Scholars Journal of Physics, Mathematics and Statistics

OPEN ACCESS

Abbreviated Key Title: Sch J Phys Math Stat
ISSN 2393-8056 (Print) | ISSN 2393-8064 (Online)
Journal homepage: https://saspublishers.com

Energy Prediction Model of PSO-BP Neural Network Three-dimensional Clusters based on Atomic Coordinates

Zhan Zhang[1*], Rong Huang[2], Zhenglong Li[3]

[1]College of Chemical Engineering,Qinghai University, Xining 810016, Qinghai , P.R.China
[2]College of Automation, Nanjing Institute of Technology, Nanjing 210009, Jiangsu, P.R.China
[3]College of Education, Shaanxi Normal University, Xian 710062, Shanxi, P.R.China

DOI: 10.36347/sjpms.2021.v08i06.002 | **Received:** 19.04.2021 | **Accepted:** 28.05.2021 | **Published:** 04.06.2021

***Corresponding author:** Zhan Zhang

Abstract **Original Research Article**

Energy prediction for different cluster structures is the basis for finding and predicting the global optimal structure of clusters. The current methods for predicting the energy of the ground state structures of different clusters include theoretical prediction methods and optimized simplified potential energy function methods. The accuracy of the theoretical prediction method is high, but its calculation amount is too large. Therefore, this paper proposes a PSO-BP neural network three-dimensional cluster energy prediction model based on atomic coordinates, and uses different types of Euclidean distances between atoms as input variables, and the energy of clusters with different structures as output variables. Select gold cluster Au20 and boron cluster B45-part of the sample data as the training set to build the model, and predict the rest of the samples, and finally get: the prediction accuracy of the PSO-BP neural network model is higher than that of the traditional BP neural network model. The cluster energy prediction model is feasible.

Keywords: cluster structure energy prediction; atomic coordinates; Euclidean distance; PSO-BP neural network.

1. INTRODUCTION

Atomic clusters are submicroscopic structures between atoms and molecules formed by the interaction between atoms [1]. The study of clusters emerged in the 1970s. In 1984, Knight et al. observed the magic number structure of alkali metals with electron shell characteristics. In 1985, HW Kroto and RESmalley et al. found that C60 won the Nobel Prize in Chemistry. The study of clusters has attracted great attention from the physics community and the materials community. The basic problem of atomic cluster research is how clusters are developed from atoms step by step, that is, to predict the ground state structure of clusters. The methods of predicting the ground state structure of clusters include theoretical prediction method and optimized simplified potential energy function method. The theoretical prediction method has high accuracy, but due to its large amount of calculation, when the number of atoms is relatively large, the simplified potential energy function method is often optimized to predict its ground state structure [2]. Even so, the prediction of the cluster ground state structure is still NP difficult. Generally, the structure of atomic clusters can be determined by the three-dimensional coordinates of each atom. Predicting the energy level of the atomic clusters under the known structure is the main content of this paper. It is also the construction of a PSO-BP neural network based on atomic coordinates in this paper. The basis of the energy prediction model for three-dimensional clusters.

2. DATA PROCESSING

In engineering practice, the data we obtain usually lack values, repeated values, etc., so the data needs to be preprocessed before use.

2.1 Euclidean Distance

Euclidean distance is used in many algorithms. It is a common distance measurement algorithm that refers to the natural length between two points in an m-dimensional space. In particular, Euclidean distance in two-dimensional space or three-dimensional space refers to the actual distance between two points. As shown in Figure 1. Euclid's calculation formula is as follows

$$d(X,Y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2}=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$$

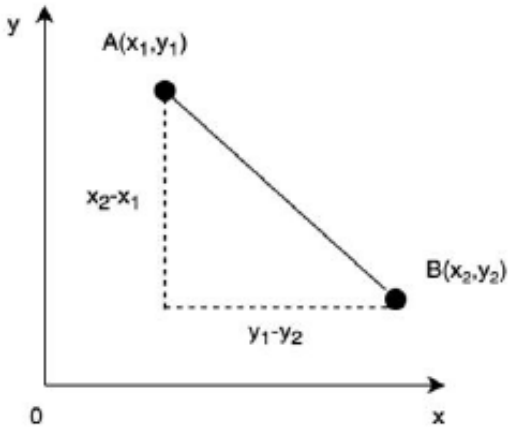

Fig-1: Diagram of Euclidean distance

Euclidean distance is often used to measure the distance between two variables to measure their degree of correlation. The smaller the Euclidean distance, the closer the two variables are; the larger the Euclidean distance, the farther the two variables are.

2.2 Construct input layer indicators

In this question, we need to use the PSO-BP neural network model to predict the cluster energy. Generally, the higher the dimensionality of the training set, the greater the risk of overfitting. Therefore, we need to extract feature indicators from the known atomic cluster structure data. The final selected input variables are as follows:

(1) The number of atoms.

(2) The longest Euclidean distance between atom i and atom j.

(3) The shortest Euclidean distance between atom i and atom j.

(4) The average Euclidean distance between atoms.

Finally, the data is standardized according to the requirements of the model, and we adopt the z-score standardization method.

3. ESTABLISHMENT OF PSO-BP NEURAL NETWORK MODEL

3.1 BP neural network model

BP neural network is a biological intelligent scientific behavior imitated from the perspective of bionics. It is a multi-layer forward neural network based on error back propagation. It has good adaptation, self-organization and strong learning, association, fault tolerance and resistance. Interference ability [3]. BP neural network can simulate any linear and nonlinear function, so it has good predictive ability. It includes input layer, hidden layer, output layer, each layer has some neurons, and the adjacent layer. All units are connected to each other, and the connection strength between the two units is called "weight". The main steps are as follows:

① Determine the structure of BP neural network. Set the input layer, hidden layer, output layer, and the number of nodes of each layer of BP neural network according to the specific research problem.

② Initialize the relevant parameters of the BP neural network, including the weights and thresholds between the connected layers, and the learning rate.

③ Calculate the network output error according to the hidden layer output H and the prediction error O.

④ Determine whether the error meets the requirements, if not. According to the error, adjust the error of each layer, update the network connection weight and threshold.

⑤ Judge whether the iteration is over, if not, return to.

3.2 Particle Swarm Optimization (PSO) algorithm

Particle Swarm Optimization (PSO) Algorithm Particle Swarm Optimization (PSO) is an evolutionary computing technology developed by Kennedy Eberhart [4-5] and others in 1995, which is derived from the simulation of a simplified social model. The global optimal solution is searched through collaboration and information sharing between individuals in the group. The information sharing principle of following two optimal positions provides a new idea for optimization algorithm design. One of these two positions is the optimal solution a currently found by the particle g_{best}, and the other is the optimal solution currently found by the individual in the neighborhood P_{best} [6].

In the particle swarm algorithm, particles are particles with no mass and volume in the D-dimensional search space, but with position and speed information, and fly at a variable speed in the search space. The flight speed is based on the individual's flight experience and the flight of the group. Dynamic adjustment of experience [7]. Assuming that there are n particles in the search space, the position of the i-th individual particle can be expressed as

$X_i = (X_{i1}, X_{i2}, \cdots, X_{in})$;

Flight speed $V_i = (V_{i1}, V_{i2}, \cdots, V_{in})$;

The position of the best fitness value experienced by particle i is $P_i = (P_{i1}, P_{i2}, \cdots, P_{in})$;

The optimal position experienced by the particle group is $P_g = (P_{g1}, P_{g2}, \cdots, P_{gn})$

The position and velocity of the particles are updated by the following formula [8]:

$$V_{id}^{k+1} = W \times v_{id}^{k+1} + c1 * rand1(p_{id} - x_{id}^{x}) + c2 * rand2(p_{id} - x_{id}^{x}) \quad (1)$$

$$X_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1} \quad (2)$$

Among them: W is the inertia coefficient, which represents the influence of the historical speed of the particle on the current speed; $c1$ and $c2$ represent the acceleration factor, $c1$ adjusts the particle to fly to its optimal position, $c2$ adjusts the particle to fly to the global optimal position, usually $c1 = c2 = 2$; rand1 and rand2 are uniformly distributed random numbers, ranging from 0 to 1, including 0 and 1, and x_{id}^{k} is the d-dimensional component of the position vector of the k-th iteration particle i[9]; velocity formula (The second item of 1) indicates the cognitive ability of the microparticle individual; the third item indicates the social cognitive ability. The particle swarm algorithm based on formulas (1) and (2) is called the basic particle swarm algorithm [10].

The algorithm flow chart of standard PSO is shown in Figure 2.

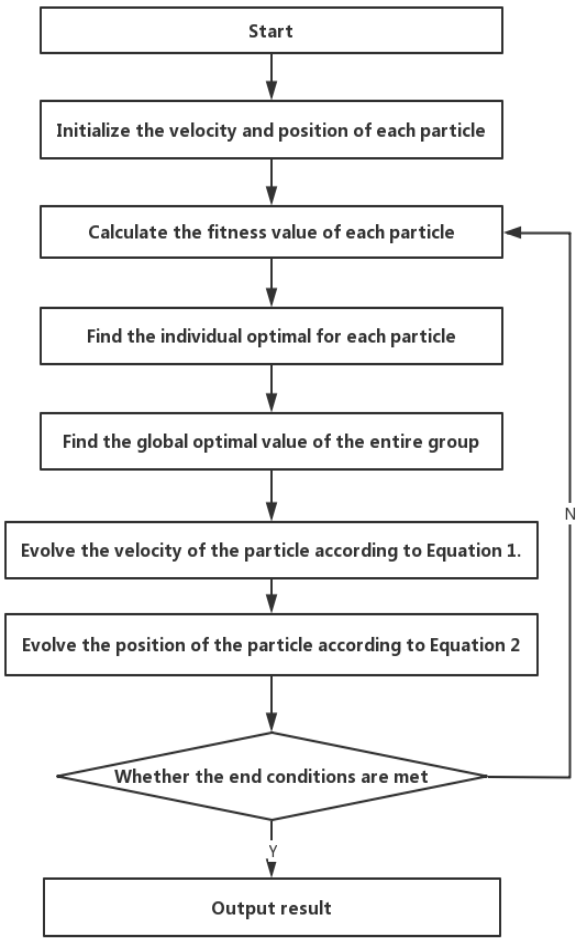

Fig-2: PSO algorithm flow chart

3.3 Research based on PSO-BP algorithm

The particle swarm optimization algorithm is efficient, simple in structure, fast convergence speed, easy to implement, and stable in performance. The particle swarm algorithm to optimize the weights and thresholds of the BP neural network is more stable than the structure obtained by the BP neural network algorithm randomly generating weights and thresholds Sex and superiority.

The steps of the PSO-BP algorithm are as follows:

Step 1: Determine the number of nodes in the input layer, hidden layer and output layer of the BP neural network through training samples.

Step 2: According to the BP neural network structure determined in Step 1, establish the mapping relationship between the BP neural network weights and thresholds and the particle dimensional space of the particle swarm, that is, the total number of neural network connection weights and thresholds is equal to the dimension of the particle swarm, Initialize the relevant parameters of the particle swarm: inertia weight W, acceleration constants C1 and C2, initial position X and initial speed V, set the maximum and minimum speed, population size and evolutionary algebra.

Step 3: Setting $error = \sum_{i=1}^{i=n} \left| Oi_{\text{prediction}} - Oi_{\text{actual}} \right|$ means that the training error and the function are used as the fitness evaluation function of the BP neural network. When the value of error is the smallest, the fitness of the particle swarm reaches the highest, and the problem gets the optimal solution. $Oi_{\text{prediction}}$ is the predicted output of the neural network, and Oi_{actual} is the actual output of the neural network.

Step 4: According to the training error of the BP neural network and Step3, the fitness value of the particle swarm is calculated.

Step 5: Compare the fitness value of each particle with its historical optimal position. If it is better, update the fitness value of the particle as the optimal position of the particle, namely P_{best}。

Step6: Compare the fitness value of each particle with the historical best position g_{best} of the group. If it is better, update g_{best}.

Step 7: Adjust the speed of the particles according to formula 1 and formula 2 in 2 and update the position of the particles.

Step 8: Adjust the relevant parameters of the BP neural network according to the optimal position of the population, mainly the connection weight and threshold.

Step 9: Perform fitness evaluation on the updated particle swarm to determine whether the number of iterations reaches the end condition. If the end condition is met, the algorithm stops. If it is not met, then it should go to Step5 to continue iterating until the end condition is met.

4. MODEL SOLVING

In this paper, based on the collected data, a PSO-BP neural network three-dimensional cluster energy prediction model based on atomic coordinates is established to predict the energy of gold clusters Au20 and boron clusters B45-.

First, the data is preprocessed, the coordinate data of the gold cluster Au20 is converted, and the Euclidean distance between each atom under the different structure of the gold cluster Au20 is calculated, and then the atom i and the atom j are obtained respectively. The longest Euclidean distance; the shortest Euclidean distance between atom i and atom j; the average Euclidean distance between atoms. Then select the number of atoms; the longest Euclidean distance between atom i and atom j; the shortest Euclidean distance between atom i and atom j; the average Euclidean distance between each atom is used as the input of the model Variable, the energy of gold cluster Au20 is used as the output variable. Select 990 sample data as the training set to build a model, and make predictions on the remaining 10 samples.

Matlab is a high-performance visual scientific and engineering calculation software. It provides a complete neural network toolbox. It can greatly save the time of programming neural network learning algorithms and also improve the accuracy of network output. Make the established network more reliable and effective. So the experiment was carried out in the environment of Matlab.

The results obtained by the prediction of the PSO-BP neural network model are compared with the results obtained by the original BP neural network. The comparison results are shown in Table 1. The comparison between the predicted value and the actual value of the PSO-BP neural network model is shown in Figure 3.

Table-1 Comparison of prediction results between PSO-BP neural network and traditional BP neural network

Sample number	PSO-BP neural network prediction value	BP neural network prediction value	Actual value
990	-1548.59	-1547.63	-1549.148376
991	-1550.36	-1551.28	-1549.998946
992	-1550.15	-1547.82	-1549.54879
...	...	...	...
999	-1553.27	-1554.39	-1553.358472

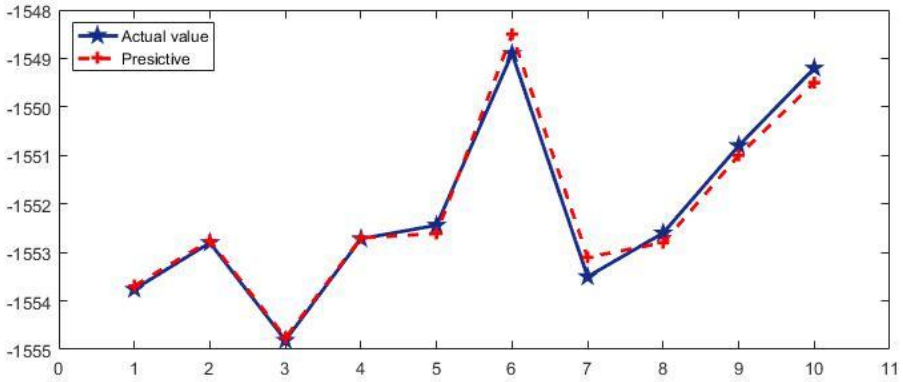

Fig-3: Comparison of predicted and actual values of Au20 by PSO-BP neural network

Secondly, a boron cluster B45-energy prediction model based on PSO-BP neural network is established. The calculation steps are the same as the Au20-energy prediction model for gold clusters based on PSO-BP neural network. First, data preprocessing is performed, and the boron cluster B45 is calculated. - The Euclidean distance between each atom under different structures, and then the longest Euclidean distance between atom i and atom j; the shortest Euclidean distance between atom i and atom j; The average Euclidean distance between atoms. Then select the number of atoms; the longest Euclidean distance between atom i and atom j; the shortest Euclidean distance between atom i and atom j; the average Euclidean distance between each atom is used as the input of the model Variable, the energy of the boron cluster B45- is used as the output variable. Select 3741 sample data as the training set to build the model, and make predictions on the remaining 10 samples. Figure 4 shows the comparison between the predicted value and the actual value.

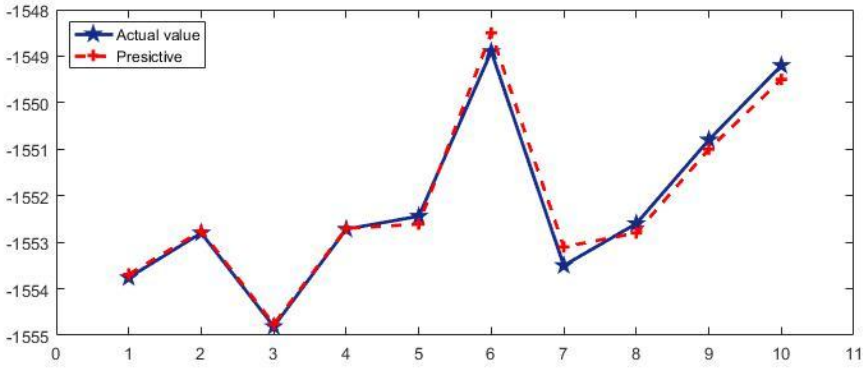

Fig-4: Comparison of PSO-BP neural network predicted value and actual value of boron cluster B45-

From Table 1 and Figure 3 and Figure 4, it can be seen that the prediction accuracy of the PSO-BP neural network model is relatively high, while the prediction accuracy of the traditional BP neural network model is moderate. On the whole, compared with the traditional BP neural network, the PSO-BP neural network model is more accurate, has smaller errors, and has stronger network generalization ability. It can effectively optimize the BP neural network prediction model. PSO-BP neural network. The network is more suitable as the energy prediction model of Au20 and B45- for gold clusters.

5. SUMMARY

BP neural network is widely used and has its own unique advantages, such as approximating nonlinearity, etc. In addition to advantages, BP network also has shortcomings that affect wider applications, such as: slow convergence; falling into a local minimum; It is difficult to determine the number of hidden layer neurons and other issues. But based on the PSO-BP neural network model can effectively part of the BP neural network's shortcomings. The particle swarm algorithm has no crossover and mutation calculations, and the search speed is fast; it has memory and transmits the best position in history to other particles; the structure is simple, the adjustment parameters are few, and it is easy to implement; the real number coding is directly determined by the solution of the problem; Strong search ability, fast convergence and high precision. The particle swarm algorithm to optimize the weights and thresholds of the BP neural network is more stable and superior than the structure obtained by the BP neural network algorithm randomly generating weights and thresholds. When using the PSO-BP neural network model to predict the energy of gold cluster Au20 and boron cluster B45-, compared with traditional BP neural network, it is more accurate, has smaller errors, and has stronger network generalization ability. It can be used for BP neural network. The prediction model is effectively optimized, and its own fitting effect is very good, and the prediction accuracy is high. Therefore, the PSO-BP neural network can be used as an energy prediction model for Au20 clusters and B45-B clusters.

REFERENCES

1. Kangxi. (2020). Research based on Ag (29) clusters and their derived clusters [D]. Anhui University.
2. Li Yanle. 2018Theoretical research on the structure and catalytic activity of gold and palladium clusters [D]. Nanjing University.
3. Zhang Hui. (2012). Housing price prediction based on particle swarm optimization BP neural network [J]. Value Engineering, (14).
4. Kennedy J, Eberhart R. (1995). Part icle sw arm op t im izat ion [A]. P roc IE E E Int Conf on N eu ral N etw orks [C].Perth,. 194221948.
5. Eberhart R, Kennedy J. (1995). A new op t im izer using part iclesw arm theory [A]. P roc 6th Int S ymposium on MicroMachine and H um an Science[C]. Nagoya, 39243.
6. Xie Xiaofeng, Zhang Wenjun, Yang Zhilian. 2003Overview of particle swarm optimization [J]. Control and Decision, (2): 129-131.
7. Liu Yu, Qin Zheng, Shi Zhewen. (2006). Reduced particle swarm optimization algorithm [J]. Journal of Xi'an Jiaotong University, (8):883-887.
8. Yuan Dailin, Cheng Shijuan, Chen Qiu. (2008). A new form of particle swarm algorithm [J]. Computer Engineering and Applications, 44(33): 57-59.
9. Wang Anna, Tao Ziyu, Jiang Maofa, et al. (2006). Intelligent prediction of molten steel temperature in LF furnace based on PSO and BP network [J]. Control and Decision, 21(7): 814-817.
10. Wang Hui. (2010).A new type of multi-swarm particle swarm algorithm [J]. Computer Engineering and Applications, (35):45-48.