the equation for simple fluids where $f^{(1)}$ is a correction function that presents the effect of an acentric factor on the second virial coefficient. They correlated $f^{(0)}$ and $f^{(1)}$ from experimental data to obtain the following equation:

$$\frac{BP_c}{RT_c} = 0.083 - \frac{0.422}{T_r^{1.6}} + \omega\left(0.139 - \frac{0.172}{T_r^{4.2}}\right) \quad (2)$$

Tsonopoulos [9] have modified Pitzer's equation for evaluation of polar compounds second virial coefficient. This correlation with a significant number of constants is able to predict the second virial coefficients of some non-polar and polar compounds, but it could be utilized for limited compounds [2].

$$\frac{BP_c}{RT_c} = f^{(0)}\left(\frac{T}{T_c}\right) + \omega f^{(1)}\left(\frac{T}{T_c}\right) + f^{(2)}\left(\frac{T}{T_c}\right) \quad (3)$$

$$f^{(0)} = 0.1445 - \frac{0.33}{T_R} - \frac{0.1385}{T_R^2} - \frac{0.0121}{T_R^3} - \frac{0.000607}{T_R^8} \quad (4)$$

$$f^{(1)} = 0.0637 - \frac{0.331}{T_R^2} - \frac{0.423}{T_R^3} - \frac{0.008}{T_R^8} \quad (5)$$

$$f^{(2)} = \frac{g(\mu_R)}{T_R^6} - \frac{k(\mu_R)}{T_R^8} \quad (5)$$

In Eq.6 μ_R is reduced dipole momentum.

In 2014, *Nicola* and *coworkers* [10] used An Improved Group Contribution Method for the Prediction of Second Virial Coefficients. In this method, empirical parameters are not used where make it possible to utilize it for a broad range of compounds. In 2016, an accurate and simple semi-emprical correlation was developed by *Nicola et. al.* to correlate the second virial coefficients of refrigerants [11]. In 1988, *Vetere* [12] proposed a reliable model for the evaluation of pure compounds and their mixtures second virial coefficients. This model was derived from a new equation of state. *Vetere* [13, 14] published two revised versions of his original method in 1990 and 1999. It was claimed that the accuracy of these revisions have increased considerably in comparison to the original method.

In 2003, *Mathias* [15] proposed a simple corresponding-states equation for the second virial coefficient at relatively high temperatures (reduced temperatures higher than 0.8). This two-parameter correlation was based upon the Redlich-Kwong equation. The proposed equation is:

$$\frac{BP_c}{RT_c} = 0.08664035 - \frac{0.42748025}{T_r^{1.6}} \quad (7)$$

In 2004, *Meng et al.* [2] published an improved form of the Tsonopoulos [9] correlation. The authors showed that the present correlation is more accurate than the Tsonopoulos' [9] original correlation. Details of their proposed equation were described in their published paper.

Estela-Uribe and *Jaramillo* [16] presented a generalized virial equation of state in 2005 that has a reliable performance for natural gas systems. Their model

was based on the corresponding states principle and involved 12 adjustable coefficients. In 2012, *Oreski* compared the performance of the neural network and empirical models in estimatetion of polar compounds second virial coefficient [4]. His results showed that accurate results were obtained with the neural network model.

Thermodynamic properties of gases could be obtained by the help of statistical mechanics in terms of an intermolecular potential. Based on this approach following equation is given in terms of the intermolecular potential U(r) for evaluation of the second virial coefficient:

$$B(T) = -2\pi N\int_0^{\infty}\left(e^{-\left\{\frac{U(r)}{kT}\right\}}-1\right)r^2dr \qquad (8)$$

Where *k* is Boltzmann's constant, *N* is the Avogadro constant and *r* is the intermolecular distance. Several researchers have tried to predict the second virial coefficient by solving the aformentioned integral, but lack of knowledge of intermolecular potential usually leads to poor predictions.

Vapnik [17] has introduced Support Vector Machine (SVM) which is a relatively novel technique based on Statistical Learning Theory (SLT). SVM is a powerful method for the problems characterized by small samples, nonlinearity, high dimension, and local minima [18, 19]. The Empirical Risk Minimization (ERM) principle is generally employed in classical methods. In SVM, the ERM is replaced by the Structural Risk Minimization (SRM) principle [20]. Finding the final SVM model can be very difficult because a set of nonlinear equations have to be solved (quadratic program). *Suykens* and *Vandewalle* [21] proposed a modified version of SVM called Least Square Support Vector Machine (LSSVM), which resulted in a set of linear equations instead of a quadratic program. Solving this modified version is usually much easier.

LSSVM accuracy is highly dependent on proper parameter setting of it [22]. There are several methods for optimization of parameters of LSSVM. One of the most important methods for optimization is the Genetic Algorithm (GA) technique. The most advantages of the GA is that it simultaneously solve different complicated equations and search for different regions in defined space [23]. In this method, after iterative computations accurate solution is obtainable. We describe this method in detail in the next sections.

Because of the high ability of LSSVM for modeling the nonlinear phenomenon, during the last decade, it has been used for many applications such as prediction of a coal-fired boiler NO x emissio [24], prediction of toxicity of nitrobenzene [25], liquid desiccant dehumidifiers [26], modeling of polyelectrolyte membrane fuel cell [27], modeling of penicillin fermentation process [28], modeling of isomerization of C8 aromatics [29], modeling of different substances solubility in supercritical carbon dioxide [30] and many other applications [31].

In this study, a GA-based LSSVM is used for estimation of 98 pure compounds second virial coefficient. Then, the prediction ability of the GA-based LSSVM method and four well-known empirical correlations (*Pitzer* [8], *Tesonopolos* [9], *RK Gasanov* (RK) and *Long Meng*) was evaluated.

THEORTICAL SECTION

Least squares support vector machine

SVMs, which are based on the Statistical Learning Theory (SLT) and the Structural Risk Minimization (SRM) principle, were introduced by *Vapnic* [17]. The solution is obtained by solving the Quadratic Programming (QP). This method avoids the local minima and provides an advantage over other regression techniques [32]. As it was mentioned before, finding the final SVM model can be very difficult because it requires the solution of a set of nonlinear equations (quadratic program). As a simplification, *Suykends* and *Vandewalle* [21] proposed a modified version of SVM called Least Square Support Vector Machine (LSSVM), which resulted in a set of linear equations instead of a quadratic program. The standard LSSVM algorithm has been described as follows. Given a set of training data like this:

$$\left[(x_1,y_1)\ldots(x_N,y_N)\right]\subset R^N\times R \qquad (9)$$

The following regression model is constructed by using nonlinear mapping function φ(x), which maps the input data to a higher dimensional feature space:

$$y = w^T\cdot\varphi(x)+b \quad \text{with} \quad w\in R^N,\ b\in R \qquad (10)$$
$$\varphi: R^N\rightarrow R^M,\quad M\rightarrow\infty$$

Where *w* is the weight vector and *b* is the bias. When the LSSVM is used as an approximation function, a new optimization problem is created in the case of SRM.

The quadratic loss function is selected in LSSVM. The optimization problem of LSSVM is created as:

$$\mathrm{Min}\, J(w,e) = \frac{1}{2} w^T \cdot w + \frac{1}{2}\gamma \sum_{k=1}^{N} e_k^2 \quad (11)$$

The constraint of these equations is:

$$y = w^T \cdot \varphi(x), + b + e_k \quad k = 1....N \quad (12)$$

Where γ is the regularization parameter that balances the model's complexity and the training error, and e_k is the desired error. In order to solve the constrained optimization problem, a Lagrangian is constructed as:

$$L(w,b,e,\alpha) = J(w,e) - \sum_{k=1}^{N} \alpha_k \left\{ w^T \cdot \varphi(x) + b + e_k - y_k \right\} \quad (13)$$

In this equation α_k is Lagrange multipliers and called support value. The solution of the above equation can be obtained by partially differentiating with respect to each variable.

Genetic algorithm

Genetic Algorithms (GA), to obtain a fast search and optimization technique, use the "survival of the fittest" principle of natural evolution with the genetic propagation of characteristics. The most important aspect of a GA is that it determines many possible solutions simultaneously and explores different regions in the desired space chosen by the user [33].

GA, uses a direct analogy to Darwinian natural selection and genetics in biological systems, is a promising alternative to conventional traditional methods. Based on the Darwinian principle of "survival of the fittest", GA can obtain the optimal solution after a series of iterative computations. The search process is composed of artificial mutation, crossover, and selection [34]. The adjusting processes of GA include three steps:

Chromosome design: in this step, γ and δ^2 arecoded to form the chromosome. The chromosome X was presented as $X=\{p_1, p_2\}$ where p_1 and p_2 are γ and δ^2 in this work, respectively.

Population generation: in this step, the randomly initialized population of possible solutions is generated.

Fitness study: in this step, a fitness function is evaluated. In the present study, the average relative deviation of testing data was used as the fitness function. Steps of the GA learning algorithm are detailed in the literature [35].

These three steps generate a new population of possible solutions, which as compared to the previous population; usually leads to better satisfying the optimization objective. The best string obtained after repeating the above described loop forms the solution to the optimization problem.

Preparation of data set

Second virial coefficient data of 98 compounds were collected from literature in order to asset the abilities of GA-LSSVM model in prediction of the second virial coefficient [36]. The compounds were classified to the different groups such as paraffin, aldehydes, non-polar compounds, ethers, amines, heterocyclics with N, esters, alcohols, and phenols, ketones and nitrils, and non-cyclic hydrocarbon with F, Cl, Br, I.

Investigating the results of published literature, 9 parameters (temperature, molecular weight, critical temperature, number of carbon in the compound, number of carbon branch, density, critical pressure, and normal boiling temperature) were devoted to the model as inputs.

In Table 1, Input variables and data range of GA-LSSVM model are presented. It is important to realize that none of the experimental data from the literature were excluded and all published data were considered correct.

All data were normalized to prevent the larger number from overriding smaller ones[37]. Generally, normalization can be done by different equations. In this research, data were scaled between [0.1-0.9] by means of Eq. 14.

$$\mathrm{Scaled}_{value} = 0.1 + \frac{(\mathrm{Unscaled})_{value} - \min_{(\mathrm{unscaled\ value})}}{\max_{(\mathrm{unscaled\ value})} - \min_{(\mathrm{unscaled\ value})}} \times 0.8 \quad (14)$$

Optimization of LSSVM based on GA

In present work, total available data were randomly divided into two parts: 65% of data (as training data) and 35 % data (as test data). This dividing was down randomly several times to reach a condition in which there was no local accumulation of training or testing data. The purpose was to acquire a uniform distribution fo training and testing data. To asset accuracy of model,

Table 1: Input data range of GA-LSSVM model.

No.	Input parameter	Unit	Data range
1	Molecular weight	g/mol	16-338
2	Critical temperature	K	190.56-694.20
3	Critical pressure	MPa	1.83-8.08
4	Acentric factor	-	0.011-0.665
5	Number of carbon in the compound	-	1-8
6	Number of carbon branch	-	0-3
7	Density	g/ml	0.001-2.28
8	Boiling Point	K	111.30-454.40
9	Temperature	K	110.83-673.31

for each compound, Average Relative Deviation (ARD) of testing data was used (which is calculated by means of Eq. (15)).

$$\mathrm{ARD} = \frac{100}{\mathrm{N}} \times \sum_{\mathrm{i}=1}^{\mathrm{N}} \left| \frac{\mathrm{y}^{\mathrm{exp}} - \mathrm{y}^{\mathrm{cal}}}{\mathrm{y}^{\mathrm{exp}}} \right| \qquad (15)$$

Where y^{exp} is experimental data, y^{calc} is the calculated second virial coefficient, and N is the number of experimental data.

By the object of minimization of total ARD on training and testing dataset, Radial Basis Function (RBF) was used for GA-LSSVM method.

According to the published literature, γ and δ^2 can take any value between 0-4000 and 0-100 respectively. The parameter setting of GA-LSSVM model was carried out by LSSVMLab 1.6 free toolbox and GA Toolbox of MATLAB R2008b. The general architecture of GA-LSSVM model and GA optimization process of LSSVM parameters is presented in Fig. 1. At first, an initial population will be set. This initial population contains random solutions including random values for γ and δ^2. Then, this initial population will undergo evolution until termination criteria are achieved. The cost function as the criteria for selecting the best individual solution is the ARD. In each generation, the population will undergo the GA operators, the main of which are cross over and mutation. In the end, the best solution will be selected.

RESULTS AND DISCUSSION

The GA-LSSVM was utilized to predict the second virial coefficient of pure compounds. The first part of data (65%) was used for the training and optimization of model parameters. After training and optimization of GA-LSSVM parameter, its capability for prediction of pure compounds second virial coefficient was checked. In this regard, the cross-plot figure was utilized to evaluate the performance of the suggested architecture based on the testing and training data (Fig. 2).

The distance between each point and the diagonal line is a criterion for prediction precision. In this regard, models with a high rate of accumulation near the diagonal line are considered more accurate As Fig. 2 indicates, points show training and testing data and diagonal of the Fig. 2 is the locations of exact predictions.

Accuracy of GA-LSSVM model can be compared with empirical models. The accuracy of GA-LSSVM model presented in this study as well as the ones previously discussed for estimating the second virial coefficient of compounds is given in Table 2. This table shows the calculations of the Average Relative Deviation (ARD) for GA-LSSVM and empirical models. Fig. 3 also shows ARD of GA-LSSVM model for different classes of compounds.

Total ARD of GA-LSSVM and Pitzer, Tesonopolos, RK, and Long Meng methods are depicted in Fig. 4.

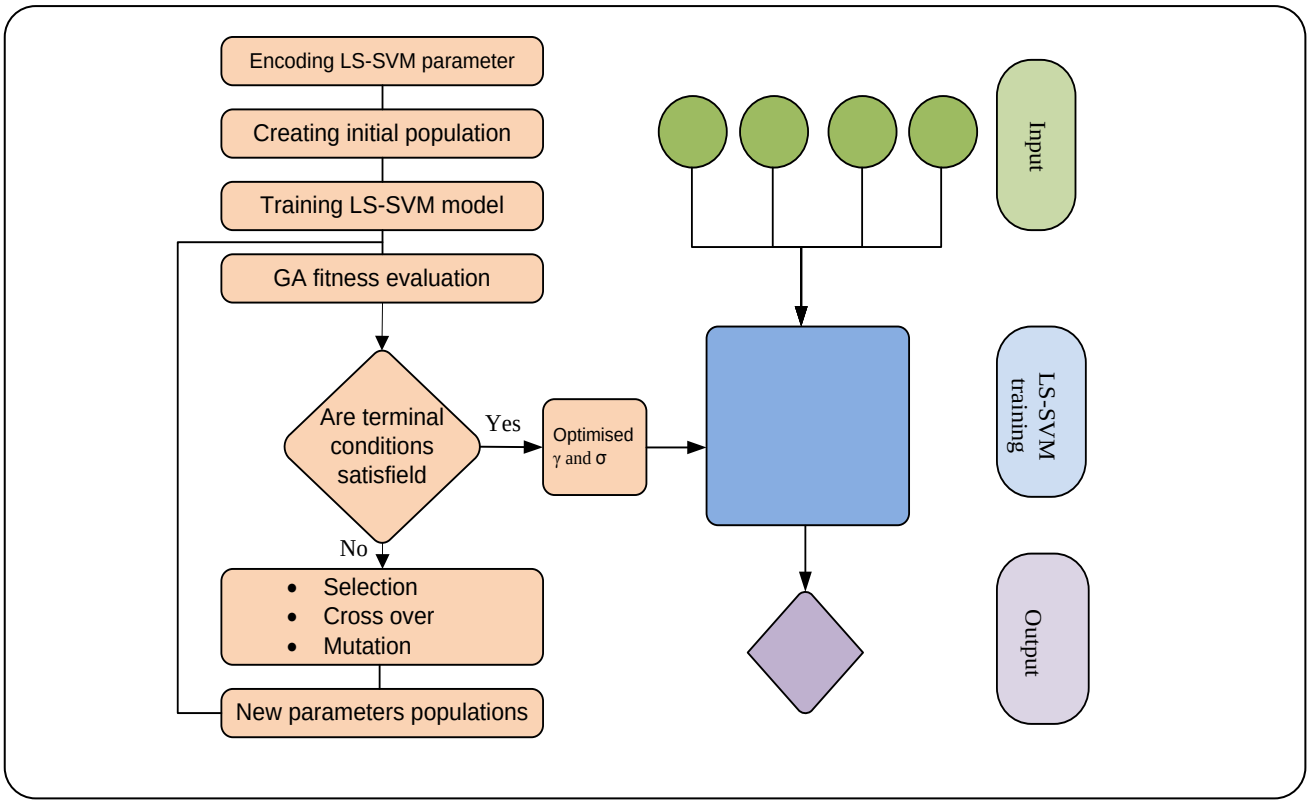

Fig. 1: General architecture of LSSVM model and GA optimization process of LSSVMmodel parameters.

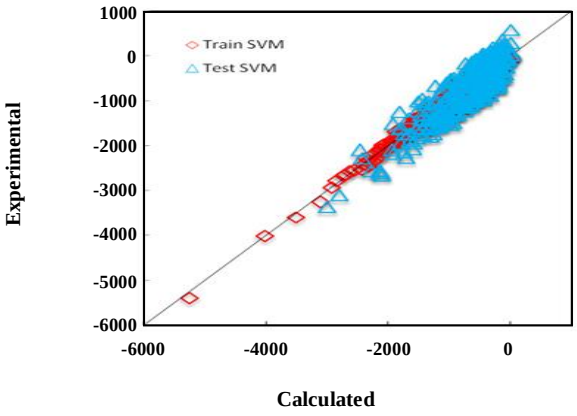

Fig. 2: Comparison of experimental and calculated values for testing and training data.

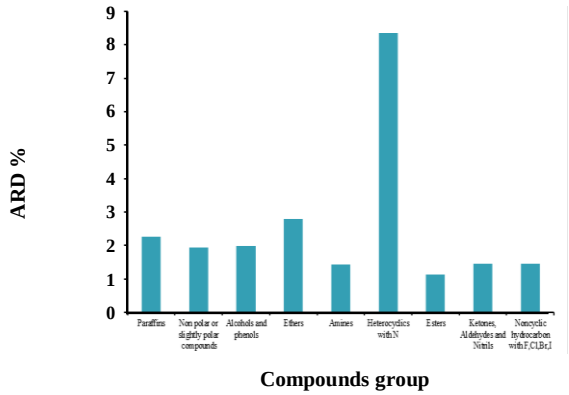

Fig. 3: ARD of GA-LSSVM method for differentclasses of compounds.

Results showed that in all classes of compounds, GA-LSSVM method was more accurate than these four empirical equations. Detailed results are shown in Table 2 and Fig. 4. It is obvious that the present study is more accurate and the accuracy of four empirical correlations is nearly equal. It may also be deduced that as the molecular structure of the compounds become more complex the accuracy of GA-LSSVM method over traditional correlations becomes more evident. For example, compare the results for group two (non-polar or slightly-polar compounds) and group 3 (alcohols and phenols). It is clear that the accuracy GA-LSSVM method for these two groups are almost the same (ARD=1.94% for the second group and 1.99% for the third group), while the accuracy of the empirical correlations differ widely. For example, Pitzer correlation predicts the second virial coefficient of

Table 2: Comparing the accuracy of GA-LSSVM and empirical models for differentclasses of compounds.

No.	Compound Group	Method ARD%				
		GA-LSSVM	Pitzer	Tsonopoloulos	Long Meng	RK
1	Paraffins	2.26	5.52	4.86	4.68	9.28
2	Non polar or slightly polar compounds	1.94	5.76	6.33	5.42	8.77
3	Alcohols and phenols	1.99	15.23	13.88	11.99	21.83
4	Ethers	2.80	12.95	14.30	10.08	10.67
5	Amines	1.42	11.35	11.58	11.11	14.35
6	Heterocyclics with N	8.35	26.78	26.31	29.85	31.08
7	Esters	1.12	9.21	9.71	21.78	15.94
8	Ketones, Aldehydes and Nitrils	1.46	40.60	40.89	42.53	46.73
9	Noncyclic hydrocarbon with F,Cl,Br,I	1.46	11.24	10.56	10.73	11.48
Average ARD%		2.53	15.40	15.38	16.46	18.90

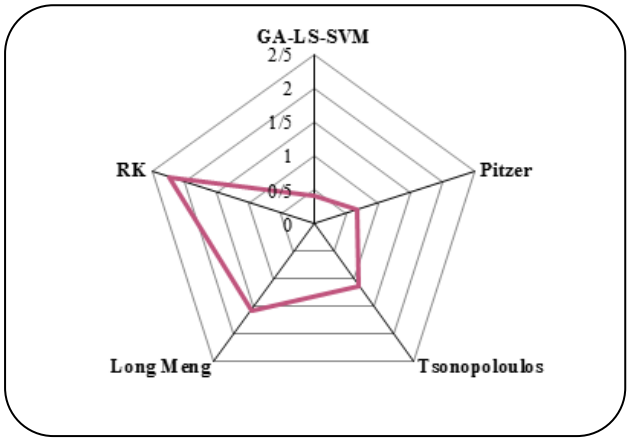

Fig. 4: ARD of GA-LS-SVM method and four empirical equations.

the second group with an ARD of 5.76%, while the ARD for the third group rises sharply to 15.23%. The ARDs of the Tsonopoloulos correlation also shows a similar trend. The results for halogenated noncyclical hydrocarbons (group 9) also clearly show that for these compounds the current work predicts the second virial coefficient with much more accuracy. It is evident that for simpler molecules the difference in accuracy is not so sharp. For example, the ARD of paraffin group for the present work is found to be 2.26% and it is clearly more accurate than *Pitzer* (ARD=5.52%), Tsonopoloulos (ARD=4.86%), Long Meng (4.68%), and RK (9.28%).

The performance of the GA-LSSVM model and four well-known empirical models (Pitzer, Tesonopolos, RK, and Long Meng) was compared in Fig. 5. It is shown in Fig. 5 that results of GA-LSSVM model are the least scattered around the reference 45º line. Our results are consistent with the finding of Oreski [4], which pointed neural network model is accurate than semi-empirical models such as Tesonopolos for prediction of the second virial coefficient.

CONCLUSIONS

In summary, we have presented a new technique, GA-LSSVM, to estimate the second virial coefficient of 98 compounds in different classes. The GA-LSSVM model results were compared to Pitzer, Tesonopolos, RK, and Long Meng models based on the 980 experimental data points from the literature. The calculations show that the present work is proved to be more accurate than traditional empirical correlations. The ARD% of the overall data set is 2.53 for the GA-LSSVM model while the best ARD% for empirical models (Tesonopolos) is 15.38. For empirical models, the difference in accuracy becomes sharper as the molecules become more complex. The proposed GA-LSSVM model predicts good results for classes of compounds that empirical correlations usually fail to give good estimates. For example, Pitzer correlation predicts the second virial coefficient of the second group with an ARD of 5.76%, while the ARD for

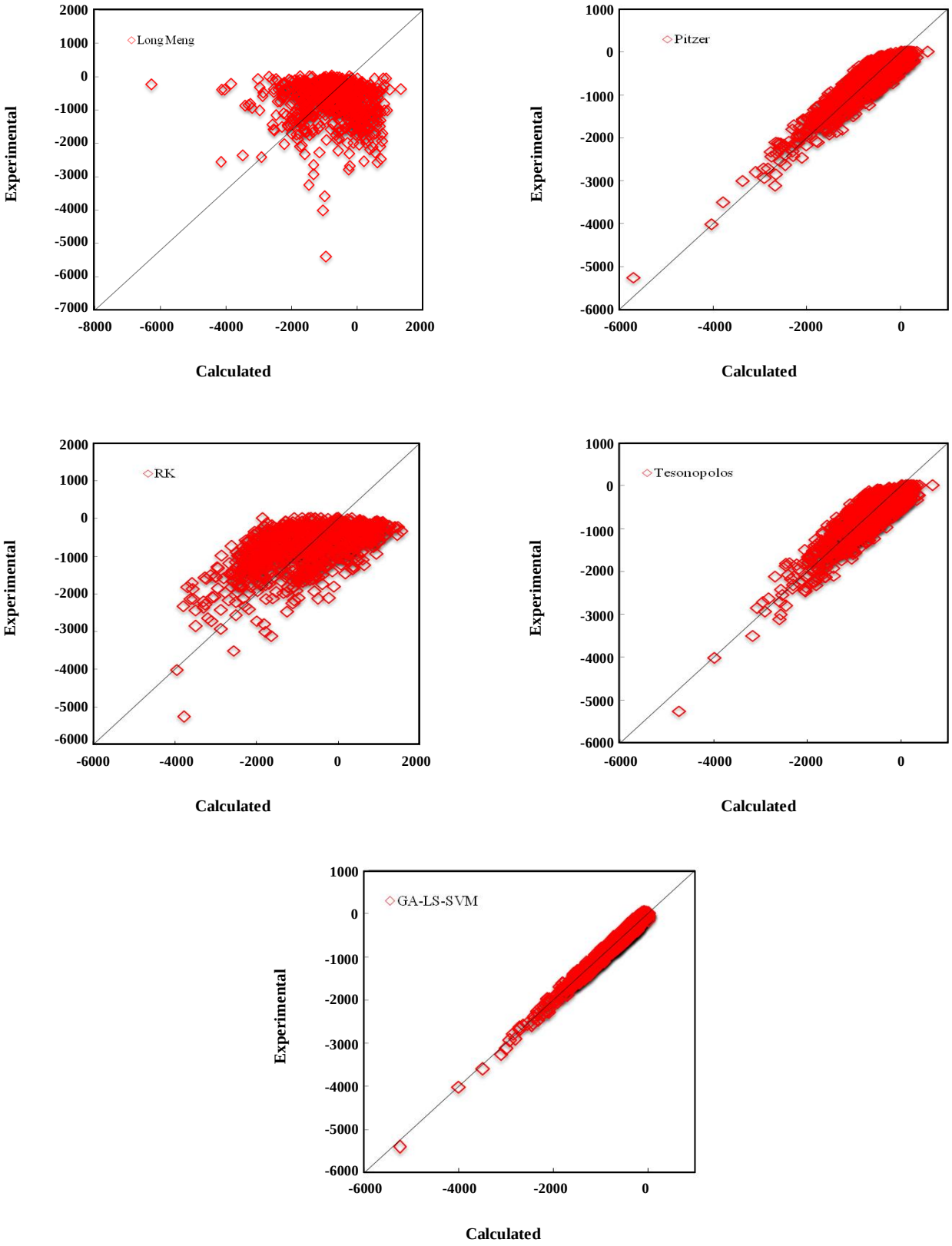

Fig. 5: Cross plot of GA-LSSVM and empirical models.

the third group rises sharply to 15.23% where GA-LS_SVM model predicts the second virial coefficient of both groups with ARD of 1.94 and 1.99.

Received : Sep. 5, 2016 ; Accepted : Oct. 30, 2017

REFERENCES

[1] Meng L., Duan Y.Y., Prediction of the Second Cross Virial Coefficients of Nonpolar Binary Mixtures, *Fluid Phase Equilib.*, **238**(2): 229-238 (2005).

[2] Meng L., Duan Y.Y., Li L., Correlations for Second and Third Virial Coefficients of Pure Fluids, *Fluid Phase Equilib.*, 226(10): 109-120 (2004).

[3] Rashtchian D., Ovaysi S., Taghikhani V., Ghotbi C., Application of the Genetic Algorithm to Calculate the Interaction Parameters for Multiphase and Multicomponent Systems, *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, **26**(3): 89-102 (2007).

[4] Oreški S., Comparison of Neural Network and Empirical Models for Prediction of Second Virial Coefficients for Gases, *Procedia Engineering*, **42**(1): 303-312 (2012).

[5] Rakel N., Bauer K.C., Galm L., Hubbuch J., From Osmotic Second Virial Coefficient (B22) to Phase Behavior of a Monoclonal Antibody, *Biotechnol. Progr.*, **31**(2): 438-451 (2015).

[6] Tomberli B., Goldman S., Gray C., Predicting Solubility in Supercritical Solvents Using Estimated Virial **Coefficients** and Fluctuation Theory, *Fluid Phase Equilib.*, 187: 111-130 (2001).

[7] Herhut M., Brandenbusch C., Sadowski G., Modeling and Prediction of Protein Solubility Using the Second Osmotic Virial Coefficient, *Fluid Phase Equilib.*, **422**(25): 32-42 (2016).

[8] Pitzer K.S., Curl Jr R., The Volumetric and Thermodynamic Properties of Fluids. III. Empirical Equation for the Second Virial Coefficient1, *J. Am. Chem. Soc.*, 79(10): 2369-2370 (1957).

[9] Tsonopoulos C., An empirical Correlation of Second Virial Coefficients, *AIChE Journal*, **20**(2): 263-272 (1974).

[10] Di Nicola G., Falone M., Pierantozzi M., Stryjek R., An Improved Group Contribution Method for the Prediction of Second Virial Coefficients, *Industrial & Engineering Chemistry Research*, **53**(35): 13804-13809 (2014).

[11] Di Nicola G., Coccia G., Pierantozzi M., Falone M., A Semi-Empirical Correlation for the Estimation of the Second Virial Coefficients of Refrigerants, *International Journal of Refrigeration*, **68**: 242-251 (2016).

[12] Vetere A., An Empirical Model for Calculating Second Virial Coefficients, *Chem. Eng. Sci.*, **43**(12): 3119-3127 (1988).

[13] Vetere A., On the Calculation of the Second Virial Coefficients of Pure Compounds and Mixtures, *Chem. Eng. Sci.*, **46**(7): 1787-1794 (1991).

[14] Vetere A., An Improved Method to Predict the Second Virial Coefficients of Pure Compounds, *Fluid Phase Equilib.*, **164**(1): 49-59 (1999).

[15] Mathias P.M., The Second Virial Coefficient and the Redlich-Kwong Equation, *Indu. Eng. Chem. Res.*, **42**(26): 7037-7044 (2003).

[16] Estela-Uribe J., Jaramillo J., Generalised Virial Equation of State for Natural Gas Systems, *Fluid Phase Equilib.*, **231**(1): 84-98 (2005).

[17] Vapnik V., "Statistical Learning Theory", Wiley, New York, (1998).

[18] Torabi Dashti H., Masoudi-Nejad A., Zare F., Finding Exact and Solo LTR-Retrotransposons in Biological Sequences Using SVM, *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, **31**(2): 111-116 (2012).

[19] Peng Z., Yin H., ECT and LS-SVM Based Void Fraction Measurement of Oil-Gas Two-Phase Flow, *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, **29**(1): 41-50 (2010).

[20] Adankon M.M., Cheriet M., Support Vector Machine, *Encyclopedia of Biometrics*, 1504-1511 (2015).

[21] Suykens J.A.K., Vandewalle J., Least Squares Support Vector Machine Classifiers, *Neural process. lett.*, **9**(3): 293-300 (1999).

[22] Yang Z., Gu X., Liang X., Ling L., Genetic Algorithm-Least Squares Support Vector Regression Based Predicting and Optimizing Model on Carbon Fiber Composite Integrated Conductivity, *Materials & Design*, **31**(3): 1042-1049 (2010).

[23] Afshari S., Aminshahidy B., Pishvaie M.R., Well Placement Optimization Using Differential Evolution Algorithm, *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, **34**(2): 109-116 (2015).

[24] Lv Y., Liu J., Yang T., Zeng D., A Novel Least Squares Support Vector Machine Ensemble Model for NOx Emission Prediction of a Coal-Fired Boiler, *Energy*, **55**(15): 319-329 (2013).

[25] Niazi A., Jameh-Bozorghi S., Nori-Shargh D., Prediction of Toxicity of Nitrobenzenes Using ab Iinitio and Least Squares Support Vector Machines, *J. Hazard. Mater.*, **151**(2): 603-609 (2008).

[26] Zendehboudi A., Tatar A., Li X., A comparative Study and Prediction of the Liquid Desiccant Dehumidifiers Using Intelligent Models, *Renewable Energy*, **114**: 1023-1035 (2017).

[27] Li C.H., Zhu X.J., Cao G.Y., Sui S., Hu M.R., Identification of the Hammerstein Model of a PEMFC Stack Based on Least Squares Support Vector Machines, *J. Power Sources*, **175**(1): 303-316 (2008).

[28] Wang X., Chen J., Liu C., Pan F., Hybrid Modeling of Penicillin Fermentation Process Based on Least Square Support Vector Machine, *Chem. Eng. Res. Des.*, **88**(4): 415-420 (2010).

[29] Li L., Su H., Chu J., Modeling of Isomerization of C8 Aromatics by Online Least Squares Support Vector Machine, *Chin. J. Chem. Eng.*, **17**(3): 437-444 (2009).

[30] Mehdizadeh B., Movagharnejad K., A Comparative Study between LS-SVM Method and Semi Empirical Equations for Modeling the Solubility of Different Solutes in Supercritical Carbon Dioxide, *Chem. Eng. Res. Des.*, **89**(11): 2420-2427 (2011).

[31] Torabi Dashti H., Masoudi-Nejad A., Mining Biological Repetitive Sequences Using Support Vector Machines and Fuzzy SVM, *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, **29**(4): 1-17 (2010).

[32] Li J., Liu H., Yao X., Liu M., Hu Z., Fan B., Structure-Activity Relationship Study of Oxindole-Based Inhibitors of Cyclin-Dependent Kinases Based on Least-Squares Support Vector Machines, *Anal. Chim. Acta*, **581**(2): 333-342 (2007).

[33] Tayyebi S., Shahrokhi M., Bozorgmehry Boozarjomehry R., Fault Diagnosis in a Yeast Fermentation Bioreactor by Genetic Fuzzy System, *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, **29**(3): 61-72 (2010).

[34] Zhang Y., An Improved QSPR Study of Standard Formation Enthalpies of Acyclic Alkanes Based on Artificial Neural Networks and Genetic Algorithm, *Chemometr. Intell. Lab.*, **98**(2): 162-172 (2009).

[35] Sastry K., Goldberg D.E., Kendall G., "Genetic Algorithms, In: Burke E., Kendall G., editors Search Methodologies", Springer, Boston, MA (2014).

[36] Dymond J.H., Marsh K.N., Wilhoit R.C., Wong K.C., Frenkel M.D., "Virial coefficients of pure gases In: Frenkel M, Marsh KN, editors. Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology, Group 4, Vol. 21, Subvolume A ", Springer-Verlag, New York (2002).

[37] Mehdizadeh B., Movagharnejad K., A Comparison between Neural network Method and Semi Empirical Equations to Predict the Solubility of Different Compounds in Supercritical Carbon Dioxide, *Fluid Phase Equilib.*, **303**(1): 40-44 (2011).