

Prediction of PM_{2.5} concentration based on BP neural network optimized by bee colony algorithm

Jie Heng, and Min Li*

School of information engineering, Mianyang Normal University, Mianyang 621000, Sichuan

Abstract. According to the ambient air pollutants data and meteorological conditions data of Mianyang City in 2017, the BP neural network model based on MATLAB is established to predict the daily average PM_{2.5} concentration of Mianyang City in the next two days. However, the traditional BP network has the disadvantages of slow convergence speed and easy to fall into local optimum. In order to improve the prediction accuracy of the model, an optimization algorithm is added to the prediction model to avoid the model falling into local minimum. In this paper, the bee colony algorithm is added to the prediction model to improve the accuracy of BP neural network prediction model. The data from January to November are used for training, and the data from December are used as the verification results. The results show that the optimization model can accurately predict the daily average PM_{2.5} concentration of Mianyang City in the next two days, which provides a new idea for the prediction of PM_{2.5} concentration of the city, provides a theoretical basis for the early warning and decision-making of air pollution, and also provides more reliable prediction services for people's daily travel.

Keywords: Haze, Prediction model, BP neural network, Bee colony algorithm.

1 Introduction

In 2016, 253 out of 338 cities in China still had substandard air quality, accounting for 74.8 percent of the total, of which 80.2 percent were days with PM_{2.5} as the primary pollutant, triggering the red alert. In 2018, the proportion of excessive PM_{2.5} days exceeded 50 percent in Beijing, Tianjin and Hebei, and the average concentration of PM_{2.5} in the Yangtze River Delta economic urban agglomeration was 44 g/m³ in the whole year of 2018, falling short of the National Air Quality Standard^[1].

This paper chooses the city of Mianyang, in the northwest of Sichuan Basin, as the research area. Mianyang coordinates: 103 degrees, 45 minutes east longitude-105 degrees, 43 minutes east longitude, 42 minutes 30th parallel north-33 degrees, 03 minutes east longitude, at the edge of the Tibetan Plateau and Chengdu Plain. The air pollutant data in

* Corresponding author: 562863252@qq.com

this paper were collected from the Mianyang Environmental Monitoring Station, including the daily mean of SO₂, CO, O₃, PM₁₀, PM_{2.5} and AQI. The data of meteorological conditions are obtained from Mianyang Meteorological Bureau, including the daily average of the maximum temperature, the minimum temperature, the average temperature, the pressure, the maximum wind speed, the maximum wind speed, the average wind speed and the relative humidity. The data taken are for the whole year from 1 January 2017 to 30 December 2017. In this paper, data from January to November of the same year were selected for training, and data from December of the same year were used as the verification results. This paper focuses on the establishment of neural network, explores the optimization method of neural network model, compares the optimized model with training, analysis and comparison to get the best prediction model. First, collect and sort out the data. The vacancy data was checked and eliminated. After that, the prediction model of BP neural network is established [2]. In the next step, we add the bee colony algorithm into the prediction model, and make use of the advantages of the parallel global search of the optimization algorithm to improve the stability, learning and efficiency.

2 BP neural network prediction model

2.1 Principle of BP neural network.

Using BP artificial neural network as PM_{2.5} concentration prediction model, its network topology structure is shown in figure 1 from the diagram, we can see that the basic structure of the network structure is divided into three layers, they are the input layer, the hide layer and the output layer^[3].

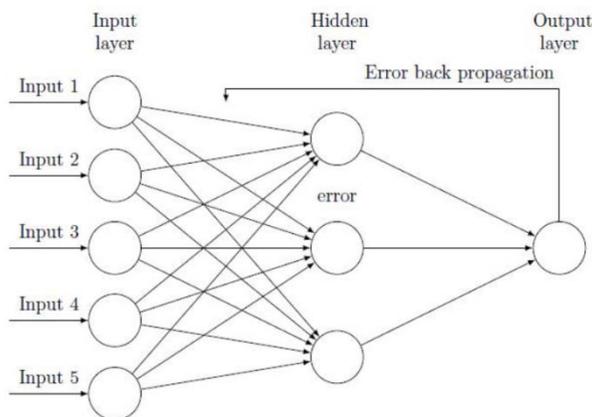


Fig. 1. Neural network structure diagram.

BP neural network, also called Back Propagation Neural Network (BP), is an error Back Propagation algorithm. It is characterized by forward transmission of the input signal and forward propagation of the input signal along the neural network. If there is an error signal that does not satisfy the condition in the process, the error will propagate backward along the neural network, the forward and backward propagation stages are repeated until the calculated error is within the set condition and the training is terminated. The error signal propagates in the reverse direction, and the initial parameters of the model are corrected by using the gradient descent to transmit the error signal, so that the output value of the network keeps approaching the sample data^[4].

2.2 The limitation of BP neural network

BP Neural Network solves the problem of multi-layer weight adjustment of nonlinear continuous function by using the advantages of its model. However, the traditional BP neural network still has some limitations in the prediction, such as: the ability to optimize the initialization parameters is limited, BP algorithm itself is a local search method, its network parameters are randomly generated by the system initialization, this will have an impact on the Internet. Secondly, the convergence speed is slow, because the complexity of the problem to be solved is different, for some more complex problems, BP algorithm may need longer training time, its convergence speed will be slow^[5]. Finally, the structure of BP network is uncertain. At present, three-layer neural network structure is commonly used, but the number of middle layer or hidden layer can be single layer or multi-layer, which leads to the uncertainty of network topology structure. The current research shows that the number of hidden layers of BP network, we can only rely on experience to explore some experience and continuous experimental verification to determine, the lack of theoretical basis required. If there are too many hidden layers in the model, the training time will be prolonged without improving the precision, and if there are too few hidden layers, the whole neural network will converge too early.

3 Optimization design based on bee colony algorithm

3.1 Principles of swarm algorithm

The Artificial Bee Colony Algorithm (ABC), proposed by Karaboga in 2005, is based on the idea that bees work together to gather honey. In the searching process, this algorithm only takes fitness function as the evolutionary basis and optimizes the initial value^[6]. Artificial bee colony algorithm is a kind of artificial intelligence model which simulates the self-organization and self-adaptation behavior of biological community in nature without control, its main advantage is that it searches in the process of each global iteration, so that the probability of finding the optimal solution is greatly increased, the optimized algorithm was added to the BP neural network model to predict the average daily concentration of PM_{2.5} over the next two days in Mianyang.

3.2 The training process of bee colony algorithm

(1).Bee Colony initialization, the algorithm to initialize the parameters set. The parameters include: the number of solutions (NS), that is, the number of honey sources, the number of bees, the maximum limit number (limit) and so on.

$$x_{ij}=L_{minj} + \text{rand}(0,1)(U_{maxj} - L_{minj}) \quad (1)$$

U_{maxj}, L_{minj} , represents the upper and lower limits of the j dimensional search, respectively, at which point the scouts begin to search for the nectar source after it is produced.

(2).Scout Stage. The honey source location was updated by formula (2).

$$v_{ij}=x_{ij} + \Phi(x_{ij} - x_{kj}) \quad (2)$$

In the formula, J is a random number in $[1, J]$, $k = [12... N_s]$, which means a random number in the range of $[-1,1]$ is generated randomly in the NS.

(3). Following the bee phase. After learning the information of the source, the selection probability of the different source or the different solution is calculated according to the formula (3).

$$p_i = \frac{\text{fit}(x_i)}{\sum_{k=1}^{NS} \text{fit}(x_k)} \quad (3)$$

where P_i is the selection probability of the I solution, for the moderate solution, the larger the fit value, the higher the fitness, the better the quality of the honey^[7].

(4). Determine if honey source X_i needs to be abandoned.

(5). Scout Bees continue to update the nectar source location according to form (2).

Repeat step (3) (4)

(6). To find the perfect source of nectar. At this point, the search for the optimal solution is completed, and the optimal solution is set as the initial weight and threshold of BP neural network.

4. Experiment and analysis

4.1 Prediction data of traditional BP neural network

The average error percentage of PM_{2.5} concentration in Mianyang is 8.9% by using the traditional BP neural network. The biggest error is 31.9% on December 19, the lowest error is 0.9% on December 13, eleven of those days had a percentage error of 5% or less. As figure 2 shows:

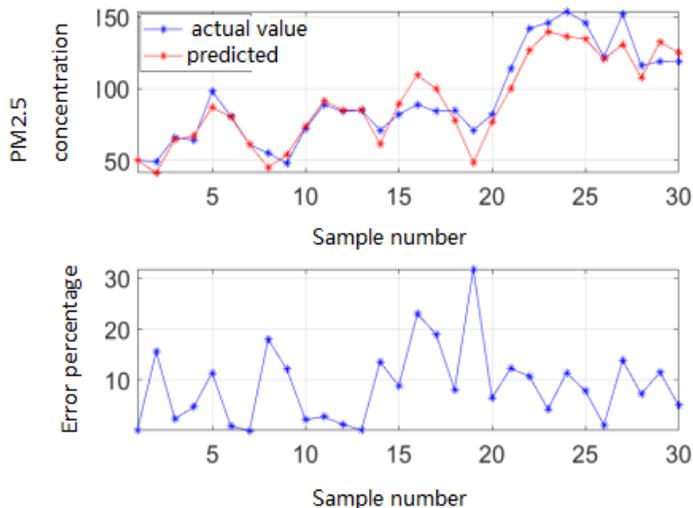


Fig. 2. Error map of PM_{2.5} concentration predicted by BP neural network on the first day of the future in Mianyang.

By using BP neural network, the error of the second day in Mianyang is larger than that of the first day. As shown in figure 3, the average percentage error of predicting PM_{2.5} concentrations in Mianyang by BP network on the 2nd day was 15.2% , with the largest error of 35.9% on the 9th of December, and the prediction error of 5% or less on the 8th day.

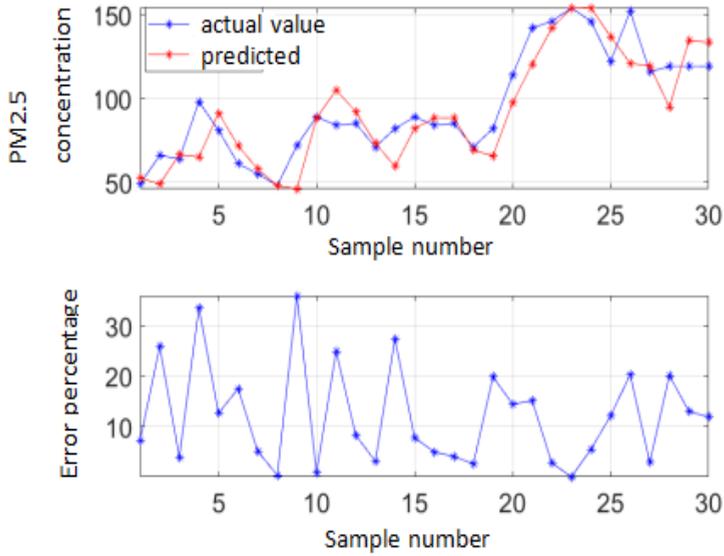


Fig.3. Error map of PM_{2.5} concentration predicted by BP neural network on the 2nd day in Mianyang.

4.2 Prediction data of BP neural network based on bee colony algorithm

As can be seen from figure 4, the BP neural network optimized by the bee colony algorithm was able to predict the average daily concentration of PM 2.5 on the first day in Mianyang with an average error of 4.2%, and the percentage error of traditional models in predicting PM_{2.5} concentrations in Mianyang on the first day of the future was reduced by 4.7% .

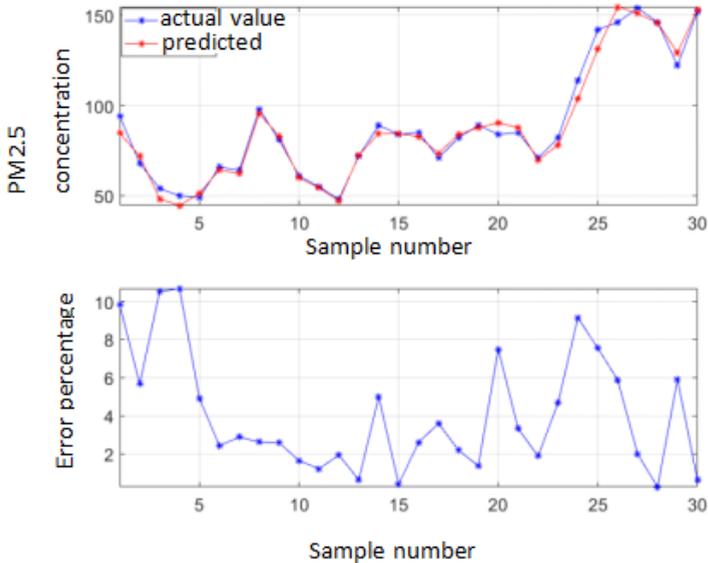


Fig. 4. ABC-BP forecast of PM_{2.5} concentration and error on the first day in Mianyang.

As can be seen from figure 5, the ABC optimized BP neural network predicted the average daily concentration of PM_{2.5} over the next 2 days in Mianyang. As can be seen, the

data error rate of the second day of prediction was not as accurate as that of the first day, but it is still much better than the traditional model. The average error is 9.5% after the ABC optimization, which is 5.7% lower than the traditional model.

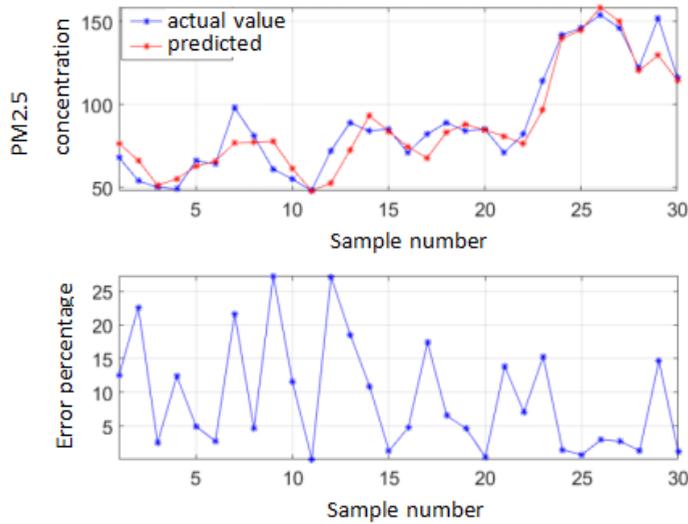


Fig. 5. ABC-BP projection of PM_{2.5} concentration and error for the second day in Mianyang.

5 Conclusion

In this paper, a prediction model of BP Neural Network based on Bee Colony Algorithm is proposed. The most commonly used evaluation index for prediction results is the loss function, there are MSE, RMSE, MAE, MBE, as shown in figure 6. When the model is run, the BP and ABC-BP loss function values are compared in the command line, and the higher the loss function values are, the worse the predictive power of the model [8].

```

Optimal validation data in future day 1 BP forecast
MSE: 50.02 RMSE: 7.07 MAE: 5.95 MBE: 0.09
Optimal validation data in abc-bp forecast on the first day of the future
MSE: 21.50 RMSE: 4.64 MAE: 3.53 MBE: -0.90
Optimal validation data in BP on the next day
MSE:160.03 RMSE:10.18 MAE:9.02 MBE: -0.97
Optimal validation data in abc-bp forecast for the next day
MSE:76.60 RMSE:6.83 MAE:7.52 MBE: -0.67
Worst validation data in BP forecast for the first day in the future
MSE: 1888.10 RMSE:43.45 MAE: 37.77 MBE: -34.60
Worst validation data in abc-bp forecast on the first day of the future
MSE: 272.88 RMSE: 16.52 MAE:10.65 MBE: -7.14
Worst validation data in BP forecast for the next day
MSE: 879.04 RMSE:29.65 MAE:24.62 MBE:4.96
Worst validation data in abc-bp forecast for the next day
MSE: 333.43 RMSE: 18.26MAE:14.32 MBE: -3.25
    
```

Fig. 6. BP, ABC-BP loss function value contrast chart.

When the model trains the prediction ability of a network, our goal is to find the point where the loss function is at the minimum. These functions can be minimized when the predicted value of the model is infinitely close to the true value of the sample. The

evaluation criteria of the neural network model, I chose the value of the loss function to measure the accuracy of the model prediction, in the two-day prediction of PM_{2.5} concentration in Mianyang, whether it is the best or worst case, according to the model evaluation formula to make qualitative analysis of the model, prediction performance, we can intuitively see that the BP neural network prediction model optimized by the bee colony algorithm is more stable and the prediction accuracy is the highest.

References

1. Helen Thompson, Zhen Zhang. How does urban form influence PM 2.5 concentrations: Insights from 350 different-sized cities in the rapidly urbanizing, 2010[J]. *Cities*, 2010, 98.
2. James T. Kelly, Adam Reff, Brett Gantt. A method to predict PM2.5 resulting from compliance with national ambient air quality standards [J]. *Atmospheric Environment*, 2017, 162.
3. Weihao Tang, Yanying Li, Yang Yu, Zhongyu Wang, development of models predicting biodegradation rate rating with multiple linear regression and support vector machine algorithms [J]. *Chemosphere*, 2020, 253.
4. Sobhan Kumar Kompalli, Sreedharan Krishnakumari Satheesh, et al. Seasonal contrast in size distributions and mixing state of black carbon and its association with PM chemical composition from the eastern coast of India. 2020, 20(6):3965-3985.
5. Chengbin Du, Wenhui Zhao, Shouyan Jiang, Xiaodong Deng. Dynamic XFEM-based detection of multiple flaws using an improved artificial bee colony algorithm [J]. *Computer Methods in Applied Mechanics and Engineering*, 2020, 365.
6. Pankaj Goswami, Haina Brahma. River Water Level Prediction Modelling using Artificial Neural Network and Multiple Linear Regression. 2019, 9(6):23-31.
7. Genetic Algorithm and BP Neural Network [J]. *Applied Mechanics and Materials*, 2014, 3360 (599):726-730.
8. James T. Kelly, Adam Reff, Brett Gantt. A method to predict PM2.5 resulting from compliance with national ambient air quality standards [J]. *Atmospheric Environment*, 2017, 162.