ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

Prediction Model of Air Pollution Index in Anyang City in Winter based on BP Neural Network

Haoran Fu[1, a], Huahui Li[2, b]

[1]School of Economics, Anyang Normal University, Anyang 455000, China.

[2]School of Mathematics, Anyang University, Anyang 455000, China.

[a]hao3681@foxmail.com, [b]853391460@qq.com

Abstract

In this paper, the winter air pollution index of Anyang city is taken as an example, and the prediction model of winter air pollution index of Anyang city based on BP neural network is constructed by using the theoretical knowledge and application method of BP neural network. By simulating the process of biological nerve processing information, BP neural network makes the signal propagate forward and the error propagate backward, so that the network can be fully trained and can carry out large-scale calculation. In order to build a neural network that can accurately predict the air pollution index, this paper selects the daily air pollution index of Anyang City in winter of 2017 and 2018 to train the network, and takes the daily air pollution index of winter of 2019 as the test data. Tests show that the model is feasible to predict the air pollution index after being fully trained, and has high accuracy for prediction except for severe pollution. Therefore, BP neural network can be used as a simple, feasible and effective tool to study the prediction model of air pollution index.

Keywords

Air pollution index (API); BP neural network; Network training.

1. INTRODUCTION

1.1. Air Pollution Index

Air Pollution Index (API) simplifies the concentration of several air pollutants that can be monitored into a unified conceptual index value. Air pollution index classifies the air quality status of a region, which is suitable as an important index to describe the air quality status of a city for a period of time [4].

Air pollution index (API) is a daily monitoring index related to people's outdoor environment besides weather forecast. The air pollution monitoring values we see every day, like weather forecast, are based on the simulation results of the forecasting system, that is, except for some extreme cases, the air pollution index is also based on a specific model and implemented in the simulation system. The model also has differences in stability and prediction accuracy. For the general public, the function of air pollution index prediction is roughly the same as that of weather forecast, which is convenient for us to take effective preventive measures in time to reduce the impact of possible air pollution on human health.

The air pollution index is divided into six grades, namely excellent air quality, good air quality, mild pollution, moderate pollution, severe pollution and severe pollution. The corresponding air pollution index ranges from 0 to 50, 51 to 100, 101 to 150, 151 to 200, 201 to 300 and over 300 in turn. The influence of pollution on the health of different people will expand with the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

increase of air pollution index. The range of specific air pollution index and its corresponding impact on human health are shown in Figure 1.

Air quality index	Air quality status	Represent color	Impact on health	Proposed measures to be taken
0-50	Excellent		Satisfactory air quality, basically no air pollution and no harm to health.	All kinds of people can take part in outdoor activities and breathe fresh air.
51-100	Good		Except for a few people who are particularly sensitive to certain pollutants, it will not cause harm to human health.	Except for a few people who are particularly allergic to certain pollutants, other people can do outdoor activities normally.
101-150	Light pollution		The symptoms of sensitive people will be slightly aggravated, which has no obvious effect on healthy people.	Children, the elderly and patients with heart disease and respiratory diseases should minimize outdoor activities with heavy physical exertion.
151-200	Moderate pollution		The symptoms of sensitive people are further aggravated, which may affect the heart and respiratory system of healthy people.	Children, the elderly and patients with heart disease and respiratory diseases should try to reduce going out and staying indoors, and the general population should reduce outdoor sports appropriately.
201-300	Severe pollution		The poor air condition will do serious harm to everyone's health.	Children, the elderly and patients with heart disease and respiratory diseases should stay indoors and stop outdoor sports, and the general population should minimize outdoor activities.
300	Serious pollution		The air condition is extremely poor, and all people's health will be seriously endangered.	Children, the elderly and patients should stay indoors to avoid physical exertion. Except those with special needs, the general population should try not to stay outdoors.

Figure 1. Scope and influence of air pollution index (AQI)

1.2. Present Situation of Air Pollution in Anyang

Anyang is located in the northernmost part of Henan Province, at the intersection of Henan, Hebei and Shanxi provinces. Because of its special geographical location, Anyang has been included in the "2+26 Urban Air Pollution Prevention and Control Plan" by the State Council. The plan requires Beijing, Tianjin, Hebei and surrounding areas to make concerted efforts to solve the environmental pollution problem, coordinate and control the total emissions of greenhouse gases and jointly improve the air quality.

In 2017, the Ministry of Environmental Protection in Anyang City, informed criticism, was informed that Beijing, Tianjin, Hebei and its surrounding areas suffered from large-scale air pollution. In this large-scale air pollution, Anyang is not only the first polluted place, but also the area with the longest pollution time. As a matter of fact, Anyang is not the first time to be

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

informed criticism by the Ministry of Environmental Protection because of its poor prevention and control of air pollution. As early as 2013-2014, Anyang was one of the cities with the most serious air pollution in Henan Province. In the following years, the air pollution situation in Anyang has not been significantly improved, and continuous large-scale air pollution often occurs in winter. It can be seen that whether Anyang's air pollution index is up to standard is even more important for the implementation of air pollution prevention and control in Beijing, Tianjin and Hebei and surrounding areas. The main reasons why Anyang has become the city with the longest duration of heavy pollution are as follows: many enterprises have exceeded emissions and failed to implement emission reduction measures as required.

After Anyang Municipal Government strengthened the governance measures, strengthened the construction of monitoring and control, strictly refined emergency management and control, and improved the environmental supervision grid. According to the announcement of Anyang Municipal People's Government Office, the urban air pollution index in 2018 has improved compared with the same period of the previous year, but the air pollution index remains high, which seriously affects local residents' daily outdoor activities.

2. DATA MATERIALS

The air pollution index data of Anyang City in this paper are all from the public data of the Ministry of Environmental Protection. In addition, the air pollution index in winter in Anyang city is of high level in previous years, so it is more valuable to choose the air pollution index in winter as the research object. Therefore, the training data of the model in this paper is the daily air pollution index in winter of 2017 and 2018 (from November 7th of each year to February 4th of next year). The winter of 2019 (November 7th to February 4th of next year) is taken as the test data [5].

3. NORMALIZATION SCHEME

The data selected in this paper has a large span, so it is necessary to process the data to avoid the "over-fitting problem" in calculation, which is also convenient for network training. The commonly used data preprocessing method is standardization. Comparing the data processing effects of two standardized methods-zero mean standard deviation standardization method and normalization method, the fitting effect of normalization method is obviously better than that of zero mean standard deviation standardization method, so normalization method is chosen to process the data [3].

In addition, the value after data preprocessing should be concentrated in the range of 0.2-0.8 to ensure that the established model has strong extrapolation ability. To achieve this goal, the following data preprocessing and anti-preprocessing algorithms are adopted:

$$y = (0.8-0.2)*(x-\min(x))/(\max(x)-\min(x))+0.2$$

$$x = (y-0.2)*(\max(x)-\min(x))/(0.8-0.2)+\min(x)$$

x is the original data of Anyang air pollution index, and y is the result of normalization of Anyang air pollution index.

4. BACK PROPAGATION

4.1. Principle of BP Neural Network

Artificial neural network is an intelligent method to simulate animal neural network, and it is also one of the most basic artificial intelligence. The principle of artificial neural network simulating the thinking mode of human brain is that human beings constantly adjust the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

connection form between the cell structures of the brain through learning, so that the network function can be continuously improved through learning.

The BP (Back Propagation) neural network studied in this paper is a kind of artificial neural network. BP neural network uses reverse error algorithm to train and optimize the network. The learning rule of BP algorithm adopts gradient descent method, so that the network can adjust the weight of each layer through learning and train a lot in a short time. The learning process includes two propagation modes: forward and backward. Under the action of transfer function, the input information propagates forward and the error propagates backward. According to the error returned in the reverse direction, the weights and thresholds of each layer are constantly adjusted, so that the error of the model is continuously reduced in this process, thus obtaining more accurate results. BP neural network contains input layer nodes, output layer nodes and one or more hidden layer nodes, among which hidden layer nodes may also contain multiple layers. For a BP neural network, it can contain multiple intermediate layers at the same time or only one intermediate layer, and we call the BP neural network with only one intermediate layer the most basic three-layer BP neural network. Figure 2 shows its topological structure [1].

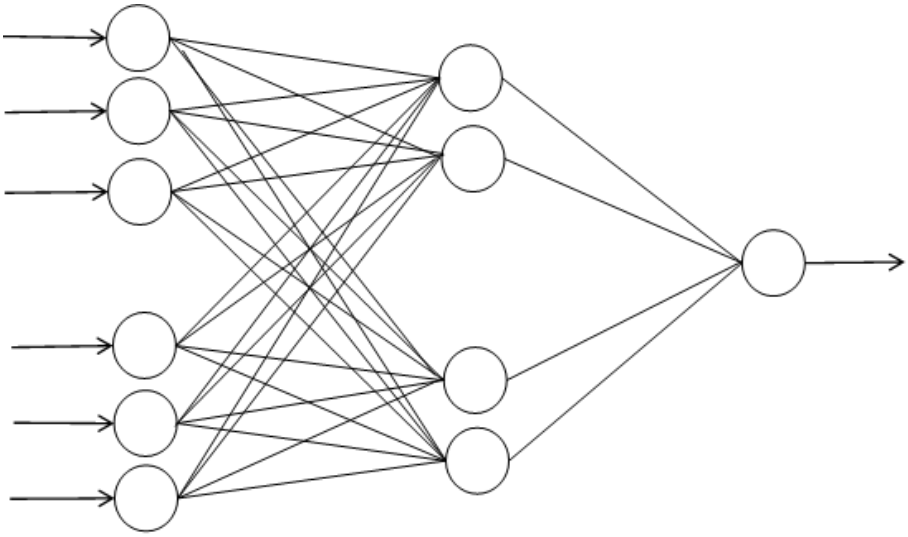

Figure 2. Topological structure of BP neural network

Every neuron of BP neural network has n input values $x_1, x_2, \cdots, x_n$. Its output value is $y = f\left(\sum_{i=1}^{n} \omega_i x_i\right)$, where ω_i is the weight coefficient to be determined. Each neuron needs to transmit information to the next neuron, and the output value needs to meet a certain critical value, which is the threshold value. The output value $f\left(\sum_{i=1}^{n} \omega_i x_i\right)$ is the result of layer-by-layer transfer function. The commonly used transfer function is Sigmoid function, and the function image is shown in Figure 3.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

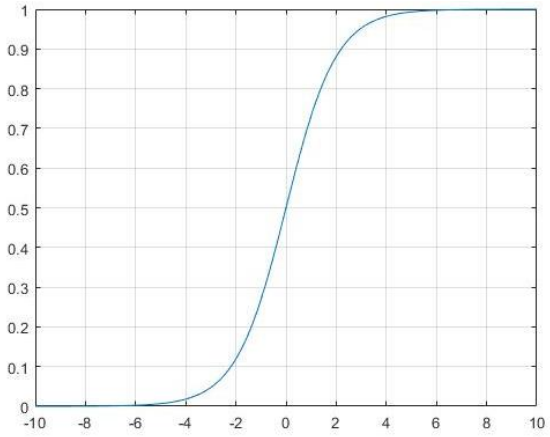

Figure 3. S-type function image

There are mainly two kinds of Sigmoid functions: logsig function and tansig function. Logsig function is a logarithmic function, and its functional expression is $f(x)=1/(1+\exp(-1))-1$; Tansig function is hyperbolic tangent function, and its expression is $f(x)=2/(1+\exp(-2x))-1$. The former trend changes more gently than the latter [2].

4.2. Network Training

The training process of BP neural network is as follows: there are several samples produced by given input-output relationship, and then these data are used to train the neural network. A set of output values is obtained from the input sample data, and there is a big error between the first output value and our ideal output value. Therefore, according to the difference between the actual output value and the ideal output value, we constantly modify the weights and thresholds to reduce the difference. After repeated training for many times, the actual output value and the ideal output value tend to be consistent, thus obtaining the model we need. The flow chart of the training process is shown in Figure 4 [6].

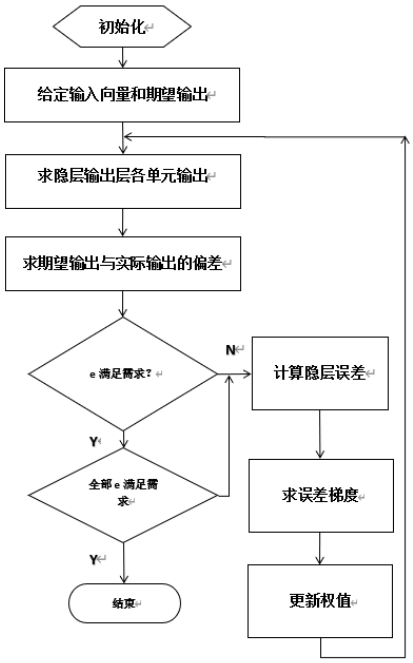

Figure 4. Schematic diagram of training process

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

The research object of this paper is the air pollution index of Anyang city in winter. by processing the historical data of daily air pollution index in winter of 2017 and 2018, the model is established by Neural Net Fitting in Matlab toolbox, and then the preset test data is obtained by combining the historical data with the trained model, so that the model can be popularized to get the actual predicted value. When establishing BP neural network, it is usually necessary to preset the structural parameters of neural network. In this paper, the preset input values, test values, the number of nodes in each layer, the selection of transfer functions and training methods and the reasons are as follows.

First of all, it is the selection of input value test value. In this paper, the daily air pollution index from 2017 to 2018 winter is used as the training sample, that is, the input value, and the data distribution line chart is shown in Figure 5; The daily air pollution index in winter of 2019 is used as the test sample, that is, the expected test value, and the data distribution line chart is shown in Figure 6.

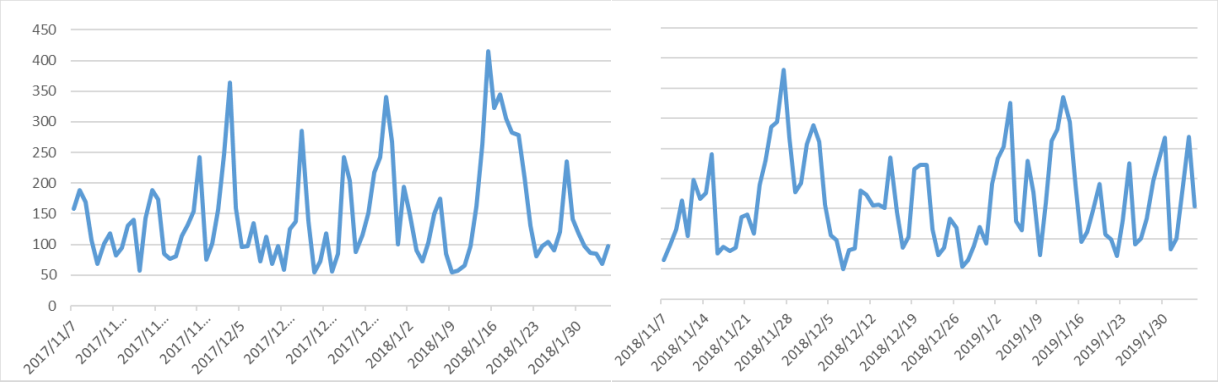

Figure 5. Line chart of daily air pollution index in winter from 2017 to 2018

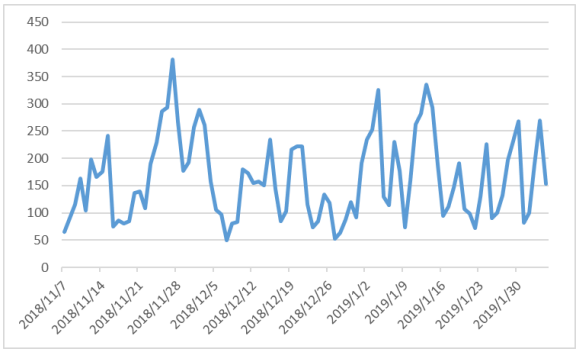

Figure 6. Line chart of daily air pollution index in winter of 2019

According to Kolmogorov's theorem about the number of neurons in the hidden layer of BP neural network, the continuous function in any closed interval can be approximated by a single hidden layer neural network. The air pollution index involved in this paper can be regarded as a continuous function on a closed interval, so the number of hidden layers in the network is one by default in training, that is, the network structure selected in this paper is the most basic three-layer BP neural network.

After the network structure is determined, the transfer function needs to be selected. In this paper, the exhaustive method is adopted when selecting transfer functions, which exhausts the arrangement and combination of two commonly used transfer functions. After testing the influence of different combinations on the effect of neural network, the conclusion about the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

selection of transfer functions is drawn: tansig function is selected in the input-hidden layer, and the fitting effect is best when general linear function is used in the hidden layer-output layer.

Because BP neural algorithm adopts the steepest gradient descent method, the following problems usually exist in the application process: the convergence speed is too slow, and local minima are easy to occur. In order to avoid the above problems in training, this paper chooses Levenberg-Marquardt method of least square fitting to optimize the established model, so that the original learning time is shortened and the effect of network training is better [7].

4.3. Simulation Results

In this paper, the daily air pollution index (180 data in total) of Anyang city in winter in 2017 and 2018 is used as the training sample of neural network, and the BP neural network model of winter air pollution index prediction is established by Neural Net Fitting in Matlab toolbox. The daily air pollution index of Anyang City in winter of 2019 (90 data in total) was used as test samples, and 70% of the data were used as training data, 15% as inspection data and 15% as test data.

After analyzing the training results, the analysis chart of training results is obtained. It can be seen from Figure 7 that the BP neural network established in this paper has smaller error when the air pollution index is lower, and larger error when the air pollution index is greater than 300. Therefore, the BP neural network can well predict the air pollution index except for serious air pollution (index greater than 300), which proves that the model has certain practicability and can well predict the air pollution index except some extreme cases.

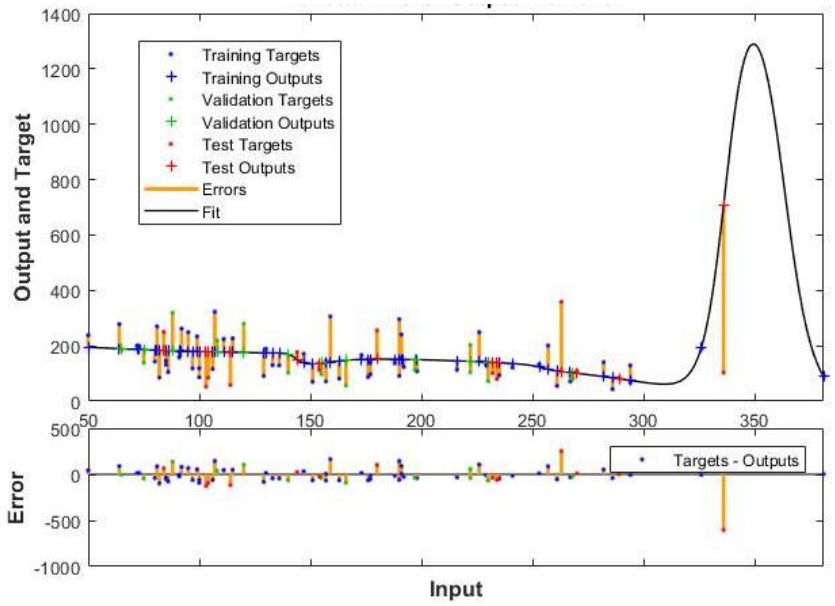

Figure 7. Training result analysis chart

By comparing the difference between the target value and the predicted value, the error histogram is shown in Figure 8. The histogram shows that the error of BP neural network is normal distribution. From the angle of error analysis, it can be shown that the BP neural network has a good training effect on the prediction model of air pollution index in Anyang in winter.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

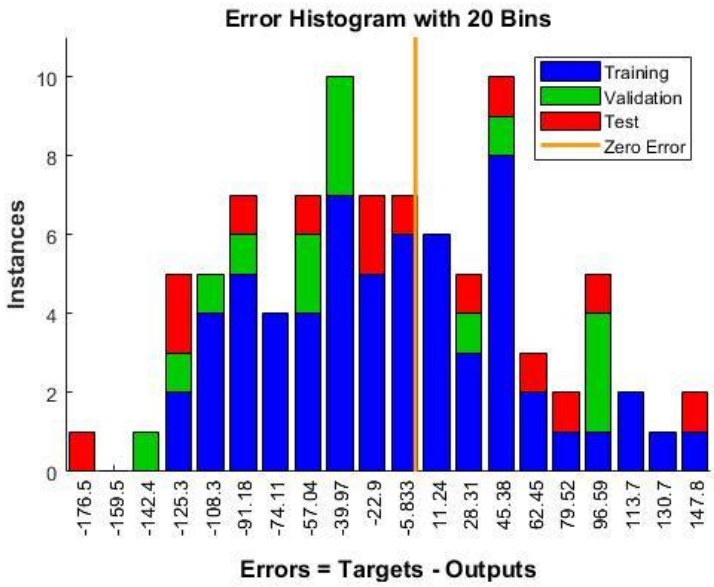

Figure 8. Error histogram

The gradient of test data and learning times are shown in Figure 9. The gradient and learning times of the verification data recorded 11 times, and according to the records, it can be concluded that it is an ideal state in the sixth hidden layer unit.

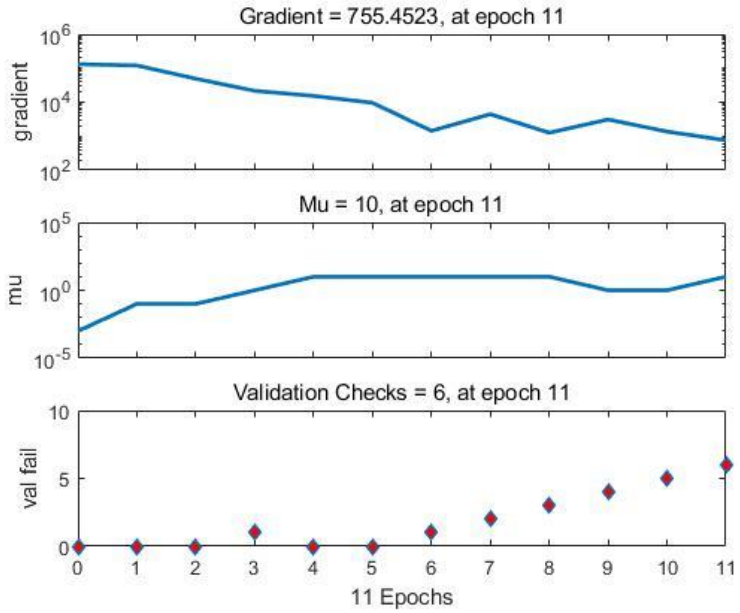

Figure 9. Gradient of test data and learning times

See figure 10 for the training record. The blue curve in the figure represents the performance of mean square error (MSE) in the training process of BP neural network in each generation. The green curve represents the performance of the mean square error (MSE) of the cross-validation process in each generation. The red curve represents the performance of the mean square error (MSE) in each generation during the test, that is, the training result of the BP neural network. The intersection of dashed lines shows that the results of BP neural network training to the fifth generation are the best.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

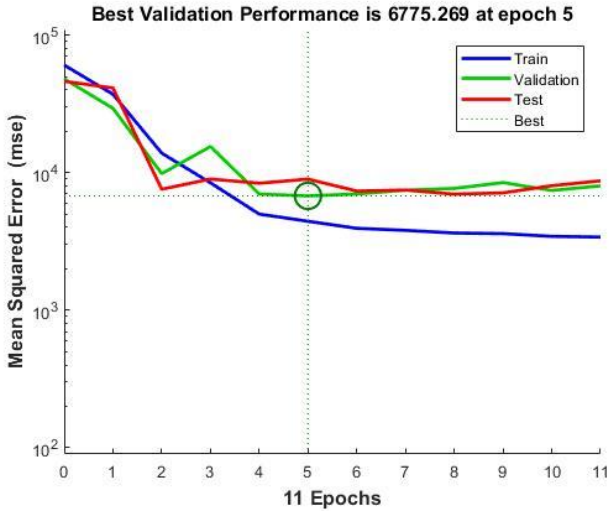

Figure 10. Relationship between gradient of validation data and learning times

The quantitative relationship between the input value and the target value is determined by regression analysis. in fig. 11, there are training data, test data and regression analysis between all data and the target value. the r value of each data must be compared with 1. it can be seen that there is a certain linear relationship between each data and the target value, which shows that the BP neural network obtained in this paper can predict the air pollution index.

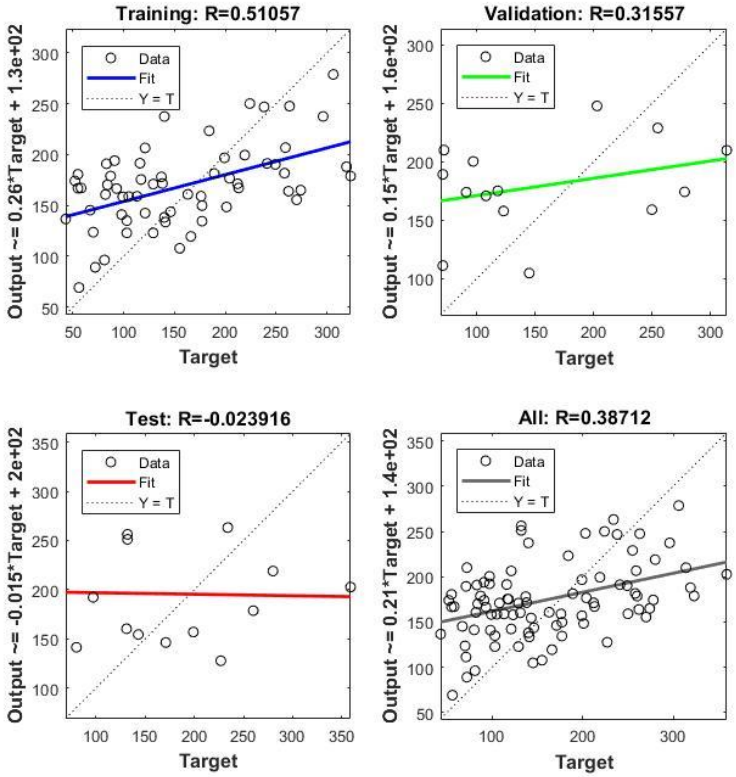

Figure 11. Regression analysis chart

The forecast situation of BP neural network after training is analyzed from five angles. The results show that the daily air pollution index in winter of 2017 and 2018 is used as the training data, and the daily air pollution index in winter of 2019 is used as the test data, which has a

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

good training effect on the forecast model of Anyang's winter air pollution index based on BP neural network.

By analyzing the process and results of establishing BP neural network model, the results show that the neural network model can be used to describe the nonlinear relationship between air pollution index and its influencing factors after full training, and the model can obtain higher prediction accuracy except for the case of serious air pollution (index greater than 300). Therefore, it is simple and feasible to apply the fully trained BP neural network to the prediction of air pollution index, except for some extreme cases with large errors.

4.4. BP Neural Network Prediction

The trained BP neural network is used to predict the daily average concentration of PM_2.5 from January 1 to 6, 2019. The prediction results are shown in the following table and the following figure.

Table 1. Comparison of actual and predicted daily average concentration of $PM_{2.5}$ from January 1 to 6, 2019

Time	1	2	3	4	5	6
predicted value	124.4361	129.7966	130.1584	130.5212	128.8851	134.2496
Actual value	126	132	130	128	129	132

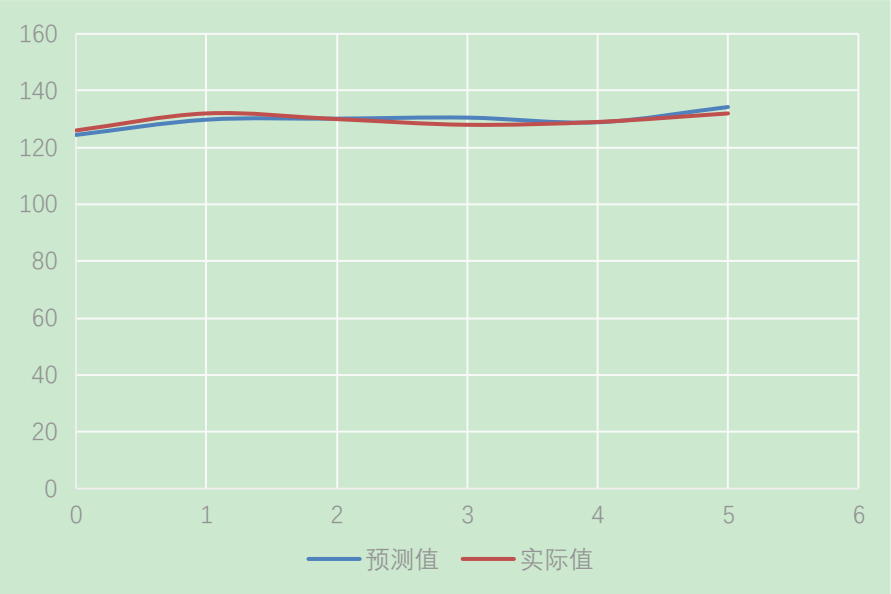

Figure 12. Comparison of $PM_{2.5}$ actual and predicted values from January 1 to 6, 2019

Comparing the real-time predicted value with the actual value, the errors are 1.5639, 2.2034, 0.1584, 2.5212, 0.1149 and 2.2496, all within 3. It can be seen from the above figure that the fitting degree between the predicted value and the actual value is quite high. In this period, the error of BP neural network prediction is smaller than the actual value, and the overall error is more balanced.

5. SUMMARY

In this paper, a prediction model of air pollution index based on BP neural network is established according to the air pollution index of Anyang City in two winters. The neural network is learned by training sample data, and then the model is tested by testing sample data.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0034

The conclusion shows that the model designed in this paper has smaller error and higher accuracy for air pollution with lower level when it is fully trained. Therefore, the next step still needs to optimize the neural network, make the network fully trained by adding training data, and reduce its prediction error of extreme air pollution, especially the prediction ability of air pollution with serious air pollution (index greater than 300).

REFERENCES

[1] Wang Geng, Wang Minsheng. Modern mathematical modeling methods[M]. Beijing: Science Press, 2008.

[2] Han Zhonggeng. Mathematical modeling method and its application [M]. Beijing: Higher Education Press, 2009.

[3] Zhang Jingxiang, Hu Manfeng, et al. Mathematical model[M]. Beijing: Science Press, 2018.

[4] Xue Zhigang, Liu Yan, Chai Fahe, et al. Urban air pollution index improvement plan and demonstration[J]. Environmental Science Research, 2011, 24(2): 125-132.

[5] Wang Jian, Hu Xiaomin, Zheng Longxi, et al. Research on air pollution forecasting method based on BP model[J]. Environmental Science Research, 2002, 15(5): 62-64.

[6] Cao Lan. Research on BP Neural Network Forecast of PM10 Concentration in the Air[J]. Environmental Research and Monitoring, 2010, 23(2): 29-32.

[7] Wan Xianlie, Yang Fenglin, Wang Huiqing. Using artificial neural network to predict the O3 concentration in the air[J]. China Environmental Science, 2003, 23(1): 110-112.