Fluid Phase Equilibria 446 (2017) 1–8

Contents lists available at ScienceDirect

Fluid Phase Equilibria

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

Predicting heat capacity of gas for diverse organic compounds at different temperatures

Jingchen Yin [a], Qingzhu Jia [a, *], Fangyou Yan [b], Qiang Wang [b]

[a] School of Marine and Environmental Science, Tianjin University of Science and Technology, 13St. 29, TEDA, 300457 Tianjin, PR China
[b] School of Chemical Engineering and Material Science, Tianjin University of Science and Technology, 13St. 29, TEDA, 300457 Tianjin, PR China

ARTICLE INFO

Article history:
Received 20 March 2017
Received in revised form
28 April 2017
Accepted 5 May 2017
Available online 6 May 2017

Keywords:
Heat capacity of gas
Thermodynamic property
Variable temperatures
QSPR
Norm index

ABSTRACT

In this work, a model was proposed for predicting heat capacity of gas (C_p) for diverse organic compounds including aromatic, acyclic aliphatic hydrocarbons, ketones and acetates at variable temperatures, range from 50 to 1500 K. At the same time, a new norm descriptor based on molecular characteristic matrix was developed. Validation results showed that this norm-index-based model had satisfactory performance for predicting C_p at various temperatures with high squared correlation coefficient R^2 of 0.9976 and high squared relation coefficient of the cross validation Q^2 of 0.9976. The results of external validation, $R^2_{training}$ and R^2_{test}, were 0.9977 and 0.9970, respectively. Meanwhile, Y-randomization test and mean absolute errors test could indicate the stability and accuracy of this model. In addition, comparison results could further demonstrate the high predictive performance and the wide applicability domain of this work; and the big strength of this work is that it could predict gas heat capacity down to 50 K.

1. Introduction

Heat capacity (C_P) is considered as a measure of how well the substance stores heat [1]. Being as a crucial property in thermodynamics [2], C_P values in the gaseous phase at variable temperatures are extremely important for the estimation of the entropy at any other temperature [3]. Moreover, heat capacities of gas have often been applied to the study of solute–solvent interactions through the heat capacities of solvation [4]. Though lots of experimental data of heat capacity of gas for many common organic compounds at 298.15 K or at different temperature ranges (very limited) could be desirable, due to the increased discovery or synthesis of new chemical compounds, such data would be usually in urgent need, especially the temperature-dependent C_P values and their availability [1].

Quantitative structure–property relationship (QSPR) is an efficient tool in the correlation and prediction of diverse thermodynamic properties of compounds from their structures. For instance, Toropov et al. modeled the Gibbs free energy [5], alkane enthalpies [6] and enthalpies of formation for organometallic compounds [7] by means of QSPR at constant temperature. Also, it is necessary to predict the heat capacity of gas especially at broad range of temperatures of chemical compounds by using theoretical approaches, and several models based on QSPR have been successfully established to calculate the values of gas heat capacity. For example, Xue et al. [1] utilized multiple linear regression, radial basis function networks and support vector machine methods to develop QSPR models for predicting gas heat capacity of 182 compounds at 298.15 K. Khajeh et al. [8] established a QSPR model for the prediction of heat capacity of gas covered a diverse set of 1174 compounds at 298.15 K by utilizing nonlinear genetic function approximation and adaptive neuro-fuzzy inference system method. Though the work of Xue et al. [1] and Khajeh et al. [8] could give satisfying prediction results, both works predicted the values of heat capacity of gas only at 298.15 K.

As a matter of fact, heat capacity of gas is usually strongly affected by the temperature. Hence, it is more important and valuable to predict heat capacity of gas at different temperatures. Group contribution methods have often been widely utilized for C_p prediction at various temperatures, such as the first-order group contribution method [9,10]_ENREF_8, the second-order method [11–13]_ENREF_10_ENREF_10. Recently, Mondejar et al. [14] predicted ideal gas C_p of organic fluids containing chlorine and/or fluorine at 300 K and 400 K using two group contribution

---

* Corresponding author.
E-mail address: jiaqingzhu@tust.edu.cn (Q. Jia).

http://dx.doi.org/10.1016/j.fluid.2017.05.006
0378-3812/

approaches. Although these group contribution methods are simple and able to be widely used, these methods could not be utilized to predict the heat capacity of gas of compounds that contained the groups for which contributions were unavailable. Also, many other methods have been reported for predicting gas C_p at different temperatures [4,15–18], like the density functional theory (DFT) in quantum chemistry combined with statistical thermodynamic theory [16] and approximating correlation [17]. Based on DFT and statistical thermodynamic, Cervinka et al. [4] proposed a new wavenumber and bond-dependent set of scale factors for calculation of ideal-gas C_p and entropies for a set of 93 rigid molecules. Lastovka et al. [19] proposed a similarity variable and then developed correlations with the values of heat capacity of gas. Although, this model was established based on large data set and had good accuracy, it had too many fitting parameters. Very recently, Bruel et al. [2] developed a model for predicting heat capacity of gas at different temperature ranges. Though high prediction accuracy had been obtained, in their work, yet only 14 compounds were involved and its samples set was too small, therefore, it might be difficult for their model to be further stretched for the efficient prediction of other kinds of chemicals.

Our previous works suggested that molecular characteristic matrix-based norm descriptors proposed in our group, could effectively describe the structure-property relationship of organics and, based on which some QSPR models had been successfully developed for predicting aquatic toxicity [20], affinity of 5-HT$_{1A}$ receptor ligands [21] and solubility of fullerene C$_{60}$ in organic solvents [22]. These results demonstrated that the method proposed by our group had high accuracy and might be widely used.

Therefore, in order to predict the heat capacity of gas for a diverse set of organic compounds at a wide temperature range, a model based on various chemical structure information was proposed. Also, a new molecular characteristic matrix norm descriptor based on the optimized structure of diverse organic compounds was developed in this work. Meanwhile, leave-ten-out cross validation and Y-randomization approaches were used to evaluate the accuracy and stability of this model.

2. Method

2.1. Data of heat capacity of gas

A data set of 2028 heat capacity data points at constant pressure under a wide temperature range (50–1500 K) for 114 diverse organic compounds (shown as in Supplementary Material Table S1) was collected from the gas phase thermochemistry data of the National Institute of Standards and Technology (NIST) Standard Reference Database 69: NIST Chemistry WebBook [23]. The molecular weight range of these organic compounds was from 86.13 to 398.45; organic structures involved in this work included aromatic hydrocarbons, acyclic aliphatic hydrocarbons, ketones and acetates. Here, it should be noticed that experimental heat capacity values obtained through direct (calorimetric) or indirect measurements (for example speed-of-sound measurements) are very limited. And most of the heat capacity data could be obtained by methods of statistical thermodynamics using spectral data [24], especially for aromatic aliphatic hydrocarbons. Therefore, in this work, gas heat capacity values recommended by NIST were used. Some of the gas heat capacity values could be verified as experimental values, and other gas heat capacity values recommended by NIST were calculated by reliable methods.

2.2. Model proposed

In order to obtain the most stable molecular structures, the HyperChem 7.0 Software [25] was used to optimize the structure of chemicals in the data set. In the process of optimization, each molecule in the data set was energetically minimized using *ab initio* method in quantum chemistry at ST0-3G level.

In this work, based on the molecular chemical graphs, the adjacent matrix and a matrix deduced from Euclidean spatial distance of the molecules were firstly generated, in which only the route and distance between atom and atom were taken into account, and the type of atom and bond had been neglected.

The two matrices (M_m, m= (a, d)) mentioned above were described as following:

$$M_a = (a_{ij}) \qquad a_{ij} = 1 \qquad \text{atom } i \text{ and } j \text{ are adjacent} \tag{1}$$

$$M_d = \begin{bmatrix} d_{ij} \end{bmatrix} \qquad d_{ij} = \begin{cases} \dfrac{1}{r_{ij}} & i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

r_{ij} is the Euclidean spatial distance between atom i and j

In order to distinguish atomic types, a property matrix (M_E) was further defined, and some atoms' basic properties as van der Waals radius, atom weight, molecular weight and atom charge were distinctively included.

$$M_E = \begin{bmatrix} ai \\ amr \\ ac \end{bmatrix}^T \tag{3}$$

$$ai = atomic\ intension = van\ der\ Waals\ radius \times atom\ weight \tag{4}$$

$$amr = atomic\ mass\ ratio = atom\ weight/molecular\ weight \tag{5}$$

$$ac = atom\ charge \tag{6}$$

Here, the descriptor $ai = atomic\ intension = van\ der\ Waals\ radius \times atom\ weight$, was newly proposed in this work.

Based on the above matrices (M_a, M_d and M_E), three new matrices (M_x, x= (1, 2, 3)) were produced by the adjustment as follows:

$$M_1 = \{[M_E(:, n), M_m]\} \qquad m = (a,\ d); n = (1, 2, 3) \tag{7}$$

$$M_2 = \left\{\left[M_E(\,:, n) \times M_E(\,:, n)^T\right] + M_m\right\} \qquad m = (a,\ d); n = (1, 2, 3) \tag{8}$$

$$M_3 = \left\{\left[M_E(\,:, n) \times M_E(\,:, n)^T\right] . \times M_m\right\} \qquad m = (a,\ d); n = (1, 2, 3) \tag{9}$$

Then, the descriptors were calculated based on M_1, M_2 and/or M_3 by using norm indexes-norm $(M, 1)$ and norm (M, fro), which were defined as follows:

$$norm(M, 1) = \max\left\{\sum |m_{i1}|, \sum |m_{i2}|, \ldots\ldots, \sum |m_{ij}|\right\} \tag{10}$$

$$norm(M, fro) = \left(\sum\sum m_{ij}^2\right)^{1/2} \tag{11}$$

Here, norm (M, fro) is the Frobenius norm of the M matrix.

Aimed at simplifying the model, only the most determinative