Research Article

An Accurate and Efficient Method to Predict Y-NO Bond Homolysis Bond Dissociation Energies

Hong Zhi Li,^{1,2} Lin Li,¹ Zi Yan Zhong,¹ Yi Han,¹ LiHong Hu,¹ and Ying Hua Lu^{1,3}

¹ School of Computer Science and Information Technology, Northeast Normal University, Changchun 130017, China

² School of Life Sciences, Northeast Normal University, Changchun 130024, China

³ Institute of Functional Material Chemistry, Faculty of Chemistry, Northeast Normal University, Changchun 130024, China

Correspondence should be addressed to LiHong Hu; lhhu@nenu.edu.cn and Ying Hua Lu; luyh@nenu.edu.cn

Received 6 June 2013; Accepted 23 July 2013

Academic Editor: William Guo

Copyright © 2013 Hong Zhi Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The paper suggests a new method that combines the Kennard and Stone algorithm (Kenstone, KS), hierarchical clustering (HC), and ant colony optimization (ACO)-based extreme learning machine (ELM) (KS-HC/ACO-ELM) with the density functional theory (DFT) B3LYP/6-31G(d) method to improve the accuracy of DFT calculations for the Y-NO homolysis bond dissociation energies (BDE). In this method, Kenstone divides the whole data set into two parts, the training set and the test set; HC and ACO are used to perform the cluster analysis on molecular descriptors; correlation analysis is applied for selecting the most correlated molecular descriptors in the classes, and ELM is the nonlinear model for establishing the relationship between DFT calculations and homolysis BDE experimental values. The results show that the standard deviation of homolysis BDE in the molecular test set is reduced from 4.03 kcal mol⁻¹ calculated by the DFT B3LYP/6-31G(d) method to 0.30, 0.28, 0.29, and 0.32 kcal mol⁻¹ by the KS-ELM, KS-HC-ELM, and KS-ACO-ELM methods and the artificial neural network (ANN) combined with KS-HC, respectively. This method predicts accurate values with much higher efficiency when compared to the larger basis set DFT calculation and may also achieve similarly accurate calculation results for larger molecules.

1. Introduction

In the past few decades, quantum chemistry has attracted remarkable attention and made significant improvements along with the corresponding fields of physics, mathematics, and computer science [1–7]. Quantum chemical methods have captured the physical essence of molecules through numerous research efforts, to the point where various molecular properties (including geometry, dipole moment, polarizability, thermodynamics, and excited states) can be obtained without experiments. However, quantum chemical calculations for large molecules can be quite demanding, in particular for high-accuracy calculations that need to solve the Schrödinger equation thoroughly. Therefore, accuracy often has to be sacrificed for practical reasons, for example, introducing approximations when solving the wave function for large molecular systems. Despite these approximations, the size of molecule is still limited. For a large system with a regular structure, the quantum chemical methods can only give a qualitative description. For complex systems, it is hard

to calculate with existing quantum chemical methods; even if the calculation can be carried out, the accuracy might be very low and the error is far beyond the specifications in scientific research [8, 9]. Therefore, alternative methods for large or complex systems need to be developed to meet the challenges of molecular calculations.

Artificial intelligence methods can resolve the accuracy problem of quantum chemical calculations for molecular bond properties by establishing the relationship between the calculated and experimental values instead of strictly solving wave functions with very high cost. By this method, high-precision calculation results can be obtained without wasting time on advanced quantum chemical methods and ultra-large basis sets because this combination-type calculation method can take mutual advantages of two methods to reduce the systematic errors induced by defects of theories and functions with simple physical parameters adopted in artificial intelligence methods. Thus quantum chemical calculations with low-level methods and small basis sets are sufficient for obtaining valid data from artificial intelligence methods.

Meanwhile, small basis sets enable quantum chemical calculations to apply to larger molecules, further increasing the chance to design novel lead molecules.

Although the artificial intelligence method is a relatively new multidisciplinary research field, in the past decade, it has received a lot of attention and thus has developed rapidly. The combination strategy consists of the quantum chemical method first being used to obtain the molecular properties and then take the quantum chemical calculation results as inputs for the statistical methods to establish the relationship between the experimental and calculation values. There are many choices on statistical methods including linear methods such as linear regression [10–14], nonlinear methods such as neural networks [15–37], and support vector machines [38–40]. Usually, the linear regression method is simple and intuitive, but problems in molecular systems are mainly nonlinear; for the same descriptor inputs, neural networks can better solve complicated nonlinear problems that might be too difficult to be modeled by mathematical equations [27]. However, the traditional neural networks have the following major drawbacks. (1) The speed of training is slow. The gradient descent method generally takes multiple iterations to converge, so the training process takes a long time. (2) Because it is easy to fall into a local minimum, the global minimum can be missed during the calculation. (3) The selection of learning rate η is sensitive. The learning rate η has a great impact on the performance of the neural network, so the proper η must be chosen to obtain a stable network. If η is too small, the rate of algorithm convergence is very slow and the training process takes a long time; on the other hand, if it is too large, the training process might be unstable (divergent). Therefore, it is important to explore a training algorithm where the learning rate is comparably high, which increases the possibility of acquiring the global optimum solution. This leads to better generalization for the prediction. The extreme learning machine (ELM) method avoids these defects, thus exhibiting a better performance than neural networks [3, 4].

Nitric oxide (NO) plays an important physiological role in the human life cycle [41–50]. It is difficult to accurately obtain homolysis BDE values in NO carrier molecules due to the complexity of the experiments. Professor Cheng's group recently devoted considerable efforts to measuring Y-NO bond homolysis BDE in solvents [51–61], which provides a great opportunity for designing NO molecule carriers in silico. Hence, we herein use artificial intelligence and DFT to achieve high-accuracy homolysis BDE calculation results to better design potentially novel NO carriers.

The content of this paper is arranged as follows. First, it describes the HC, ACO, and ELM methods. Second, it illustrates the technology roadmap of the KS-HC/ACO-ELM method. Third, it discusses the calculation results of the Kenstone, HC, ACO, and ELM methods (i.e., the appropriate classification of the training set and the test set, clustering analyses for molecular descriptors with both the known and unknown class numbers, and the establishment of the ELM model, resp.). Finally, the results and the discussions are summarized.

2. Methods

2.1. Hierarchical Clustering (the Known Number). The basic idea of the hierarchical clustering (HC) method is to classify samples in terms of distances. Firstly, it defines the distance between samples and classes and combines the nearest two classes into classes of samples, and then it recalculates the distance between the new class and other classes and classifies them according to the minimum distance. By reducing one class at a time, this process is repeated until all of the samples come into one class. The chart showing the clustering process is called a clustering chart. One of the features of the HC method is that when samples are classified into a class in a certain level, they will always belong to the same class in the subsequent divisions.

There are a variety of methods to define the distance between classes. Different definitions generate different HC analysis methods, which include the following: the shortest-distance method, the longest-distance method, the middle-distance method, the center of gravity method, the group-average method, the variable group-average method, the variable method, and the variance and sum method. The recurrence formula for these methods is

$$D_{ir}^2 = \alpha_p D_{ip}^2 + \alpha_q D_{iq}^2 + \beta D_{pq}^2 + r |D_{ip}^2 - D_{iq}^2|, \quad (1)$$

where D_{ij} is the distance between classes. The values of parameters α , β , and γ vary according to the different methods and are adopted in the program. The default values of the parameters are used in the calculations.

2.2. Ant Colony Optimization (the Unknown Number). One of the clustering problems is that the number of clusters is unknown. In the ant colony optimization (ACO) clustering method, the samples are thought of as ants with different attributes, and the cluster center is the food sources that the ants seek. Therefore, the sample clustering is the mechanism of ants looking for food sources.

If X is a data set to be classified and N is the feature set of the samples, the clustering algorithm is as follows.

- (1) Initially, allocate N samples into N classes because one sample belongs to one class.
- (2) Calculate Euclidean distances between classes w_i and w_j :

$$d_{ij} = \sqrt{\sum_{k=1}^n \left(\overline{X}_k^{w_i} - \overline{X}_k^{w_j} \right)^2}, \quad (2)$$

$$\overline{X}^{w_i} = \frac{1}{N_i} \sum_{k=1}^{N_i} X_k, \quad (3)$$

where d_{ij} represents the Euclidean distance between classes i and j , $\overline{X}^{w_i}$ is the cluster center vector, and N_i is the number of samples in class w_i .

- (3) Calculate the amount of pheromone on each path. Let r be the radius of the cluster. $\tau_{ij}(t)$ is the residual

pheromone at time t on the path from w_i to w_j . The pheromone on the path (i, j) is

$$\tau_{ij}(t) = \begin{cases} 1, & d_{ij} \leq r, \\ 0, & d_{ij} > r, \end{cases} \quad (4)$$

where $r = A + d_{\min} + (d_{\max} - d_{\min}) \cdot B$ and A and B are constants.

(4) Calculate the probability of w_i merging into w_j :

$$p_{ij}(t) = \frac{\tau_{ij}^\alpha(t) \eta_{ij}^\beta(t)}{\sum_{i \in S} \tau_{is}^\alpha(t) \eta_{is}^\beta(t)}, \quad (5)$$

where $S = \{s \mid d_{ij} \leq r, s = 1, 2, \dots, j-1, j+1, \dots, M\}$, s stands for a number of a certain class, S represents a collection of all numbers of classes, where the distance is less than or equal to r in the class j , M is the total number of classes, and $\eta_{is}(t)$ is the weight parameters.

(5) If $p_{ij}(t) \geq p_0$, then w_i is subsumed in the w_j , the number of classes minus 1. p_0 is a given probability value. Recalculate the merged cluster center.

(6) Determine whether there is merging or not. If there is no merging, stop the cycle; otherwise, go to step (2) and continue the iteration.

2.3. Extreme Learning Machine. The extreme learning machine (ELM) was proposed by Huang et al. in 2006 [62]. The core principle of ELM is a single-hidden layer feed-forward neural network (SLFN), a typical structure of which is shown in Figure 1. This network consists of an input layer, a hidden layer, and an output layer. The neurons in the former two and latter two are fully connected. There are n neurons corresponding to the input variables in the input layer; there are l neurons in hidden layer; there are m neurons corresponding to m output variables in the output layer.

Briefly, suppose the connection weight ω between the input layer and the hidden layer is given by

$$\omega = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2n} \\ \vdots & \vdots & & \vdots \\ \omega_{l1} & \omega_{l2} & \cdots & \omega_{ln} \end{bmatrix}_{l \times n}, \quad (6)$$

where ω_{ij} represents the weight between neuron i in the input layer and neuron j in the hidden layer.

The connection weight matrix β between the hidden layer and the output layer is given by

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & & \vdots \\ \beta_{l1} & \beta_{l2} & \cdots & \beta_{lm} \end{bmatrix}_{l \times m}, \quad (7)$$

where β_{jk} represents the weight value between the j neuron in the hidden layer and the k neuron in the output layer.

FIGURE 1: A typical structure of a single-hidden layer feed-forward neural network (SLFN).

The threshold b of the neuron in the hidden layer is given by

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_l \end{bmatrix}_{l \times 1}. \quad (8)$$

The input matrix X and the output matrix Y in the training set with Q samples are given by

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1Q} \\ x_{21} & x_{22} & \cdots & x_{2Q} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nQ} \end{bmatrix}_{n \times Q}, \quad (9)$$

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1Q} \\ y_{21} & y_{22} & \cdots & y_{2Q} \\ \vdots & \vdots & & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mQ} \end{bmatrix}_{m \times Q}.$$

The activation function of the neuron in the hidden layer is $g(x)$; the output of net T can be seen in Figure 1. Consider

$$T = [t_1, t_2, \dots, t_Q]_{m \times Q},$$

$$t_j = \begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{mj} \end{bmatrix}_{m \times 1} = \begin{bmatrix} \sum_{i=1}^l \beta_{i1} g(\omega_i x_j + b_i) \\ \sum_{i=1}^l \beta_{i2} g(\omega_i x_j + b_i) \\ \vdots \\ \sum_{i=1}^l \beta_{im} g(\omega_i x_j + b_i) \end{bmatrix}_{m \times 1}, \quad (10)$$

$j = 1, 2, \dots, Q,$

where $\omega_i = [\omega_{i1}, \omega_{i2}, \dots, \omega_{in}]$, $x_j = [x_{1j}, x_{2j}, \dots, x_{nj}]^T$. Equation (10) can be expressed as

$$H\beta = T', \quad (11)$$

where T' is the transpose of the matrix T and H is the output matrix of the hidden layer of the neural network. Consider

$$H(w_1, w_2, \dots, w_l, b_1, b_2, \dots, b_l, x_1, x_2, \dots, x_Q) \\ = \begin{bmatrix} g(w_1x_1 + b_1) & g(w_2x_1 + b_2) & \cdots & g(w_lx_1 + b_l) \\ g(w_1x_2 + b_1) & g(w_2x_2 + b_2) & \cdots & g(w_lx_2 + b_l) \\ \vdots & \vdots & & \vdots \\ g(w_1x_Q + b_1) & g(w_2x_Q + b_2) & \cdots & g(w_lx_Q + b_l) \end{bmatrix}_{Q \times l} \quad (12)$$

Because the number of neurons in the hidden layer is equal to the number of samples in the training set, SLFN can approach the training samples with zero error for any w and b ; that is,

$$\sum_{j=1}^Q \|t_j - y_j\| = 0, \quad (13)$$

where $y_j = [y_{1j}, y_{2j}, \dots, y_{mj}]^T$, $j = 1, 2, \dots, Q$

However, when the number of the training set Q is relatively large, the number of the hidden layer neurons K is usually smaller than Q to reduce the computational complexity. Meanwhile, the training error of SLFN can approach any $\varepsilon > 0$; that is,

$$\sum_{j=1}^Q \|t_j - y_j\| < \varepsilon. \quad (14)$$

Therefore, when the activation function $g(x)$ can be infinitely differentiable, not all of the parameters of SLFN need to be adjusted. The values of w and b can be chosen randomly before training and remain the same in the rest of the selection process. The connection weights β between the hidden layer and output layer can be obtained by solving the equations to obtain the least squares solution:

$$\min_{\beta} \|H\beta - T'\|, \quad (15)$$

the estimated solution is $\hat{\beta} = H^+T'$, where H^+ is the Moore-Penrose generalized inverse of the output matrix of the hidden layer H .

From the above analysis, ELM can randomly generate w and b before training. It is only needed to determine the number of hidden layer neurons and the activation function of the hidden layer neuron (infinitely differentiable), and then β can be calculated. Generally, ELM learning algorithms follow the next steps.

- (1) Determine the number of the neurons in the hidden layer and randomly set the connection weight w between the input layer and the hidden layer and the bias b of the neurons in the hidden layer.
- (2) Select an infinitely differentiable function as the activation function of the neurons in the hidden layer, and then calculate the matrix H of the hidden layer output.
- (3) Calculate the output layer weight $\hat{\beta} : \hat{\beta} = H^+T'$.

3. Technology Roadmap

The technology roadmap of the ELM method based on KS-HC/ACO to improve the accuracy of the molecular bond energies calculated by DFT is shown in Figure 2. It consists of three parts. Part one is the collection of the data, including molecular bond data and the molecular descriptors. The second part describes the theoretical calculation. Ninety-two important organic NO carrier molecules are studied, 12 molecular descriptors are calculated by DFT B3LYP/6-31G(d) (ΔH_{homo} , Q_Y , Q_N , Q_O , N_X , μ , α , E_{HOMO} , E_{LUMO} , $E_{\text{HOMO}-1}$, $E_{\text{LUMO}+1}$, and ΔE), and homolysis BDE experimental values are included [34]. The Kenstone method is used to divide the training set and the test set. HC and ACO are taken as different classification methods; the former requires a predefined number of the classes to select the appropriate molecular descriptors that will be modeled by the ELM method, while the latter does not need to specify the number of classes. The third part aims at testing and analyzing the models. When the calculation results are not satisfied, the algorithm should be optimized repeatedly up to a reasonable accuracy. At the same time, it is suggested to determine the comparison test.

4. Results and Discussion

4.1. Kenstone Calculation Results. The basic idea of the Kenstone method [63] is to divide the training and test sets according to the distance between two data points in the data set. First, the furthest points are put into the training set; the distances between the remaining points and the training data are then calculated. The minimum distances are retained while the maximum distance is selected and put into the training data one by one until the training set population reaches the desired number. While random assignment of the training and test sets can occasionally yield a better calculation result, it cannot guarantee the best result or an outlier-free outcome. For the Kenstone method, these problems are avoided, which means that the KS-HC/ACO-ELM model always yields good results and generalizability. In this study, 80 molecules selected with the Kenstone method are regarded as the training set and 12 molecules represent the test set. Molecules 13, 19, 21, 27, 31, 35, 36, 43, 55, 70, 78, and 81 are in the test set, and the rest are in the training set. It can be seen from the Kenstone calculation results that molecules 13, 19, 21, 27, 31, 35, 36, and 43 contain N-NO; molecule 55 contains O-NO; molecules 70 and 78 contain S-NO; and molecule 81 contains C-NO. This shows that the KS method can achieve diversity and balance for the training and test sets.

4.2. Hierarchical Clustering Calculation Results. It is critical for the HC to adopt the appropriate distance and clustering method. It is best if the method can exclude the dimensional effect, and usually the standard Euclidean distance is a reasonable choice. As for the clustering method, the shortest-distance method is too contracted and the longest-distance method is too expanded. These two methods are simple, but they belong to extreme classification methods. Therefore, the more appropriate methods are the average-distance method,

FIGURE 2: The flow chart of the KS-HC/ACO-ELM model.

TABLE 1: Cophenetic correlation coefficients of various clustering methods based on the Euclidean distance.

Clustering methods	Longest-distance method	Shortest-distance method	Average-distance method	Weighted-average method	Center of gravity method	Middle-distance method	Sum of squares
Cophenetic correlation coefficients based on the Euclidean distance	0.7726	0.6685	0.8381	0.8246	0.8113	0.7739	0.7995

the weighted average method, and the center of gravity method.

There is no existing standard for evaluating the quality of clustering results, and the application of the clustering method depends on the researchers' application skills and experiences with the classification objects. If the clustering method is applied properly, the critical scale and the distance matrix could be highly correlated. The correlation is calculated by cophenetic correlation coefficients, which can be used to evaluate the clustering results. The value of the cophenetic correlation coefficient is between zero and one. If the clustering is effective, the value is large. However, the high value is only statistically significant; it does not mean that the result is physically effective. Sometimes the result is not meaningful in practice.

The cophenetic correlation coefficients are shown in Table 1. The clustering methods based on the Euclidean distance include the longest-distance method, the shortest-distance method, the average-distance method, the weighted-average method, the center of gravity method, the middle-distance method, and the sum of squares.

Based on the average-distance method of the Euclidean distance, 12 molecular descriptors are clustered into classes, and the numbers of classes are given from 2 to 11. This

paper only takes for an example the 7 classes. The clustering pedigree chart of the average-distance method is shown in Figure 3. Numbers 1 through 12 on the horizontal axis in Figure 3 correspond to twelve molecular descriptors (ΔH_{homo} , Q_Y , Q_N , Q_O , N_X , μ , α , E_{HOMO} , E_{LUMO} , $E_{\text{HOMO}-1}$, $E_{\text{LUMO}+1}$, and ΔE). The clustering calculation results are (1, 7, 2, 7, 4, 6, 4, 5, 5, 5, and 3), which indicate that ΔH_{homo} , Q_N , μ , and ΔE are clustered into classes 1, 2, 6, and 3, respectively; Q_Y and Q_O are clustered into class 7; N_X and α into class 4; and E_{HOMO} , E_{LUMO} , $E_{\text{HOMO}-1}$, and $E_{\text{LUMO}+1}$ into class 5. When the given numbers of the clustering classes are different, the analysis and principle are the same as those used in class 7.

4.3. Ant Colony Optimization Calculation Results. The HC method needs a predefined number of clusters, which includes every type of classification. The advantage is obvious when the sample (i.e., the number of molecule descriptors) is small, but when the sample is large, it is neither necessary nor practical to list every classification using an exhaustive method. Therefore, in this paper, an ACO without a predefined number of clusters is also applied to the screening of molecular descriptors to arrive at a generalized model of the ELM method for either a small or large number of descriptors.

FIGURE 3: Clustering pedigree chart of the average-distance method.

According to the definition of the ACO, the optimization function is the minimum ratio of the within-class distance to interclass distance. This means the distance among the classes should be large, but the distance between two samples within a class should be small:

$$\min J(w, c) = \frac{\sum_{i=1}^{N_j} \sum_{p=1}^n w_{ij} \|x_{ip} - c_{jp}\|^2}{\sum_{j=1}^{M-1} \sum_{p=1}^n \|c_{j+1,p} - c_{jp}\|^2}, \quad (16)$$

where

$$c_{jp} = \frac{\sum_{i=1}^{N_j} w_{ij} x_{ip}}{\sum_{i=1}^{N_j} w_{ij}}, \quad j = 1, 2, \dots, M; \quad p = 1, 2, \dots, n, \quad (17)$$

$$w_{ij} = \begin{cases} 1, & \text{sample } i \text{ belongs to class } j, \\ 0, & \text{otherwise,} \end{cases}$$

where x_{ip} is the p factor of sample i and c_{jp} is the p factor of the center of the class j .

The results of the ACO calculation of the clustering molecular descriptors (ΔH_{homo} , Q_Y , Q_N , Q_O , N_X , μ , α , E_{HOMO} , E_{LUMO} , $E_{\text{HOMO}-1}$, $E_{\text{LUMO}+1}$, and ΔE) are (1, 2, 2, 2, 3, 4, 5, 6, 6, 6, 6, and 6), where the numbers in the bracket correspond to the classes of the molecular descriptors. From the calculation results it can be observed that twelve molecular descriptors can be divided into six classes: ΔH_{homo} is class 1, Q_Y , Q_N , and Q_O are class 2, N_X , μ , and α are classes 3, 4, and 5, respectively, and E_{HOMO} , E_{LUMO} , $E_{\text{HOMO}-1}$, $E_{\text{LUMO}+1}$, and ΔE are class 6.

4.4. ELM Calculation Results. The correlation coefficients between the 12 molecular descriptors and homolysis BDE experimental values were calculated, and they are 0.64, 0.46, 0.49, 0.02, 0.12, 0.18, 0.28, 0.51, 0.17, 0.43, 0.05, and 0.27. According to the magnitude of the correlation coefficients (e.g., by HC calculation) when Q_Y and Q_O cluster into a class, Q_Y is correlated higher with homolysis BDE than Q_O and will therefore be selected as the input of ELM. Similarly, when N_X and α are classified into one class, α will be selected as the input of ELM. In the cluster consisting of E_{HOMO} , E_{LUMO} , $E_{\text{HOMO}-1}$, and $E_{\text{LUMO}+1}$, E_{HOMO} is selected as the input of

ELM. ΔH_{homo} , Q_N , μ , and ΔE are clustered into one class. Therefore, ΔH_{homo} , Q_N , μ , ΔE , Q_Y , α , and E_{HOMO} are selected as the final inputs of ELM by using the HC method. In the same way, ΔH_{homo} , Q_N , N_X , μ , α , and E_{HOMO} are selected as the final inputs of ELM after using the ACO clustering method. The errors of the test set in different methods are shown in Figure 4(a).

To assess the KS-HC/ACO-ELM method, the results calculated by the molecular descriptors without screening KS-ELM (using all descriptors), KS-HC-ELM (best results are obtained when the molecular descriptors cluster into 7 classes), and KS-ACO-ELM are compared with the results calculated by B3LYP/6-31G(d). A three layered artificial neural network (KS-HC-ANN) calculation was also performed for comparison. The ANN structure uses the same number of inputs and 6 hidden neurons (shown in Figure 4(b)). The corrected results show that the systematic errors of the calculation value are eliminated, that the homolysis BDE are significantly improved by the KS-ELM, KS-HC-ELM, and KS-ACO-ELM methods, the range of error is narrowed down, the STD calculated by the KS-ELM, KS-HC-ELM, and KS-ACO-ELM methods is much smaller than the calculation error by DFT B3LYP/6-31G(d), the errors are in a Gaussian distribution, and the values of most of the errors are approximately 0. As is clearly seen in Figure 4(b), the results from KS-HC-ELM approach the homolysis BDE experimental values better than those obtained from the ANN, and the calculation time is also much less than that in the ANN because the ANN learning algorithm requires a number of network training parameters set by manual work. ANN also easily leads to a local optimal solution, and its calculated results are very unstable. The ELM method only requires setting the number of hidden layer nodes of the network. It does not need to adjust the network input weights and hidden bias in the implementation process of the algorithm, and it can generate a unique optimal solution, so it has both high learning speed and good generalization performance.

The STD corrected by the KS-ELM method is much smaller than the error of the B3LYP/6-31G(d), and the STD calculated by KS-HC-ELM and KS-ACO-ELM is smaller than that by KS-ELM. That means that it is necessary to screen molecular descriptors to eliminate redundant descriptors. The STD is reduced from 4.03 to 0.30, 0.28, 0.29, and 0.32 kcal mol⁻¹ after correction by the KS-ELM, KS-HC-ELM, KS-ACO-ELM, and KS-HC-ANN methods, respectively. It is quite simple to see that the result corrected by the KS-HC-ELM method is much closer to the experimental results. If some trivial features are introduced into ELM, the accuracy of the model might decrease, and if the parameters are selected inappropriately, overfitting could also appear. In this experiment, although the calculation results using the KS-HC-ELM method are superior to those by the KS-ACO-ELM method, it cannot be assumed that this will be the case under every circumstance because each method has its own advantages. The clustering methods can be divided into two categories, where one requires a predefined number of categories, while the other does not. Because the ELM inputs affect its performance and the best combination of molecular descriptors is difficult to determine, it is recommended

FIGURE 4: Errors of the calculated homolysis BDE and experimental values for the test set. (a) Errors calculated by the B3LYP/6-31G(d), KS-ELM (including all the molecular descriptors), KS-HC-ELM, and KS-ACO-ELM methods for twelve molecules in the test set. (☆), (◇), (○), and (★) represent methods B3LYP/6-31G(d), KS-ELM, KS-HC-ELM, and KS-ACO-ELM, respectively. (b) Experimental homolysis BDE and the calculated homolysis BDE values determined using the KS-HC-ELM and KS-HC-ANN methods (kcal mol⁻¹).

that when the number of clustering molecular descriptors is smaller, both the HC method and the ACO method are suitable and that when the number of the molecular descriptors is larger, the ACO method is more appropriate.

5. Conclusions

Nitric oxide (NO) is a very important signaling molecule in mammalian systems. When there are problems resulting in the release of NO molecules in the body, NO carrier molecules may be taken as a drug to treat the disease. It is complicated to measure NO molecular bond homolysis BDE in various experimental methods, and it is rather difficult to achieve chemical accuracy. The DFT method has been very popular in computational chemistry for the past two decades because many computational studies have shown that DFT calculations can capture the physical essence of the molecules. For the calculation of small molecules, the accuracy can reach a high level, but for large molecules, the cost of calculation is too expensive and the calculation accuracy is quite poor.

In this paper, the combination of KS-HC/ACO-ELM and DFT calculation methods succeeds in improving the DFT calculation accuracy of NO bond homolysis BDE. The results show that for three KS-HC/ACO-ELM methods, the STD of homolysis BDE of 92 organic molecules in the test set decreases from 4.03 to 0.30, 0.28, and 0.29 kcal mol⁻¹. This proves that the KS-HC/ACO-ELM method based on B3LYP/6-31G(d) can remarkably improve homolysis BDE calculations, and its result may be used as the reference value to experimental results with high accuracy. The adoption of Kenstone makes the model more generalized. Although the model is ambiguous and the space of the descriptors

is complex, the reproducibility of the calculation is very important. The selection of the training set and test set using the Kenstone method makes the model more robust and reproducible and makes the calculation results more convincing. The selection of appropriate molecular descriptors using the HC/ACO can eliminate the subjective bias on some input descriptors. Compared to the ANN method, the nonlinear model created by ELM can not only avoid sensitive parameters and local minimum problems, but it can also achieve more accurate calculation results with less calculation time and computing resources.

The KS-HC/ACO-ELM method expands the feasibility and applicability of the B3LYP/6-31G(d) method, and the more experimental data and molecular descriptors there are, the higher calculation accuracy the KS-HC/ACO-ELM will achieve, leading to more obvious advantages. It is important that such a high-precision method can be used to design novel NO-releasing drug molecule. We believe that it is significantly effective for the KS-HC/ACO-ELM method to improve the accuracy of DFT B3LYP/6-31G(d) calculations, even with a smaller basis set (e.g., STO-3G, etc.). Meanwhile, this method can be applied to correct other properties such as Heterolytic Bond Dissociation Energies, Absorbed Energies, Ionization Energies, and heat of formation. Further studies are ongoing. This study provides an effective tool (the combination of the KS-HC/ACO-ELM and DFT methods) to improve and predict highly accurate homolysis BDE of the molecular NO carrier systems.

Acknowledgments

The authors gratefully acknowledge the financial support from the Science and Technology Development Planning of

Jilin Province (20130522109JH, 20110364, and 20125002) and the Fundamental Research Funds for the Central Universities (11QNJJ008 and 11QNJJ027).

References

- [1] X. W. Zhao, W. Y. Zhang, X. Xu, Z. Q. Ma, and M. H. Yin, "Prediction of protein phosphorylation sites by using the composition of k-spaced amino acid pairs," *PLOS ONE*, vol. 7, no. 10, Article ID e46302, 2012.
- [2] T. Zou, M. Yin, and W. Gu, "Reverse bridge theorem under constraint partition," *Mathematical Problems in Engineering*, vol. 2010, Article ID 617398, 18 pages, 2010.
- [3] J. N. Wang, J. L. Jin, Y. Geng et al., "An accurate and efficient method to predict the electronic excitation energies of BODIPY fluorescent dyes," *Journal of Computational Chemistry*, vol. 34, no. 7, pp. 566–575, 2013.
- [4] J.-N. Wang, H.-L. Xu, S.-L. Sun et al., "An effective method for accurate prediction of the first hyperpolarizability of alkalides," *Journal of Computational Chemistry*, vol. 33, no. 2, pp. 231–236, 2012.
- [5] M. Yin, Y. Hu, F. Yang, X. Li, and W. Gu, "A novel hybrid K-harmonic means and gravitational search algorithm approach for clustering," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9319–9324, 2011.
- [6] X. T. Li, J. N. Wang, and M. H. Yin, "Enhancing the performance of cuckoo search algorithm using orthogonal learning method," *Neural Computing and Applications*, 2013.
- [7] X. Zhao, X. Li, Z. Ma, and M. Yin, "Prediction of lysine ubiquitylation with ensemble classifier and feature selection," *International Journal of Molecular Sciences*, vol. 12, no. 12, pp. 8347–8361, 2011.
- [8] K. K. Irikura and D. J. Frurip, *Prediction and Estimation of Molecular Thermodynamics*, American Chemical Society, Washington, DC, USA, 1998.
- [9] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, Wiley, West Sussex, UK, 2002.
- [10] X.-M. Duan, G.-L. Song, Z.-H. Li, X.-J. Wang, G.-H. Chen, and K.-N. Fan, "Accurate prediction of heat of formation by combining Hartree-Fock/density functional theory calculation with linear regression correction approach," *Journal of Chemical Physics*, vol. 121, no. 15, pp. 7086–7095, 2004.
- [11] X.-M. Duan, Z.-H. Li, H.-R. Hu et al., "Linear regression correction to first principle theoretical calculations—improved descriptors and enlarged training set," *Chemical Physics Letters*, vol. 409, no. 4–6, pp. 315–321, 2005.
- [12] D. Jacquemin, V. Wathelet, and E. A. Perpète, "Ab initio investigation of the $n \rightarrow \pi^*$ transitions in thiocarbonyl dyes," *The Journal of Physical Chemistry A*, vol. 110, no. 29, pp. 9145–9152, 2006.
- [13] D. Jacquemin, E. A. Perpète, O. A. Vydrov, G. E. Scuseria, and C. Adamo, "Assessment of long-range corrected functionals performance for $n \rightarrow \pi^*$ transitions in organic dyes," *Journal of Chemical Physics*, vol. 127, no. 9, Article ID 094102, 2007.
- [14] R. Khosrokhavar, J. B. Ghasemi, and F. Shiri, "2D Quantitative Structure-property relationship study of mycotoxins by multiple linear regression and support vector machine," *International Journal of Molecular Sciences*, vol. 11, no. 9, pp. 3052–3068, 2010.
- [15] J. Homer, S. C. Generalis, and J. H. Robson, "Artificial neural networks for the prediction of liquid viscosity, density, heat of vaporization, boiling point and Pitzer's acentric factor. Part I. Hydrocarbons," *Physical Chemistry Chemical Physics*, vol. 1, no. 17, pp. 4075–4081, 1999.
- [16] L. E. S. De Souza and S. Canuto, "Efficient estimation of second virial coefficients of fused hard-sphere molecules by an artificial neural network," *Physical Chemistry Chemical Physics*, vol. 3, no. 21, pp. 4762–4768, 2001.
- [17] L. Hu, X. Wang, L. Wong, and G. Chen, "Combined first-principles calculation and neural-network correction approach for heat of formation," *Journal of Chemical Physics*, vol. 119, no. 22, pp. 11501–11507, 2003.
- [18] X. Wang, L. Hu, L. Wong, and G. Chen, "A combined first-principles calculation and neural networks correction approach for evaluating Gibbs energy of formation," *Molecular Simulation*, vol. 30, no. 1, pp. 9–15, 2004.
- [19] J. R. Votano, M. Parham, L. H. Hall, L. B. Kier, and L. M. Hall, "Prediction of aqueous solubility based on large datasets using several QSPR models utilizing topological structure representation," *Chemistry & Biodiversity*, vol. 1, no. 11, pp. 1829–1841, 2004.
- [20] X.-M. Duan, Z.-H. Li, G.-L. Song, W.-N. Wang, G.-H. Chen, and K.-N. Fan, "Neural network correction for heats of formation with a larger experimental training set and new descriptors," *Chemical Physics Letters*, vol. 410, no. 1–3, pp. 125–130, 2005.
- [21] A. Alexandridis, P. Patrinos, H. Sarimveis, and G. Tsekouras, "A two-stage evolutionary algorithm for variable selection in the development of RBF neural network models," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, no. 2, pp. 149–162, 2005.
- [22] M. Arakawa, K. Hasegawa, and K. Funatsu, "QSAR study of anti-HIV HEPT analogues based on multi-objective genetic programming and counter-propagation neural network," *Chemometrics and Intelligent Laboratory Systems*, vol. 83, no. 2, pp. 91–98, 2006.
- [23] M. Fernández and J. Caballero, "Ensembles of Bayesian-regularized genetic neural networks for modeling of acetylcholinesterase inhibition by huprines," *Chemical Biology and Drug Design*, vol. 68, no. 4, pp. 201–212, 2006.
- [24] R. K. H. Galvão, M. C. U. Araújo, M. D. N. Martins et al., "An application of subbagging for the improvement of prediction accuracy of multivariate calibration models," *Chemometrics and Intelligent Laboratory Systems*, vol. 81, no. 1, pp. 60–67, 2006.
- [25] K. T. Lee, S. Bhatia, A. R. Mohamed, and K. H. Chu, "Optimizing the specific surface area of fly ash-based sorbents for flue gas desulfurization," *Chemosphere*, vol. 62, no. 1, pp. 89–96, 2006.
- [26] P. Zhou, F. Tian, and Z. Li, "A structure-based, quantitative structure-activity relationship approach for predicting HLA-A*0201-restricted cytotoxic T lymphocyte epitopes," *Chemical Biology and Drug Design*, vol. 69, no. 1, pp. 56–67, 2007.
- [27] J. M. Wu and X. Xu, "The XI method for accurate and efficient prediction of heats of formation," *The Journal of Chemical Physics*, vol. 127, no. 21, Article ID 214105, 8 pages, 2007.
- [28] M. Sattari and F. Gharagheizi, "Prediction of molecular diffusivity of pure components into air: a QSPR approach," *Chemosphere*, vol. 72, no. 9, pp. 1298–1302, 2008.
- [29] N. Qu, L. Wang, M. Zhu, Y. Dou, and Y. Ren, "Radial basis function networks combined with genetic algorithm applied to nondestructive determination of compound erythromycin ethylsuccinate powder," *Chemometrics and Intelligent Laboratory Systems*, vol. 90, no. 2, pp. 145–152, 2008.
- [30] W. Xu, X. Liu, S. Luan, Q. Zhang, and P. Zhang, "An ab initio potential energy surface of the $\text{He} + \text{H}_2^+ \rightarrow \text{HeH}^+ + \text{H}$ reaction," *Chemical Physics Letters*, vol. 464, no. 1–3, pp. 92–96, 2008.

- [31] J. Wu and X. Xu, "Improving the B3LYP bond energies by using the X1 method," *Journal of Chemical Physics*, vol. 129, no. 16, Article ID 164103, 11 pages, 2008.
- [32] M. D. Wodrich and C. Corminboeuf, "Reaction enthalpies using the neural-network-based X1 approach: the important choice of input descriptors," *The Journal of Physical Chemistry A*, vol. 113, no. 13, pp. 3285–3290, 2009.
- [33] T. Gao, L.-L. Shi, H.-B. Li et al., "Improving the accuracy of low level quantum chemical calculation for absorption energies: the genetic algorithm and neural network approach," *Physical Chemistry Chemical Physics*, vol. 11, no. 25, pp. 5124–5129, 2009.
- [34] H. Z. Li, W. Tao, T. Gao, H. Li, Y. H. Lu, and Z. M. Su, "Improving the accuracy of density functional theory (DFT) calculation for homolysis bond dissociation energies of Y-NO bond: generalized regression neural network based on grey relational analysis and principal component analysis," *International Journal of Molecular Sciences*, vol. 12, no. 4, pp. 2242–2261, 2011.
- [35] H. Z. Li, L. H. Hu, W. Tao et al., "a promising tool to achieve chemical accuracy for density functional theory calculations on Y-NO homolysis bond dissociation energies," *International Journal of Molecular Sciences*, vol. 13, no. 7, pp. 8051–8070, 2012.
- [36] H.-Z. Li, W. Tao, T. Gao, H. Li, Y.-H. Lü, and Z.-M. Su, "Improving the accuracy of DFT calculation for homolysis bond dissociation energies of Y-NO bond via back propagation neural network based on mean impact value," *Chemical Journal of Chinese Universities*, vol. 33, no. 2, pp. 346–352, 2012.
- [37] Y. E. Shao, "Prediction of currency volume issued in Taiwan using a hybrid artificial neural network and multiple regression approach," *Mathematical Problems in Engineering*, vol. 2013, Article ID 676742, 9 pages, 2013.
- [38] W. Ma, X. Zhang, F. Luan et al., "Support vector machine and the heuristic method to predict the solubility of hydrocarbons in electrolyte," *The Journal of Physical Chemistry A*, vol. 109, no. 15, pp. 3485–3492, 2005.
- [39] X. Zhao, Z. Ma, and M. Yin, "Using support vector machine and evolutionary profiles to predict antifreeze protein sequences," *International Journal of Molecular Sciences*, vol. 13, no. 2, pp. 2196–2207, 2012.
- [40] A. R. Lima, A. J. Cannon, and W. W. Hsieh, "Nonlinear regression in environmental sciences by support vector machines combined with evolutionary strategy," *Computers & Geosciences*, vol. 50, no. 4, pp. 136–144, 2013.
- [41] A. R. Butler and D. L. H. Williams, "The physiological role of nitric oxide," *Chemical Society Reviews*, vol. 22, no. 4, pp. 233–241, 1993.
- [42] B. A. Averill, "Dissimilatory nitrite and nitric oxide reductases," *Chemical Reviews*, vol. 96, no. 7, pp. 2951–2964, 1996.
- [43] R. M. J. Palmer, A. G. Ferrige, and S. Moncada, "Nitric oxide release accounts for the biological activity of endothelium-derived relaxing factor," *Nature*, vol. 327, no. 6122, pp. 524–526, 1987.
- [44] L. J. Ignarro, "Biosynthesis and metabolism of endothelium-derived nitric oxide," *Annual Review of Pharmacology and Toxicology*, vol. 30, pp. 535–560, 1990.
- [45] P. L. Feldman, O. W. Griffith, and D. J. Stuehr, "The surprising life of nitric oxide," *Chemical and Engineering News*, vol. 71, no. 51, pp. 26–38, 1993.
- [46] J. M. Fukuto and L. J. Ignarro, "In vivo aspects of nitric oxide (NO) chemistry: does peroxynitrite (OONO) play a major role in cytotoxicity," *Accounts of Chemical Research*, vol. 30, no. 4, pp. 149–152, 1997.
- [47] S. Moncada, R. M. J. Palmer, and E. A. Higgs, "Nitric oxide: physiology, pathophysiology, and pharmacology," *Pharmacological Reviews*, vol. 43, no. 2, pp. 109–142, 1991.
- [48] L. J. Ignarro, "Signal transduction mechanisms involving nitric oxide," *Biochemical Pharmacology*, vol. 41, no. 4, pp. 485–490, 1991.
- [49] C. T. Gnewuch and G. Sosnovsky, "A critical appraisal of the evolution of N-nitrosoureas as anticancer drugs," *Chemical Reviews*, vol. 97, no. 3, pp. 829–1013, 1997.
- [50] C. A. Whited, J. J. Warren, K. D. Lavoie et al., "Gating NO release from nitric oxide synthase," *Journal of the American Chemical Society*, vol. 134, no. 1, pp. 27–30, 2012.
- [51] J.-P. Cheng, K. Wang, Z. Yin, X. Zhu, and Y. Lu, "NO affinity. The driving force of nitric oxide (NO) transfer in biomimetic N-nitrosoacetanilide and N-nitrososulfoanilide systems," *Tetrahedron Letters*, vol. 39, no. 43, pp. 7925–7928, 1998.
- [52] J.-P. Cheng, M. Xian, K. Wang, X. Zhu, Z. Yin, and P. G. W. Peng George Wang, "Heterolytic and homolytic Y-NO bond energy scales of nitroso-containing compounds: chemical origin of NO release and NO capture," *Journal of the American Chemical Society*, vol. 120, no. 39, pp. 10266–10267, 1998.
- [53] M. Xian, X.-Q. Zhu, J. Lu, Z. Wen, and J.-P. Cheng, "The first O-NO bond energy scale in solution: heterolytic and homolytic cleavage enthalpies of O-nitrosyl carboxylate compounds," *Organic Letters*, vol. 2, no. 3, pp. 265–268, 2000.
- [54] X.-Q. Zhu, J.-Q. He, Q. Li, M. Xian, J. Lu, and J.-P. Cheng, "N-NO bond dissociation energies of N-nitroso diphenylamine derivatives (or analogues) and their radical anions: implications for the effect of reductive electron transfer on N-NO bond activation and for the mechanisms of NO transfer to nitranions," *Journal of Organic Chemistry*, vol. 65, no. 20, pp. 6729–6735, 2000.
- [55] J.-M. Lü, J. M. Wittbrodt, K. Wang et al., "NO affinities of S-nitrosothiols: a direct experimental and computational investigation of RS-NO bond dissociation energies," *Journal of the American Chemical Society*, vol. 123, no. 12, pp. 2903–2904, 2001.
- [56] X.-Q. Zhu, W.-F. Hao, H. Tang, C.-H. Wang, and J.-P. Cheng, "Determination of N-NO bond dissociation energies of N-Methyl-N-nitrosobenzenesulfonamides in acetonitrile and application in the mechanism analyses on NO transfer," *Journal of the American Chemical Society*, vol. 127, no. 8, pp. 2696–2708, 2005.
- [57] X.-Q. Zhu, J.-Y. Zhang, and J.-P. Cheng, "Mechanism and driving force of NO transfer from S-nitrosothiol to cobalt(II) porphyrin: a detailed thermodynamic and kinetic study," *Inorganic Chemistry*, vol. 46, no. 2, pp. 592–600, 2007.
- [58] X. Li, X.-Q. Zhu, X.-X. Wang, and J.-P. Cheng, "Determination of N-NO bond dissociation energies of N-nitrosoindoles and their radical anions in acetonitrile," *Chemical Journal of Chinese Universities*, vol. 28, no. 12, pp. 2295–2298, 2007.
- [59] X. Li, X.-Q. Zhu, and J.-P. Cheng, "Determination of NO chemical affinities of benzyl nitrite in acetonitrile," *Chemical Journal of Chinese Universities*, vol. 28, no. 12, pp. 2327–2329, 2007.
- [60] X. Li and J.-P. Cheng, "Determination of S-NO bond dissociation energies of S-Nitroso-N-acetyl-D, L-penicillamine dipeptides," *Chemical Journal of Chinese Universities*, vol. 29, no. 8, pp. 1569–1572, 2008.
- [61] X. Li, H. Deng, X.-Q. Zhu, X. Wang, H. Liang, and J.-P. Cheng, "Establishment of the C-NO bond dissociation energy scale in solution and its application in analyzing the trend of NO transfer from C-nitroso compound to thiols," *Journal of Organic Chemistry*, vol. 74, no. 12, pp. 4472–4478, 2009.

- [62] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, 2006.
- [63] R. W. Kennard and L. A. Stone, "Computer aided design of experiments," *Technometrics*, vol. 11, no. 1, pp. 137-148, 1969.