Prediction of PM2.5 value by BP artificial neural network in honeybee colony

jiaoyan Ai, Wenyang Gong
Guangxi university GXUSOEE, Nanning, Guangxi China

*373792051@qq.com

Abstract. In this paper, propose a BP artificial neural network optimized by artificial bee colony algorithm with a PM2.5 value for the current Nanning area. In this paper, aiming at the shortcomings of BP neural network prediction, the global optimization performance of artificial bee colony algorithm is used to provide the optimal global initial value for BP neural network and avoid the neural network falling into local optimization. To make the prediction model more accurate. PM2.5 is a small part of the atmospheric chemical composition, which has a great impact on the air quality and thus on the human body. With people's attention to their own health nowadays, it is of great practical significance to predict the PM2.5 content in a certain period in the future. The results show that the average square error and average relative error of the prediction of PM2.5 content based on BP artificial neural network are only 1.25% and 2.46% respectively. BP artificial neural network based on bee colony can be well applied to the prediction of chlorophyll a content, providing a reference for the prediction of PM2.5 content in a certain period in the future.

1. Introduction

People are now paying more and more attention to their own health problems. The environment and air quality are closely related to people's daily life, affecting human health and mental health. The content of PM2.5 in the air is a factor that can judge whether the air quality is good or bad. The prediction of PM2.5 in the air is of great guiding significance to people's daily work and life.

Traditional bp artificial neural network has good application value for numerical prediction, but traditional bp artificial neural network often leads to premature convergence, thus the prediction is not so accurate. Therefore, this paper proposes a valuable method to predict the content of PM2.5.

To sum up, the main contributions of this paper are as follows: 1) bp artificial neural network; 2) bp artificial neural network optimization; 3) bp artificial neural network is optimized by artificial bee colony optimization algorithm.

2. Related work

2.1. At the beginning

The artificial bee colony algorithm is used to improve the BP neural network, and the optimal solution (initial weight and threshold value) in the neural network is transformed into the optimal honey source of the bee colony. [1] Artificial bee colony algorithm provides the global optimal initial weights and
thresholds, and then trains the data through the optimized BP neural network. The global optimality of artificial bee colony algorithm is perfectly integrated with the generalization ability of BP neural network.

2.2. The establishment of BP neural network

Initialize parameters related to artificial bee colony algorithm. The following formula (1) has a population size \( N_c \), Honey bees that go to work \( N_e \), And another bee serving as a monitor to collect honey \( N_o \), And related sites with honey are \( N_s \), And can also be said to be possible solutions, the maximum number of bee searches is limit.[2] The relationships between \( N_c \), \( N_e \), \( N_o \), \( N_s \) and \( D \) are:

\[
N_c = N_e + N_o = 2N_s \tag{1}
\]

\[
D = N_{input} \times N_{hidden} + N_{hidden} \times N_{output} + N_{output} \tag{2}
\]

\( N_{input} \), \( N_{output} \) and \( N_{hidden} \) is where (2) represents the number of relevant nodes in the input layer, hidden layer and output layer of the neural network.[2]

2.3. The Next

Calculating the fitness value of each solution \( f(X_i) \). In the formula: \( f(X_i) \) is the fitness value of the \( I \)th solution. The fitness function is given according to the mean square error of neural network. First, the initial fitness value of the first generation is obtained.

2.4. The Next

Collecting Bees for Domain Search to Produce New Solutions \( V_{ij} \).

\[
V_{ij} = X_{ij} + \text{rand}(-1,1)(X_{ij} - X_{kj}) \tag{3}
\]

In the formula: \( j = \{1,2,\ldots,D\}, D \) is the dimension of each solution, \( k = \{1,2,\ldots,N\} \), Is arbitrary, And \( k = i \). Taking bees for domain search to generate the latter solution \( V_{ij} \). The former solution \( V_{ij} \) is compared with the latter solution \( X_{ij} \). If the adaptability of the latter solution is not less than that of the previous solution, the latter solution is used; if not, the update search is continued.

2.5. The Next

Then to calculate \( P_i \), That is to say, the range of bees searching around the given initial value. Follow bees to take values from these numerical ranges by using the wheel disc in the Macao casino to obtain a better number in the later period, then stop, otherwise continue to search within this range. the following formula (4) is the numerical range that can be obtained in the front.

\[
P_i = f(X_i) / \sum_{n=1}^{N} f(X_n) \tag{4}
\]
2.6. The Next
If the number of times I take the value in Xi in the previous formula (4) is greater than the limit of the value determined in the initial value, then the x value is subtracted, and then the following formula (5) is used to obtain the new following value and the previous number is subtracted.

\[ X_i = X_{\text{min}} + \text{rand}(0,1)(X_{\text{max}} - X_{\text{min}}) \]  

Max and min refer to the maximum and minimum values of x value.

2.7. The End
In the previous initial value setting, there is an MCN, which means that the honeybee that goes to collect honey and the honeybee that monitors the collection of bees get a solution of one cycle. If the number of cycles in the round-trip cycle is not greater than the previous initial value MCN, then they return to 2.4 to continue to obtain the following value. If it is not, it ends, and finally the number with the greatest fitness is taken as the current optimal value we can obtain. The current dominant values obtained by pushing the front waves and the back waves of the Yangtze River serve as the initial numbers of the weights and thresholds of the nodes of the neural network we are testing to obtain a better neural network for the content prediction of our PM2.5.

Can jump out of the current range of local optimal solution. When the numerical value cycle is not stopped but a better value has to be obtained, the method of searching by scouting bees avoids the search result falling into a local extreme value. [6]

3. Data set
The clustering results of the improved artificial bee colony algorithm are applied to BP neural network modeling.

3.1. Experimental environment and experimental data set
The neural network has an input layer, an implicit layer and an output layer. The number of hidden layers is set to two. The hidden layer neurons are estimated by empirical formula. [5] The formula is as follows:

\[ q = \sqrt{m + n + a} \]  

Where q is the number of neurons in the hidden layer, m is the number of neurons in the input layer, n is the number of neurons in the output layer, and a is a constant of 1-10. The output is pm2.5.

Set artificial bee colony algorithm parameters. Colony size \( N_c \), Set to \( N_c = 200 \) The number of bees collected \( N_s \), Number of following bees \( N_o \) And the number of initialsolutions \( N_s \):

\[ N_s = N_e = N_o = 100 \]  

Set the limit number of failures limit=25, Cycles MCN=50.30 samples were used for prediction, and the prediction results were compared with the real values.

In the experiment, 70 hours of training data are used to train the traditional BP neural network and the BP neural network model based on the improved bee colony algorithm proposed in this paper respectively, and the trained neural network is used to predict the data after 30 days.

3.2. Definition of error index
Average absolute error:

\[ MSE = \frac{1}{N} \sum_{i=1}^{N} (\tilde{Z}_i - Z_i)^2 \]
The smaller the MSE value in the formula, the greater the adaptability of bees, and the greater the possibility that the value can be found. Artificial Bee Colony (ABC) uses one team to gather honey and another team to follow. Bees with different identities cooperate and exchange information with each other, thus obtaining very good local searching ability. Information sharing among groups is helpful to get the best in iterative evolution.[7] The optimal solution obtained by artificial bee colony algorithm is used as the initial weight and threshold value of the BP neural network, and then trained and updated by the BP neural network, which can effectively ensure that the training results of the BP neural network jump out of the local optimum, and effectively improve the performance of prediction accuracy, convergence speed and stability.[8]

4. Experimental results and analysis
This section will show the experimental results and analyze the experimental results from the two angles of accuracy and error analysis result sample.

4.1. Accuracy analysis

![Fig. 1. comparison chart of predicted value and real value](image)

The neural network optimization algorithm proposed in this paper is closer to the real value with less error, which is in line with the experimental expectation.

As can be seen from the above experimental results. The blue box is more concentrated than the orange box, and the round orange box is more divergent. In the comparison experiment between the predicted value and the real value, the bp artificial neural network optimized by artificial bee colony algorithm has better prediction result than the traditional bp artificial neural network, which is closer to the real value, thus being more conducive to the prediction of PM2.5 in the atmosphere and guiding people's future travel.

Judging from the weighted average accuracy of categories, we can analyze from the error analysis and comparison experiment blueprint, which category has better error accuracy. Objectively analyze the bp artificial neural network optimized by artificial bee colony algorithm and the traditional bp artificial neural network from various aspects. From the error analysis shown in fig. 2, the error variation trends of the two methods are basically the same, but after 16 generations, they have certain differences. The bp artificial neural network optimized by the artificial bee colony algorithm can continue to evolve after
16 generations, but the traditional bp artificial neural network has stopped continuing to evolve after 8 generations. The bp artificial neural network optimized by the artificial bee colony algorithm can jump out of the limitation of the current range and continue to evolve, which makes the error accuracy better and the prediction result more accurate.

The experimental results show that:
Compared with the general bp neural network, the bp neural network model based on the improved artificial bee colony algorithm greatly reduces the prediction error of pm2.5 and relatively improves the prediction accuracy;
The weights and thresholds of hidden layers in the improved bp neural network are determined by using the improved artificial bee colony algorithm, which relatively reduces the complexity of adjusting parameters in the network model training process.

5. Summary and Prospect
In this article, 1. The experiment is only a comparative experiment of data sampling on the network, lacking the comparison of data collection experiments; 2. Due to other factors such as weather and
artificial factors, various meteorological data will be distorted to some extent. 3. The lack of error analysis and evaluation affects the accuracy. 1. Compare the test results with other optimization algorithms;

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

References
[1] Liu Haoran, Zhao Cuixiang, Li Xuan, et al. Research on a Neural Network Optimization Algorithm Based on Improved Genetic Algorithm [J]. Chinese Journal of Scientific Instrument, 2016, 37(07): 1573-1580.

[2] Gong Yulin, Wen Dahua, Zhang Jinghe. Study on Water Quality Evaluation Method Based on Particle Swarm Optimization Algorithm and RBF Neural Network [J]. Journal of Changchun University of Science and Technology (Natural Science Edition), 2014, 37(05): 141-145.

[3] Wang Lanxia. Study on PM10 Pollution Forecast Based on RBF Neural Network [D]. Xi'an: Xi'an University of Architecture and Technology, 2008.

[4] Yin Anqi, Lin Yuanyi, Lin Weijun, et al. Prediction of Daily Average PM10 Concentration in Guangzhou Based on PSO-BP Neural Network [J]. China Health Statistics, 2016, 33(5): 763-766.

[5] Liang Dawei. Application of RBF Neural Network in Wind Speed Forecast of Wind Farm [D]. Chongqing: Chongqing University, 2015.

[6] M. Song Sumin, Kuang Wenzhen, Xu Li, et al. Application of RBF Neural Network in Railway Freight Volume Forecast [J]. Railway Computer Application, 2017, 26(1): 47-51.