Empirical Study on Haze Forecasting Model Based on Neural Network

Zhang Pinghua
Department of computer science and technology, Hefei Technology College
Hefei, Anhui Province, China
zphpinganye@163.com

Abstract: There are many factors affecting the formation of haze, and the relationship between haze and haze is very complex. The prediction model of BP network with high precision fitting is used to predict the haze based on BP neural network. First, establishing an index system that affects the haze factors. Secondly, the neural network haze prediction model is established. Finally, the BP neural network is used to predict the number of days of the year. The results show that the model has high feasibility and rationality, and has obtained accurate and reliable results, which provides valuable reference for improving the air monitoring of relevant departments.

1. Introduction
With the continuous development of China's economy and the quality of life, the problem of environmental quality has also begun to become severe, especially in the haze weather[1-3], people's life is also seriously affected and sometimes brings traffic accidents, physical health or other problems. Take Hefei as an example. In the 2016-2017 year's latest ranking of China's haze cities, Hefei ranks twenty-eighth in the country, with an annual average of 80 micrograms per cubic meter of PM2.5.

2. RESEARCH STATUS
At present, the methods of air quality prediction are: artificial neural network; fuzzy set theory; principal component analysis and other methods[4-6]. A.K.Gupta[7-9], after studying the PM10 in Calcutta residential areas and industrial areas, found that the greatest impact on the area is coal dust and vehicle exhaust. The aerosol particles in the air of Tokyo area were analyzed by Shinichi Okamaoto and so on. The air pollution in this area was mainly caused by the excess of 4 particles in the soil. Georage D and other[10] through the study and analysis of the PM2.5 from the Epacsn sampling, the region is more complex, not only soil particulate pollutants, vehicle emission emissions, coal exhaust pollutants and iron and steel, metallurgy and biomass incineration and other pollutants. K.F.Ho[11] used principal component analysis (PCA), cluster analysis and enrichment factor methods to analyze PM2.5 in Hongkong. The influence degree and contribution value of various sources on different PM2.5 were obtained. In China, Wang Jizhi[12] and other studies analyzed the haze weather under low visibility weather conditions which have been investigated and predicted. The basic methods for analyzing and predicting the low visibility weather of haze are described, and the problems to be paid attention to and the application of the new observation technology in the weather
conditions are discussed. Li Qingchun[13] uses the optimal value difference to analyze the distribution and evolution characteristics of low visibility in Beijing area under the haze weather conditions. The possible causes of visibility are analyzed, and the possible impact of urbanization on atmospheric visibility is analyzed in depth.

3. HAZE PREDICTION MODEL

3.1. Haze prediction index system
The paper takes Hefei as the research object and analyzes the factors affecting the haze weather and the size of the influence degree of each factor and establishes the haze prediction index [14-16] and 10 two grade index system of regional environment, regional economic development, living standard and the 4 first grade index of urban development status.

Based on the analysis of the above situation, the data of the percentage system corresponding to the 10 two level indicators from 2002 -2012 are selected in the experimental simulation as shown in Table 1.

| YEAR | S   | N     | D     | G     | M     | C     | H     | P     | B     | T     |
|------|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2002 | 9.8 | 0.071 | 0.162 | 111.8 | 20.6  | 4207.8| 10860 | 1107.5| 9697.7| 142.24|
| 2003 | 11.5| 0.071 | 0.165 | 111.7 | 20.6  | 4229.2| 14729 | 1122.3| 11262.2| 161.06|
| 2004 | 11.9| 0.076 | 0.166 | 111.5 | 26    | 4330.5| 18172 | 1136.3| 13121.9| 210.52|
| 2005 | 18.3| 0.072 | 0.141 | 111.1 | 40.8  | 4648.2| 25108 | 1148.8| 14096.2| 173.18|
| 2006 | 12.3| 0.071 | 0.149 | 114.1 | 44.7  | 5139.6| 28150 | 1162.9| 14096.2| 198.89|
| 2007 | 14.7| 0.066 | 0.142 | 112.1 | 57    | 5521.9| 31736 | 1180.7| 14146.7| 216    |
| 2008 | 17.6| 0.066 | 0.161 | 113   | 71.4  | 5904.1| 34977 | 1197.6| 14145.3| 209.91|
| 2009 | 15.2| 0.066 | 0.148 | 114.5 | 79.8  | 6285  | 37203 | 1213.3| 14380.6| 235.2  |
| 2010 | 19.1| 0.049 | 0.122 | 109.1 | 87.8  | 6327.1| 42501 | 1229.9| 15572.1| 303.14|
| 2011 | 19.1| 0.053 | 0.121 | 110.2 | 114.8 | 6570.3| 44240 | 1245.8| 18065.2| 319.6  |
| 2012 | 19.2| 0.057 | 0.121 | 110.3 | 143.2 | 6954.1| 46715 | 1257.8| 19765.4| 432.8  |

3.2. The thought of haze prediction model based on Neural Network
The emergence of artificial neural network provides a new method for haze forecasting. The basic idea of BP neural network haze forecast is as follows:

- Normalized values of each evaluation index are used as input vectors of the network.
- Design the BP network structure. Design and determine the number of layers of BP network, number of neurons, preprocessing methods of input and output data, learning rate and conversion function.
- Determine training samples and train the network and judge the training results by recognition rate.
- Make accurate and effective prediction for other samples of the same type except for training samples.

3.3. Preprocessing of data samples
Before using neural network to calculate, we need to normalize data by means of zero mean standard
deviation or normalization. The prediction data of haze use the (3) type of normalized 10 evaluation indexes as input, and the S type transfer function is selected in the hidden layer of neural network, which can improve the convergence speed of network training.

In this paper, normalization of data is processed by Min-Max Scaling method, also known as Min-Max normalization, and the formula of feature quantification as shown in Equation 3.

\[
    z = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}
\]

(1)

Among them, the xi is the input value of the neural network before the conversion; the z is the input value of the transformed neural network; the max(xi) is the minimum input of the neural network before the conversion; it is the maximum input of the neural network before the conversion.

3.4. Model design of haze

- The number of neurons in the input layer
  
  Input layer: 10 indicators are used as input of neurons, that is, the number of nodes in the input layer is n=10.

- The number of neurons in the output layer
  
  Output layer: the result of haze prediction, that is, the number of nodes in the output layer m=4.

- The number of neurons in the hidden layer

  Hidden layer: the accuracy of BP network depends largely on the number of neurons in the hidden layer, but there is no general theory to determine the number of neurons. Therefore, a trial and error method is used to determine the number of them[16]. The empirical formula for calculating the number of hidden layer nodes can be referred to as follows Equation(2) - (4). In this example, (3) formula is used to calculate the number of hidden nodes.

\[
    x = \log_2 n 
\]

(2)

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

(3)

\[
    \sum_{i=0}^{n} C_i^k > k
\]

(4)

Among them, x is the number of hidden layer elements, n is the number of input units, α is a constant between [0,10], k is the sample number, and l is a constant between[0,n].

In practical applications, the number of hidden layers is determined by neural network training. First, the number of nodes in the hidden layer is determined according to the empirical formula, and a BP network with variable number of neurons in the hidden layer is designed to determine the number of the best hidden layer neural element by error comparison. By analyzing, the number of hidden layer nodes selected in the experimental simulation is 3, 5, 7 and 9 respectively. After a large number of experiments, it is proved that the number of nodes in the hidden layer of the haze prediction model BP neural network is the most suitable for this example. The recognition rate of network training is shown in Figure 1, and the convergence diagram is shown in Figure 2.
4. EXPERIMENTAL VERIFICATION AND RESULT ANALYSIS OF FOG AND HAZE PREDICTION MODEL

This case choose the following factors which affecting the haze weather: N02 annual mean, S02 total emission, annual average value of inhalable particulate matter, regional GDP, coal consumption, car ownership, heating supply area of the city, construction area, second industry tax revenue and population density of household registration. With a total of 11 data, of which 8 processed data are used as training samples to train the network, of which 3 are used as test samples to test the network.

The structure of BP network is set up, and the structure is set as: 10-5-4, the hidden layer transfer function selects logsig function, the transfer function of the output layer neuron selects purelin function, the network training adopts the traingdx function.

The maximum learning iteration number is 5000, the initial learning rate is 0.01, the desired target error minimum is 0.01, and the train () function is used to train the neural network with the maximum learning iteration number of 50 times. The diagram of the training process is shown in Figure 3, and the result is shown in Figure 4.

The comprehensive analysis shows that the haze prediction model based on neural network has certain fitting ability for dealing with small sample and high dimension problems. In the model
training process, the year with large fluctuations in the data, although the difference between the predicted value and the actual value is larger, the model is still able to predict more accurate results in energy efficiency. Therefore, the haze forecast model based on neural network has a certain reference significance for relevant departments to analyze and study the unemployment rate of university graduates.

5. CONCLUDING REMARKS
The training sample data of neural network has great influence on the training results of neural network. When selecting data samples, the scientific selection method should be adopted to select the representative training samples. This example is mainly to study and analyze the 10 factors affecting the occurrence of haze. The results are more satisfactory, but the analysis of other factors is lack. In the future, we should focus on the reasons for the formation of PM2.5 and PM10 and the prediction of air quality.

Acknowledgment
Supported By: Natural Science Key Project Foundation of Anhui Province (Grant no.KJ2018A0827, KJ2017A624).

REFERENCES
[1] Zhang Ayuan, Zhang Chong. Media reports on haze events in Beijing and Shanghai under frame construction theory [J]. Journal of Yangtze University (SOCIAL SCIENCE EDITION), 2016,39 (09): 86-88.
[2] Yin Ting Kai. Cause analysis and prevention measures of haze disaster [J]. Economic Research Guide, 2013 (13): 259-260.
[3] Zhou Guangqiang, Chen Min, Peng Li. Haze scientific monitoring and health impact [J]. science, 2013, 65 (4):56-59.
[4] A comparative study on source apportionment models of atmospheric particulates such as [4], Chen Ding.PMF, CMB, FA [D]. Changchun: Jilin University, 2011.
[5] Zheng Quan, Tian Chen. Experience and Inspiration of the haze war in Losangeles, USA [J]. China finance, 2013 (11): 70-71.
[6] Chang Ji Wen. Current situation and problems of legal prevention and control of haze pollution [J]. Chinese lawyer, 2014 (08): 26-28.
[7] SAWANT A A, SONG C, NA K, et al. Characterization of PM2.5 and selected gas-phase compounds at multiple indoor and outdoor sites in Mira Loma, California[J]. Atmospheric Environment, 2004, 38(37):6269-6278.
[8] Pan Benfeng, Wang Wei, Li Liang, and so on. Analysis of the characteristics and causes of haze pollution in the autumn and winter season of China's large and medium cities, [J]. environment and sustainable development, 2013, 38 (1): 33-36.
[9] GUPTA A K, KARAR K, SRIVASTAVA A. Chemical mass balance source apportionment of PM10 and TSP in residential and industrial sites of an urban region of Kolkata, India[J]. Journal of Hazardous Materials, 2007, 142(1):279-287.
[10] THURSTON G D, ITO K, LALL R. A source apportionment of U.S. fine particulate matter air pollution[J]. Atmospheric Environment, 2011, 45(24):3924-3936.
[11] HO K F, CAO J J, LEE S C, et al. Source apportionment of PM2.5 in urban area of Hong Kong.[J]. Journal of Hazardous Materials, 2006, 138(1):73-85.

[12] Wang Jizhi, Zhang Xiaoye, Yang Yuanqin, et al. Identification of the weather condition PLAM index of haze pollution in Sichuan basin and its mechanism analysis on the influence of aerosol concentration change [C]// China Meteorological Society annual meeting S6 atmospheric composition and weather, climate change. 2014:113.

[13] Li youth, Zhang Xiaoling, Li Lin. Analysis of the influence of Beijing urbanization on the spatial and temporal distribution of low visibility under the haze weather conditions [C]// China Meteorological Society annual meeting. Innovation driven development to improve the ability to improve the defense capability of meteorological disasters - the second city meteorological Forum - disaster, environment, influence and response. 2013:83-89.

[14] Optimization of [14] ATLAS K H. neural network and neural network for optimization [D]. Dalian: Dalian University of Technology, 2013.

[15] Gao Pengyi. BP neural network classifier optimization technology [D]. Wuhan: Huazhong University of Science and Technology, 2012.

[16] Zhan Yongping, Fu Chun Li, Duan Xiaofei, Wen Qingzhi, Li Songyan. Based on BP network analysis of the influence factors of the gas well fracturing effect in the Shanxi formation of the Daniudi gas field [J]. Journal of Yangtze University (SCIENCE EDITION), 2017,14 (19): 85-89+120.