ISSN 2616-5872 Vol.4, Issue 3: 53-58, DOI: 10.25236/AJEE.2022.040311

Application of BP Neural Network Based on the Genetic Algorithm in Secondary Modeling of Air Quality Forecast

Xinru Chen*, Ruijie Shen, Shuting Sun

Yunnan University of Finance and Economics, Kunming, Yunnan, 650221, China
**Corresponding author*

Abstract: *It is one of the effective methods improving the ambient air quality to establish air quality forecast models to know the possible air pollution process in advance and take corresponding control measures. The secondary modeling of air quality forecast was explored. By determining the definition and calculation method of air quality index (AQI) and primary pollutants, and according to the degree of influence on pollutant concentration, five meteorological conditions, namely temperature, humidity, air pressure, wind speed and wind direction, were classified by a k-means clustering algorithm in line with AQI. According to the measured data and primary forecast data of three monitoring points (A, B and C), a secondary forecast model was developed using BP neural network based on the genetic algorithm. The results show that the weather favorable for pollutant diffusion had the conditions of high wind speed and high air pressure, while the weather unfavorable for pollutant diffusion had the conditions of low wind speed and low air pressure. The secondary forecast model has good prediction accuracy, and can simultaneously predict the concentrations of multiple monitoring points, many days in the future and various pollutants, thus better forecasting the air quality. The built secondary forecast model can improve the accuracy of air quality forecast, and has high economic and ecological value.*

Keywords: *Air Quality Forecast; Secondary Forecast Model; Air Quality Index; Genetic Algorithm; Back Propagation*

1. Introduction

Air pollution is an increasingly prominent and urgent problem in the world. Monitoring and forecasting air quality, and the data analysis of atmospheric conditions is a very critical link in pollution prevention and control [1]. Exploring air pollution and establishing a reasonable air quality forecast model can provide reasonable basis for scientific decision-making, thus reducing the harm of air pollution to human health and environment and improving air quality. At present, the WRF-CMAQ simulation system (referred to as the WRF-CMAQ model for short) is commonly used to forecast air quality [2]. However, due to the uncertainty of simulated meteorological field and emission list, and the incomplete clarity of the formation mechanism of pollutants including ozone, the results of WRF-CMAQ prediction models are not ideal [3]. Therefore, from the air pollutant concentration, various meteorological conditions, the simulation results of the primary forecast and other conditions, it is a major challenge to conduct secondary modeling forecast to improve the accuracy of forecast [4].

The goal of this study is to classify meteorological conditions reasonably according to the degree of influence on pollutant concentration, and to establish a secondary forecast model which is suitable for multiple monitoring points at the same time to forecast the future air quality.

2. Establishment and solution of models

2.1. Calculation of air quality index and primary pollutants

2.1.1. Definition of air quality index and primary pollutants

The air quality index (AQI) is a new air quality evaluation standard issued by China in March, 2012. It monitors six pollutants, namely CO, SO_2, NO_2, O_3, PM_{10} and $PM_{2.5}$. AQI presents these six pollutants with a unified evaluation standard [5]. Simply put, AQI is an index that can quantitatively describe air quality. It describes the degree of air cleanliness or pollution and its impact on health [6].

ISSN 2616-5872 Vol.4, Issue 3: 53-58, DOI: 10.25236/AJEE.2022.040311

To calculate AQI, the individual air quality index (IAQI) of each pollutant should be obtained first, and the calculation formula of IAQI is as follows:

$$IAQI_P = \frac{IAQI_{Hi} - IAQI_{LO}}{BP_{Hi} - BP_{LO}}\left(C_P - BP_{LO}\right) + IAQI_{LO} \tag{1}$$

Where $IAQI_P$ is the IAQI of pollutant P, C_P is the mass concentration value of pollutant P, BP_{Hi} and BP_{LO} are the high and low values of pollutant concentration limits close to C_P respectively, and $IAQI_{Hi}$ and $IAQI_{LO}$ are the IAQIs corresponding to BP_{Hi} and BP_{LO} respectively.

Table 1 shows the concentration limits and corresponding IAQI levels of various pollutant items.

Table 1: IAQI and concentration limits of corresponding pollutant items

Pollutant items	IAQI and concentration limits of corresponding pollutant items								Units
IAQI	0	50	100	150	200	300	400	500	/
CO, 24-hour average	0	2	4	14	24	36	48	60	mg/m^3
SO_2, 24-hour average	0	50	150	475	800	1600	2100	2620	$\mu g/m^3$
NO_2, 24-hour average	0	40	80	180	280	565	750	940	
O_3, maximum 8-hour moving average	0	100	160	215	265	800	/	/	
PM_{10}, 24-hour average	0	50	150	250	350	420	500	600	
$PM_{2.5}$, 24-hour average	0	35	75	115	150	250	350	500	

Note: (1) When the maximum 8-hour moving average concentration of O_3 is higher than 800 $\mu g/m^3$, the IAQI calculation will no longer be carried out.

(2) When the concentration of other pollutants is higher than the corresponding limit value of IAQI of 500, the IAQI calculation will no longer be carried out.

AQI is the maximum value of each individual index, namely:

$$AQI = \max\left\{IAQI_{SO_2},\ IAQI_{NO_2},\ IAQI_{PM_{10}}, IAQI_{PM_{2.5}},\ IAQI_{O_3},\ IAQI_{CO}\right\} \tag{2}$$

Air quality grade ranges are divided according to AQI values, and the AQI ranges corresponding to the grades are shown in Table 2.

Table 2: Air quality grade and corresponding AQI range

Air quality grade	excellent	good	Light pollution	Moderate pollution	Severe pollution	Serious pollution
AQI range	[0,50]	[51,100]	[101,150]	[151,200]	[201,300]	[301,+∞)

When AQI is less than or equal to 50, it is said that there is no primary pollutant on that day; When AQI is greater than 50, the biggest pollutant of IAQI is the primary pollutant. If IAQI has two or more pollutants, it will be listed as the primary pollutant. Pollutants with IAQI greater than 100 are over-standard pollutants.

2.1.2. Calculation results of AQI and primary pollutants

Given the daily measured data of pollutant concentration in a monitoring point A, according to formula (1) and (2), the daily measured AQI and primary pollutants at any time can be calculated. To take the calculation of the daily measured AQI and primary pollutants on August 25th, 2020 as an example. The calculation process is as follows:

The monitoring concentrations of CO, SO_2, NO_2, O_3, PM_{10} and $PM_{2.5}$ at this monitoring point on that day were 0.5 mg/m^3, 8 $\mu g/m^3$, 12 $\mu g/m^3$, 112 $\mu g/m^3$, 27 $\mu g/m^3$ and 11 $\mu g/m^3$ respectively. According to Table 1, the corresponding concentration limits of each pollutant can be obtained; According to formula (1), the IAQI values of each pollutant can be calculated. It is as shown in Table 3 below.

Table 3: Calculation results of IAQI value of each pollutant

Pollutant items	$IAQI_{Hi}$	$IAQI_{LO}$	BP_{Hi}	BP_{LO}	C_P	$IAQI$
CO	50	0	2	0	0.5	12.5
SO_2	50	0	50	0	8	8.0
NO_2	50	0	40	0	12	15.0
O_3	100	50	160	100	112	60.0
PM_{10}	50	0	50	0	27	27.0
$PM_{2.5}$	50	0	35	0	11	15.7

ISSN 2616-5872 Vol.4, Issue 3: 53-58, DOI: 10.25236/AJEE.2022.040311

Combined with formula (2), the AQI of that day was 60, and the corresponding primary pollutant was O_3. Similarly, it can be calculated that the daily measured AQI of monitoring point A from August 25th to August 28th, 2020 was 60, 46, 109 and 138 respectively, and the primary pollutants were O_3, none, O_3 and O_3 respectively.

2.2. Classification of meteorological conditions

2.2.1. Analysis before classification

A large number of observation facts and statistical analysis show that the changes of meteorological conditions, including atmospheric pressure, relative humidity (high-altitude dew point temperature), temperature, wind speed and wind direction, etc., have obvious laws on the influence of concentration diffusion degree of air pollutants CO, SO_2, NO_2, O_3, PM_{10} and $PM_{2.5}$. Under the condition of constant pollutant discharge, when the meteorological conditions in a certain area are favorable for the diffusion or settlement of pollutants, AQI in this area will decrease, otherwise it will increase. Therefore, it is particularly important to judge which meteorological conditions are conducive to the diffusion of pollutants and classify the meteorological conditions.

Firstly, AQI was calculated according to the daily measured data of pollutant concentration in monitoring point A. Secondly, according to the hourly pollutant concentration of monitoring point A and the meteorological condition values in the meteorological measured data, the 24-hour average values of temperature, humidity, air pressure, wind speed and wind direction were taken as the representative values of the day. Finally, combined with the climatic conditions and AQI of the day, an unsupervised classification method-K-means clustering algorithm was used to classify these five climatic conditions according to AQI values, and the influence of different climatic conditions on AQI was analyzed [7].

2.2.2. Classification results of meteorological conditions

Considering that different meteorological conditions have different influences on AQI, and different meteorological characteristics have certain correlation, this study adopted the method of clustering different climatic characteristics separately, analyzed their influence trends one by one, and finally summarized two climatic types, namely, the climatic conditions conducive to the diffusion of pollutants and the climatic types unfavorable to the diffusion of pollutants.

By clustering analysis of temperature, humidity, air pressure, wind speed and wind direction respectively, the final cluster center was obtained, and AQI was averaged according to their corresponding categories, and the cluster centers of temperature, humidity, air pressure, wind speed and wind direction corresponding to different categories and AQI mean values are shown in Table 4.

Table 4: Cluster centers and AQI mean value of each meteorological condition

Items	Cluster centers			
	1	2	3	4
Temperature	10.21	25.47	17.61	30.87
AQI cluster mean	54.79	70.10	69.12	64.31
Humidity	87.03	26.41	46.41	70.57
AQI cluster mean	44.43	62.32	88.89	66.71
Air pressure	999.01	1005.84	1023.05	1015.54
AQI cluster mean	80.32	59.20	52.28	76.07
Wind speed	3.55	0.89	1.47	2.34
AQI cluster mean	50.21	89.55	59.23	40.58
Wind direction	310.86	223.25	121.55	44.87
AQI cluster mean	64.56	62.64	68.94	62.39

It can be seen from Table 4 that there is no obvious fluctuation trend of AQI with temperature. AQI was concentrated between 54 and 70, all of which belonged to "good" air quality grade. Therefore, the temperature change didn’t play a decisive role in the diffusion of pollutants; As for humidity, AQI tended to increase first and then decrease when the humidity gradually increased. Generally speaking, when the relative humidity was high, it was not conducive to the diffusion of pollutants; As far as air pressure was concerned, AQI had an obvious changing trend with the change of air pressure. When the air pressure was lower around 999 MBar, the AQI was higher, and with the air pressure reaching 1023 MBar, the AQI decreased obviously. Therefore, when the air pressure was high, it was beneficial to the diffusion of pollutants, and then AQI decreased and tended to be good. When the wind speed reached 2.34 m/s and 3.544 m/s, the air quality level was "excellent". There was an obvious trend that when the wind speed was high, AQI decreased obviously, and the air quality level tended to be "good". Therefore, when the wind speed was high, it was beneficial to the diffusion of pollutants. It doesn't seem that different wind

ISSN 2616-5872 Vol.4, Issue 3: 53-58, DOI: 10.25236/AJEE.2022.040311

directions have obvious ascending or descending effect on AQI. Because the wind direction is not the key factor to determine the diffusion of pollutants, the influence of wind direction on air quality depends on the wind speed.

To sum up, according to AQI classification of meteorological conditions, the following conclusions are drawn: when the wind speed is high and the air pressure is high, it is beneficial to the diffusion of pollutants; When the wind speed is low and the air pressure is low, it is not conducive to the diffusion of pollutants.

2.3. A secondary forecast model

2.3.1. Modeling and analysis

Under the conditions of known hourly pollutant concentration, primary meteorological forecast, measured pollutant concentration and meteorological data of each monitoring point [8], the measured data can be used to correct the predicted data, find out the error correction rule, and correct the error based on the prediction results obtained by primary forecast models through the predicted meteorological data in the next three days, so as to construct a secondary forecast model with the minimum maximum relative error of AQI forecast values and the highest prediction accuracy of primary pollutants. To predict the daily concentration values of six conventional pollutants in the next three days, and the corresponding AQI and primary pollutants need to be calculated according to formula (1) and (2).

In this study, a BP neural network model based on the genetic algorithm was applied to predict [9-10]. The traditional BP neural network is a three-layer network with input layer, hidden layer and output layer. Although it is simple and practical, and its convergence speed and execution efficiency are relatively high, its stability is poor. BP neural network based on the genetic algorithm can optimize the weight and threshold of solution, stabilize the structure of neural network and improve the stability of algorithm.

2.3.2. Establishment of a secondary forecast model

BP neural network based on the genetic algorithm includes three parts: BP neural network, the genetic algorithm and BP neural network prediction.

In the BP neural network part, the hourly pollutant concentration of each monitoring point and the meteorological index in the primary forecast were processed as input data, the hidden layer node was set to 5, and the hourly pollutant concentration of each monitoring point and the meteorological measured data were used as output data. The genetic algorithm could find the individual corresponding to the optimal fitness value through selection, crossover and mutation operations, which meets the requirement of "the maximum relative error is as small as possible". The genetic algorithm can optimize the weights and thresholds of BP neural network at the same time, which makes the established model have certain robustness. In the prediction part of BP neural network, the optimal individuals obtained by genetic algorithm were used to assign values to the initial weights and thresholds of the network, and the prediction function outputted the results after the network was trained.

2.3.3. Solution of a secondary forecast model

The BP neural network model based on the genetic algorithm can be used for prediction. Taking the hourly prediction of SO_2 concentration in a monitoring point C as an example, the results are as follows:

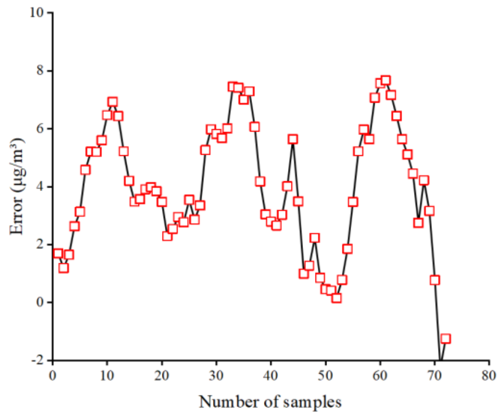

Figure 1: Prediction error of SO2 at monitoring point C

(1) When the BP neural network model based on the genetic algorithm predicted SO_2 concentration, the number of nodes in the hidden layer of the model was 5, and the actual training times were 16. The model basically reached the best fitness before 20 iterations.

(2) Because there was a big difference between the hourly predicted data and the hourly measured data, the established model cannot accurately predict the concentrations of pollutants. However, it can be seen from Figure 1 that error correction could be carried out in the model by using the error law of the model prediction, so that the output values of the secondary prediction could be obtained according to the primary prediction data. The output values obtained at this time were closer to the measured data (Figure 2).

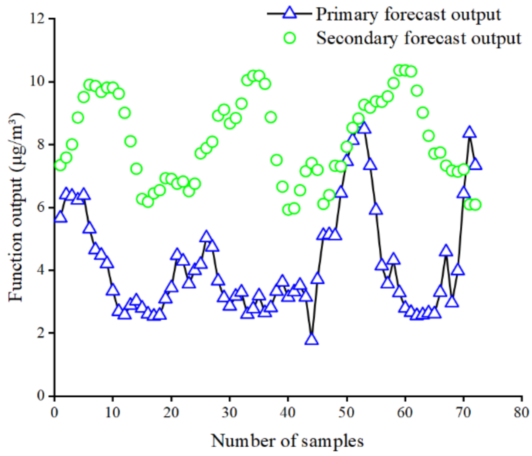

Figure 2: Prediction output of SO2 at monitoring point C

(3) The hourly concentrations of six pollutants in monitoring points A, B and C were predicted using the established model, and the data from July 13, 2021 to July 15, 2021 were treated as the average monitoring concentrations of 24 hours a day, and the AQI and primary pollutants in these three days were obtained by combining formulas (1)-(2) respectively (Table 5).

Table 5: Pollutant concentrations, AQI and primary pollutant results predicted by monitoring points A, B and C

Monitoring points	Forecast date	Daily value predictions of the secondary model							
		SO_2	NO_2	PM_{10}	$PM_{2.5}$	O_3	CO	AQI	primary pollutant
A	2021/7/13	6.09	19.90	33.68	15.46	95.05	0.53	47.53	none
	2021/7/14	5.69	20.00	29.36	12.83	74.16	0.53	37.08	none
	2021/7/15	6.71	23.93	39.75	20.13	134.51	0.56	78.76	O_3
B	2021/7/13	6.68	13.46	26.52	9.31	70.59	0.48	35.29	none
	2021/7/14	6.69	13.84	27.02	9.77	83.45	0.48	41.72	none
	2021/7/15	6.97	15.49	30.98	10.04	91.28	0.50	45.64	none
C	2021/7/13	7.94	25.04	42.49	23.76	142.10	0.57	85.08	O_3
	2021/7/14	8.00	23.78	41.12	22.82	144.88	0.58	87.40	O_3
	2021/7/15	8.50	24.83	47.95	29.26	169.64	0.66	108.76	O_3

3. Conclusion

In this study, the meteorological conditions were classified, and a secondary forecast model was established, which can forecast the future air quality of several monitoring points.

The highlight of the research lies in that the developed secondary forecast model is closely connected with the reality, which can forecast air quality by predicting the concentration of pollutants in the air combined with the measured data, and then give scientific decisions to prevent air pollution. The model has high accuracy, good universality and strong popularization. In addition, compared with the traditional BP neural network, BP neural network based on the genetic algorithm can optimize the solution weights and thresholds at the same time, and has strong nonlinear mapping ability.

However, the research still has such limitations: because there are many and complex realistic factors affecting air quality, and the meteorological conditions considered in the research are relatively simple, there may be some deviation between the predicted results of the established model and the measured results.

ISSN 2616-5872 Vol.4, Issue 3: 53-58, DOI: 10.25236/AJEE.2022.040311

On the whole, the model built by BP neural network based on the genetic algorithm in this study has strong generalization ability, accords with the actual situation and is convenient for popularization and application.

References

[1] Libin Wang. The situation of air pollution control and its existing problems and suggestions [J]. Management & Technology of SME, 2019, (07): 47-48.

[2] Peng Wang. Impacts of emission control strategies on PM2.5 and O3 concentrations over the Yangtze river delta investigated by the two-way coupled WRF-CMAQ model [D]. Zhejiang: Zhejiang University, 2018.

[3] Tingting Ma. Research on WRF-CMAQ-BP neural network air quality forecasting model [J]. Advances in applied mathematics, 2022, 11(2): 641-650.

[4] Yunhui Yang. Research progress of air quality forecast model [J]. Natural Science (Abstract Edition), 2016, 02(06): 123-123.

[5] Qingxian Gao, Junrong Liu, Wentao Li, Wenkang Gao. Comparative analysis and inspiration of air quality index between China and America [J]. Environmental Science, 2015, 36(04): 1141-1147.

[6] Yujie Gu, Yuxiu Zhao, Jian Zhou, Hui Li, Yujie Wang. A fuzzy multiple linear regression model based on meteorological factors for air quality index forecast [J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(06): 10523-10547.

[7] Donghai Zhai, Jiang Yu, Fei Gao, Lei Yu, Feng Ding. K-means text clustering algorithm based on initial cluster centers selection according to maximum distance [J]. Application Research of Computers, 2014, 31(03): 713-715+719.

[8] Xin Bo. Operation guide and case study of air quality model (SMOKE, WRF, CMAQ, etc.) [M]. Beijing: China Environment Publishing Group, 2019.

[9] Haoran Liu, Cuixiang Zhao, Xuan Li, Yanxia Wang, Changjiang Guo. Study on a neural network optimization algorithm based on improved genetic algorithm [J]. Chinese Journal of Scientific Instrument, 2016, 37(07): 1573-1580.

[10] Chunsheng Li, Xiaoye Li, Kejia Zhang. Price forecasting analysis of BP neural network based on improved genetic algorithm [J]. Computer Technology and Development, 2018, 28(08): 144-147+151.