

Research on Application of BP Artificial Neural Network in Prediction of the concentration of PM2.5 in Beijing

Yuanhua CHEN^{1,a,*}, Lisha WANG¹ and Lina Zhang¹

¹ CECEP Liuhe and Talroad Environmental Technology Co., Ltd, Beijing, 100082, China.

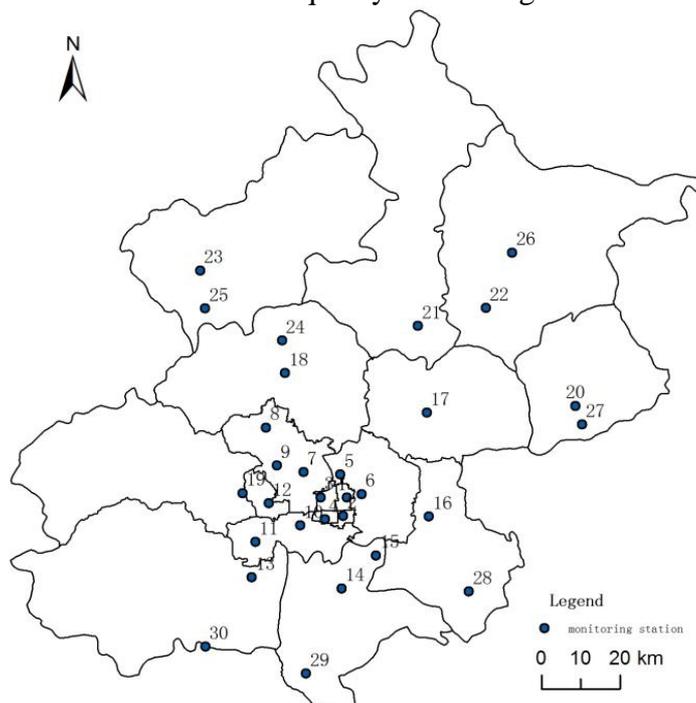
^ayuanhua.chen@talroad.com.c

Keywords: BP Artificial Neural Network, prediction model, PM 2.5 concentration.

Abstract. According to the PM2.5 hourly average monitoring concentrations of Beijing environmental air monitoring stations during may and June 2014, the BP artificial neural network prediction model was built and tested, which verifies the feasibility and accuracy of BP artificial neural network model to predict PM2.5 hourly average concentration. And the forecasting results show that the reasonable structure and suitable training algorithm are the key factors to the accuracy, reliability, and consistency of PM2.5 concentrations.

Material and methods

Data source. Data sources and processing. The hourly air quality data, including PM10, SO2, O3, NO2, CO and PM2.5 concentrations, from 30 air quality monitoring stations all over Beijing during May and June 2014 were collected from the official website of Beijing Municipal Environmental Monitoring Center[1]. The sites of the 30 air quality monitoring stations were shown in Fig. 1.



1.Dongsi, Dongcheng 2.Temple of Heaven, Dongcheng 3.Guanyuan, Xicheng 4. West Wanshou Nishinomiya, Xicheng 5.Olympic Center, Chaoyang 6.Agricultural Exhibition Hall, Chaoyang 7.Wanliu, Haidian 8.Haidian Northern New Area, Haidian 9.Beijing Botanical Garden, Haidian 10.Fengtai Garden, Fengtai 11.Yungang, Fengtai 12.Gucheng, Shijingshan 13.Liangxiang, Fangshan 14.Huangcun, Daxing 15. Yizhuang Development Zone, Daxing 16.Tongzhou New Town 17.Shunyi New Town 18.Changping Town 19.Mentougou Longquan Town 20.Pinggu Town 21.Huairou Town 22.Miyun Town 23.Yanqing Town 24.Dingling, Changping 25.Badaling, Northwest of Beijing 26.Miyun Reservoir, Northeast of Beijing 27.Donggaocun, East of Beijing 28.Yongledian, Southeast of Beijing 29.Yufa, South of Beijing 30.Liulihe, Southwest of Beijing

Fig. 1 The distribution of air quality monitoring stations in Beijing

The hourly temperature, relative humidity, atmosphere pressure, wind direction and wind speed data during May and June 2014 were collected from the Weather Underground Website[2]. The non-numerical data: wind direction was processed to numerical data, generating the sines of those directions, in order to facilitate data analysis, as shown in Table 1.

Table 1. Change of non-numerical to numerical data for wind direction

| Original data | Angle value of direction(θ) | Processed data($\sin \theta$) |
|---------------|--------------------------------------|---------------------------------|
| North | 0 | 0 |
| NNE | $\pi/8$ | 0.382 |
| NE | $\pi/4$ | 0.707 |
| ENE | $3\pi/8$ | 0.924 |
| East | $\pi/2$ | 1 |
| ESE | $5\pi/8$ | 0.924 |
| SE | $3\pi/4$ | 0.707 |
| SSE | $7\pi/8$ | 0.382 |
| South | π | 0 |
| SSW | $9\pi/8$ | -0.382 |
| SW | $5\pi/4$ | -0.707 |
| WSW | $11\pi/8$ | -0.924 |
| West | $3\pi/2$ | -1 |
| WNW | $13\pi/8$ | -0.924 |
| NW | $7\pi/4$ | -0.707 |
| NNW | $15\pi/8$ | -0.382 |

Data preprocessing. In order to make the neural network has a higher training efficiency and prediction accuracy, it needs normalized processing of raw data, a new set of dimensionless ones is transformed by the linear function, the normalization method is as follows:

$$X = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Where X is the normalized value, x is the raw data, x_{\max} is the maximum of raw data, and x_{\min} is the minimum of raw data.

BP artificial neural network. Through learning from the input and output data under study, the neural network system can get a nonlinear mapping relationship between input and output variable. Among many types of neural networks, the BP (Back propagation) neural network based on error back propagation algorithm, is the most widely used of artificial neural networks.

The basic structure of BP neural network is composed of an input layer, one or more hidden layers and an output layer. The neuron in each layer only has connection with its adjacent layer's neurons, while the neuron has no connection with ones in the same layer, as shown in figure 2.

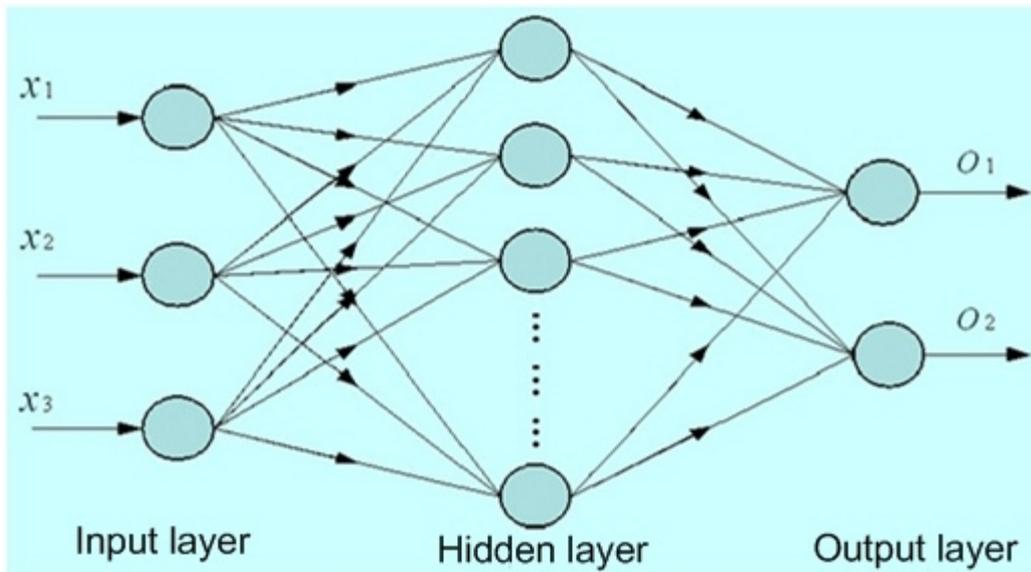


Fig.2 the structure of BP neural network

The basic process to establish a BP artificial neural network model is as follows. Firstly, a given input data will be passed from input layer to output layer units via hidden layer units, and then the result will be obtained after dealt with the output layer, which is the forward propagation process; Secondly, calculation error between actual output and desired output will be fed back along the network and correct connection weights, which is the error back propagation process; Finally, the calculation will continue until the errors of other inputs reduced to an acceptable level or the set time of training achieved[3].

Establishment of BP neural network. Step 1: design the input layer and output layer of BP neural network. This prediction model is based on hourly average monitoring concentration of air quality indexes and meteorological parameters during a certain number of days, including PM10, SO2, O3, NO2, CO, the temperature, humidity, wind speed, wind direction and pressure, and the aim of the model is to forecast PM2.5 hourly average concentration and AQI index in the future day. According to the design solution, the input matrix (P) and the output matrix (T) are shown as follows.

$$P = \begin{bmatrix} x_1(1) & x_1(2) & \dots & x_1(9) & x_1(10) \\ x_2(1) & x_2(2) & \dots & x_2(9) & x_2(10) \\ \dots & \dots & \dots & \dots & \dots \\ x_{n-1}(1) & x_{n-1}(2) & \dots & x_{n-1}(9) & x_{n-1}(10) \\ x_n(1) & x_n(2) & \dots & x_n(9) & x_n(10) \end{bmatrix}, T = \begin{bmatrix} y_1(1) & y_1(2) \\ y_2(1) & y_2(2) \\ \dots & \dots \\ y_{n-1}(1) & y_{n-1}(2) \\ y_n(1) & y_n(2) \end{bmatrix}^T$$

Thus it can be seen that the input layer is a single layer with ten neurons, and the output layer has two neurons in the single layer. The input and output data are all ones by normalized data processing of the original data.

Step 2: design the hidden layer of the neural network. Though contrasting performance of different network structure, the hidden layer adopts a two-layer structure, and the number of neurons in each hidden layer is 22-8.

Step 3: choose the training algorithm. In many training algorithms, the trainlm algorithm commonly used for training neural networks is selected.

Evaluation function of prediction performance. In this study, the mean square error (MSE) is used to evaluate the performance of BP neural network model, the evaluation function expression as follows:

$$MSE = \sum_{i=1}^N \frac{(y_i - \hat{y}_i)^2}{N} \quad (2)$$

Where \$x_i\$ is monitoring value and \$y_i\$ is forecasting value.

Results and discussion

By using 10-22-8-2 network structure and trainlm algorithm, BP artificial neural network model is established, the training sample contains 529 valid data sets in May. After 1000 iterations, the network can accomplish convergence, and the minimum mean square error (MSE) of network is 0.0058. The curves of monitoring and forecasting concentration time series is shown in figure 3, while figure 4 shows the absolute error curve.

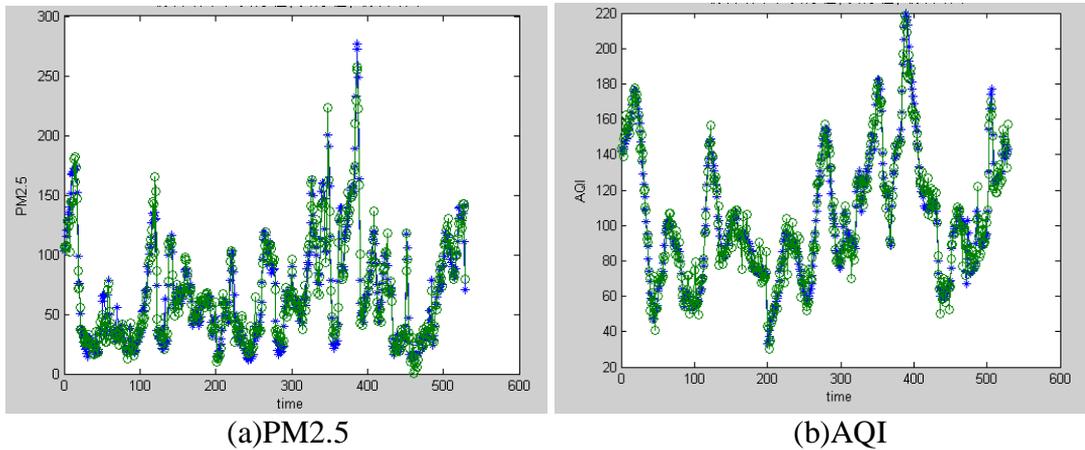


Fig. 3 monitoring and forecasting concentration of PM2.5 and AQI

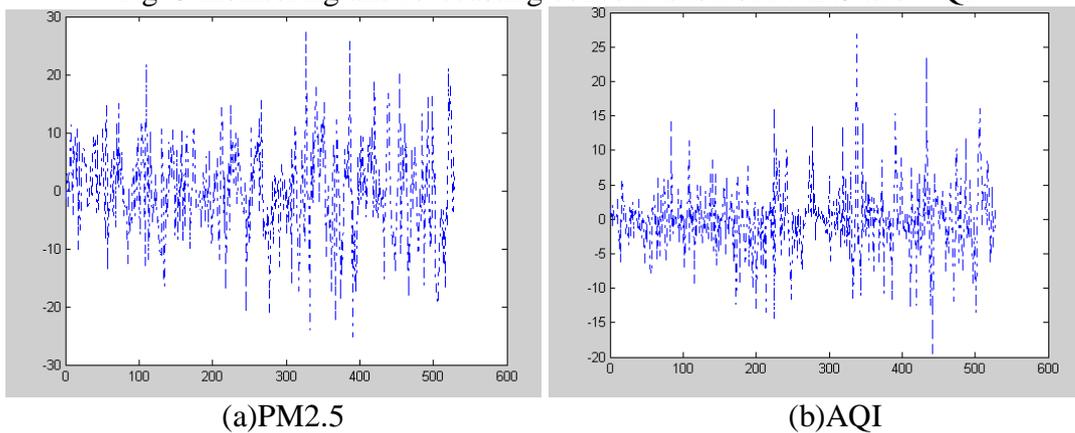


Fig. 4 the MSE of forecasting result of PM2.5 and AQI

Obviously, the trend of PM2.5 and AQI prediction values is highly matching monitoring ones in time sequence. The predict performance evaluation function is in the acceptable range except individual abnormal points, which demonstrates the BP network model's predicting effects are very good.

Conclusion

This prediction of established BP artificial neural network model for PM2.5 and AQI hourly average concentration in Beijing shows that the BP artificial neural network can well reflect the dynamic change regularity of PM2.5 concentrations, and it is feasible to forecast PM2.5 concentration based on BP artificial neural network. To achieve higher prediction accuracy, the suitable structure setting and training algorithm selection of BP neural network model is significant.

References

- [1] Information on <http://zx.bjmemc.com.cn>.
- [2] Information on <http://www.wunderground.com>.

[3] Reid B, Arnold C R. Use of ozone for water-treatment in recirculating water raceway system[J]. Progressive Fish Culturist, 1994, 56(1) :47-50.