Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2022.030408
ISSN 2790-0983 Vol. 3, Issue 4: 61-68

Artificial Neural Network in the Field of Environment

Vermpaty Charu*
Universiti Teknologi MARA, Malaysia
**corresponding author*

Keywords: Neural Network, Air Quality, Combined Model, Prediction Model

Abstract: The daily monitoring, accurate data analysis, quality prediction and visualization of air quality (AQ) are conducive to the overall control of urban AQ. This paper mainly carries on the research about the artificial neural network in the field of environment. This paper aims at exploring intelligent prediction methods, improving intelligent prediction algorithms and establishing intelligent prediction models for AQ. The combination of time series ARMA model and BP neural network(BPNN) can give full play to their advantages. In this paper, the combined model of ARMA. BPNN is established to predict AQ Index (AQI). The experimental results show that the state information displayed by the simulation is basically consistent with the actual AQ, and the accuracy is more than 85%. This shows the practicability of the model proposed in this paper, which can be used in actual AQ prediction and is conducive to the development of environmental monitoring work.

1. Introduction

Since the 21st century, thanks to the government's vigorous promotion of machinery industrialization and urban modernization, China's economic construction has developed rapidly, and it has become the world's second largest economy. For a long time, China's economic development has consumed a large amount of environmental resources and seriously damaged China's ecosystem. Excessive mechanical industrialization and urban modernization have caused serious damage to the ecological environment, and human beings have been exposed to air pollutants for a long time, which is closely related to the sharp increase in mortality in recent years [1-2]. Air pollution problem has been to residents' health and daily life caused great negative impact, cars, factories and so on the atmosphere of nitrogen oxides, hydrocarbons for children is still growing and the dangers of body function decline of the old man is particularly serious, therefore to improve the ecological environment, to the human a healthy earth is around the corner [3]. AQI is a numerical evaluation index for evaluating AQ. It combines the concentrations of several air pollutants routinely monitored by environmental quality standards and the impacts of various pollutants on human health, ecology and environment, so as to visually reflect the degree of urban air pollution [4]. Traditional AQ prediction methods only use the AQ data of the current monitoring stations, but there is spatial correlation between pollutants, and the AQI of the current monitoring stations is also affected by the AQ data of the neighboring monitoring stations to a certain extent [5]. After the world enters the Internet era, computer technology has rapidly penetrated into all walks of life, especially the development of artificial intelligence in recent years. Machine learning and deep learning technologies have been widely used in People's Daily life, such as computer vision, natural language processing, speech recognition, etc. [6]. In recent years, researchers have gradually applied the technology to AQI prediction.

Mechanism model and data model are two main types of AQ prediction. The mechanism prediction model is mainly based on physical and chemical changes, meteorological factors, and the content of pollutants in the surrounding environment [7]. On the one hand, the collected data such as pollutant discharge are inaccurate and incomplete. On the other hand, the influence of meteorological factors and physical and chemical processes on AQ has obvious nonlinearity. Therefore, the accuracy of the mechanistic model to predict AQ is not high. The era of big data has brought a huge amount of information and provided convenience for the establishment of data models. More and more scholars have conducted more accurate prediction research on AQ based on neural network models [8-9]. Some scholars have used BPNN to predict urban AQ in the western region of North China, and the experimental results show that this method can obtain excellent prediction performance [10]. Some scholars have proved that compared with BPNN, RBF neural network has better effect in AQ prediction and stronger generalization ability of the model [11-12].

The establishment of scientific and efficient AQ prediction model can help people more effectively arrange future outing plans; Secondly, it can help the relevant functional departments more accurately carry out pollution control and environmental protection work; Finally, it can also provide technical support to the country when making policy.

2. Application of Optimized BPNN to AQ Prediction

2.1. Major Air Pollutants and AQ Standards

AQ is closely related to People's Daily life. However, with the development of society and economy, whether it is the uncontrolled emission of industrial waste gas, or the increase of exhaust gas caused by the increase of private cars, the content of pollutants in urban air is constantly rising, which will threaten people's health in the long run. Air pollution is caused by a variety of substances, mainly divided into fine particles and harmful gases. For fine particles, those with aerodynamic equivalent diameter (d) $\leq 2.5\mu m$ are called PM2.5, while those with $d \leq 10\mu m$ are called PM10 [13]. Among them, PM2.5 particles have small diameter, strong adsorption capacity, can combine with many harmful chemical substances, and its small mass is not easy to settle naturally in the air, so it can stay in the atmosphere for a long time and transmit over a long distance, causing greater harm to human body [14].

As for gas pollutants, there are mainly sulfur dioxide (SO2), nitrogen dioxide (NO2), carbon monoxide (CO), ozone (O3) and other volatile organic compounds in the air [15]. Among them, the SO2 come mainly from the fuel combustion process of sulfur (coal, oil, etc.), and some non-ferrous metal ore smelting process, it is a colorless, stimulating odour gas can and PM2.5 combined with formation of aerosols in the air, is a threat to human respiratory tract, serious when can make people breathing difficulties, cause a variety of respiratory diseases and even deadly. In addition, SO2 can also combine with water vapor in the atmosphere and dissolve in water to form sulfuric acid and sulfite, leading to the formation of acid rain [16]. NO2 is the main form of nitrogen oxide pollutants in the air, which is generated by the combustion of fossil fuels, the use of nitric acid, the manufacture of nitrogen-containing fertilizers, metal smelting and other processes. Nitrogen oxides in the air will cause irritation to human lungs, induce asthma, and make it difficult for people to resist colds and other diseases. At the same time, nitrogen oxide is also one of the main components of photochemical smog. Similar to SO2, NO2 can combine with water vapor to form nitric acid rain [17]. CO is the product of incomplete combustion of carbon containing substances such as coal and oil. With the continuous development of manufacturing industry and automobile industry, both industrial oil consumption and civil gasoline and kerosene use are on the rise, resulting in the increase of CO content in the air year by year. Because CO is not easy to react with other substances in the air, it can stay in the atmosphere for 2-3 years or even longer, and because of its strong binding ability with red blood cells, if the concentration exceeds the standard, it will lead to hypoxia. O3 is produced by nitrogen oxides in the air through ultraviolet irradiation. Excessive concentration of O3 will lead to the decline of human lung function and permanent damage to plants [18].

AQ is determined by the average amount of pollutants in the air. At present, China's AQ situation is described by two dimensionless indices, AQI and AQ Sub-index (IAQI). Among them, AQI is represented by the maximum IAQI of six pollutants, PM2.5, PM10, SO2, NO2, CO and O3, and the IAQI of each pollutant is calculated by its mass concentration value.

When the AQI in the current environment is greater than 50, the air pollutant with the largest IAQI is the primary pollutant. As shown in Table 1, according to the AQI value range calculated in the above way, the AQ in the current environment can be divided into six levels.

Table 1. AQI classification

AQI range	AQI level	AQI category
0-50	Level 1	Optimal
51-100	Level 2	Good
101-150	Level 3	Light pollution
151-200	Level 4	Moderate pollution
201-300	Level 5	Heavy pollution
>300	Level 6	Serious pollution

Within the different ambient AQ levels, the national environmental protection standards give the warning color of the corresponding level, the possible harm to human health, and the recommended countermeasures for people to take. Poor AQ will not only bring a strong negative impact on people's health and work life, but also hinder our current social and economic development.

2.2. Construction of Neural Network Prediction Model

The number of hidden layers will directly affect the prediction effect and training duration of the network. Although increasing the number of hidden layers can improve the prediction accuracy, too many hidden layers will lead to the increase of model complexity. Therefore, under the premise of meeting the prediction accuracy, the number of hidden layers should be reduced as much as possible. At present, the network model with one or two hidden layers is commonly used in most studies for training. In this paper, BPNN with one hidden layer is constructed to achieve AQ prediction.

It is very important to select an appropriate number of nodes in the network input layer. Too high input dimension will increase the complexity of network training, thus increasing the training duration. Therefore, samples with a high degree of relevance to the problem to be solved should be selected as network input. AQI may be affected by many factors, such as spatial factors, meteorological conditions and human factors at that time. However, the BPNN model constructed in this paper is the analysis and prediction of AQI time series and the short-term prediction of AQI. Therefore, in the process of AQ prediction in this paper, the historical AQI series is taken as the original data. After several experiments, it is concluded that the different number of input nodes leads to the different training regression parameters of BP network. When the number of input nodes is 2, the training regression parameter of BP network is 0.875. When the number of input nodes is 4, the regression parameter of network training is 0.996. When the number of input nodes is 5, the regression parameter of network training is 0.998, which is the closest to 1. Therefore, the input layer node of BPNN in this paper is determined as 5.

The AQ prediction model in this paper is a short-term prediction of AQI. The AQI prediction model belongs to a nonlinear mapping with multiple inputs and single output, so the number of nodes in the output layer of the network can be determined as l. The selection of the number of hidden layer nodes is usually determined by testing the number of nodes in a certain range. The calculation method of its approximate value range is as follows:

$$p = \sqrt{n+q} + \alpha \tag{1}$$

Where, n and q respectively represent the number of nodes in the input and output layers, and α is taken as any real number [1,10]. According to calculation, the approximate range of the number of nodes in the hidden layer is [6.12]. After repeated experiments of gradually increasing and deleting hidden layer nodes, it is found that the number of hidden layer neurons corresponding to the minimum relative error and mean square error of the network can be determined as 12 by comparing the errors predicted by BP network with different number of hidden layer nodes.

Activation function is an important factor that affects the training performance of network models, and it is the main factor that makes networks possess nonlinear processing ability. According to the characteristics of the studied data and the characteristics of each activation function.

3. Neural Network Model Optimization

In this paper, ARMA time series model and BPNN model are combined to make short-term prediction of AQI. Because ARMA model has good linear fitting effect on time series, while BPNN is good at dealing with nonlinear mapping relationship, they have different fitting ability on the characteristics of AQI data. The combination of ARMA and BPNN can fully extract the hidden information in AQI sequence, and effectively improve the prediction accuracy and practicability of the model.

Time series Y is regarded as composed of linear structure l and nonlinear structure n. The specific algorithm of BP-ARMA combined model for time series prediction is as follows:

The ARMA model is established to predict the AQI time series, that is, the linear feature l existing in the series is fitted, and the prediction result is represented by lt. The difference between the predicted value and the observed value is calculated to form the residual series, which contains the nonlinear features existing in the historical AQI data and can be expressed as follows:

$$e_t = y_t - l_t \tag{2}$$

The residual et contains the nonlinear relationship in the AQI sequence. In order to fully explore the hidden information in the observation sequence, the powerful nonlinear fitting ability of BPNN is further used to learn the nonlinear relationship, and the residual is predicted, and the predicted value is denoted as nt. RMA model and BPNN model give full play to their respective advantages.

4. Analysis of BPNN Model Test Results

4.1. Model Testing

Using the neural network toolbox in MATLAB, the improved BPNN algorithm is used for prediction. The training set is the historical data from January to February 2021, and the predicted result AQI* is compared with the actual value.

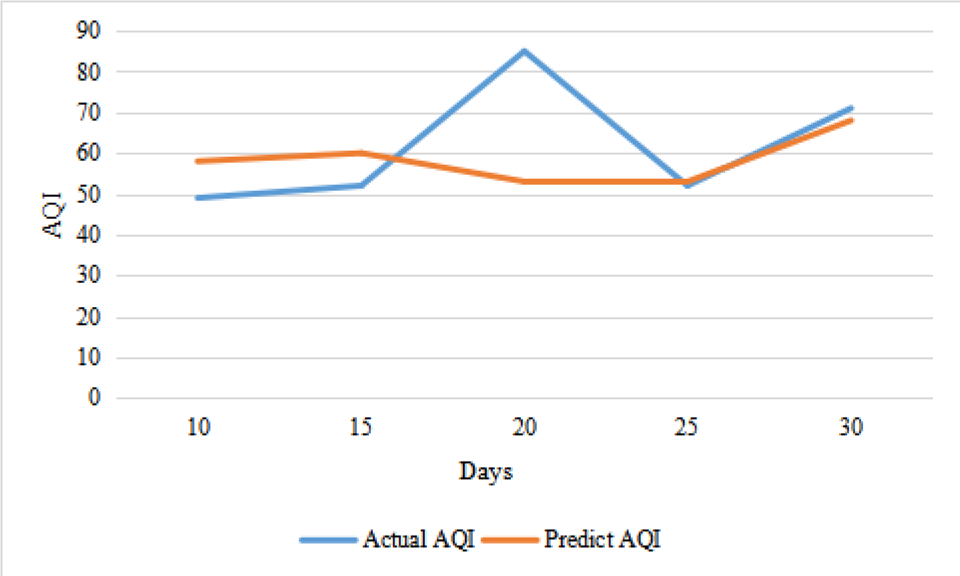

Figure 1. Prediction based on optimized BPNN

As shown in Figure 1, the overall error of the proposed model is significantly improved compared with the traditional BPNN prediction when the number of iterations is less. The generalization ability of the model has been greatly improved, which obviously has stronger practical application ability.

Table 2. Comparison of neural network training parameters before and after improvement

Neural network	Training time(s)	Training error	Accuracy of prediction
BPNN	19.02	2.32e-3	72.5%
Improved BPNN	12.78	3.16e-4	85.2%

As shown in Table 2, the training time of the improved neural network is 13.22 seconds, and the final prediction result accuracy is 87.5%.

4.2. Near-term AQ Forecast

The basic idea of this section is as follows: Firstly, the known historical data in the recent period are used for training and testing, so as to preliminarily judge the application ability of the model in the recent air climate model. Secondly, the data in the recent period are used as input to predict the future air pollution indicators.

In order to maintain the continuity in time, a total of 30 days of data in January 2022 were acquired using the same data crawling method for model training and testing.

Similar to the training and testing of the model mentioned above, the prediction results are shown in Figure 2.

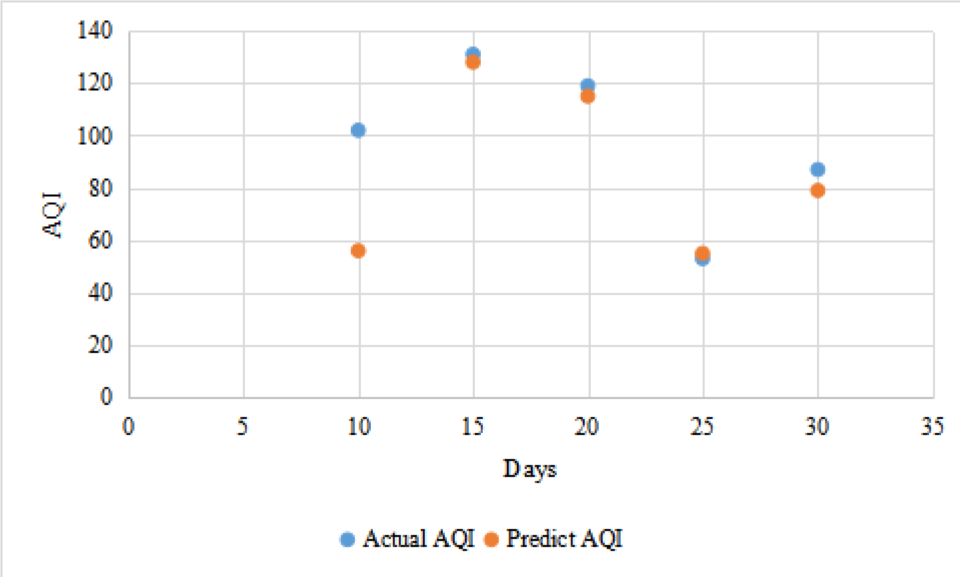

Figure 2. AQI forecasts

The prediction results verified our conjecture and were highly consistent with the actual data of March, which indeed continued the trend of improving AQ and decreasing AQI in February.

Similarly, in order to verify the effectiveness and superiority of this model training, a comparative experiment between ARIMA model and traditional BPNN model is conducted in this paper. By inputting the same historical data into ARIMA model and traditional BPNN model for training, the test results are shown in Table 3 for the accuracy of different models.

Table 3. Comparison of neural network training parameters before and after improvement

Neural network	Training time(s)	Training error	Accuracy of prediction
BPNN	14.37	2.45e-3	72.1%
Improved BPNN	12.59	2.47e-5	92.8%
ARIMA	14.18	7.21e-4	81.9%

It can be seen that the accuracy of the traditional BPNN algorithm is the lowest, only 72.1%, ARIMA is 81.9%, and the accuracy of the improved neural network is 92.8%.

5. Conclusion

With the rapid development of urban industry, air pollution is becoming more and more serious, which directly affects the health of residents. In this paper, the core part of artificial intelligence -- neural network is introduced into the AQ prediction work, which provides a preliminary basis for the subsequent intelligent AQ monitoring. In this paper, the combination of ARMA model and BP model is proposed for modeling research. ARMA model is a typical method to study linear time series. The combined model builds ARMA model to fit the linear part of AQI, and then uses BPNN to fit the nonlinear relationship implied in the residual series. Finally, the prediction result of the combined model is obtained by adding the fitting results of the two models. By comparing and analyzing the prediction results of ARMA model, BP model and combination model, the prediction results of combination model are more accurate.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Iskandaryan D, Ramos F, Trilles S. AQ prediction in smart cities using machine learning technologies based on sensor data: a review. Applied Sciences, 2020, 10(7): 2401. https://doi.org/10.3390/app10072401

[2] Wardana I N K, Gardner J W, Fahmy S A. Optimising deep learning at the edge for accurate hourly AQ prediction. Sensors, 2021, 21(4): 1064. https://doi.org/10.3390/s21041064

[3] Saini J, Dutta M, Marques G. Indoor AQ prediction systems for smart environments: A systematic review. Journal of Ambient Intelligence and Smart Environments, 2020, 12(5): 433-453. https://doi.org/10.3233/AIS-200574

[4] Benhaddi M, Ouarzazi J. Multivariate time series forecasting with dilated residual convolutional neural networks for urban AQ prediction. Arabian Journal for Science and Engineering, 2021, 46(4): 3423-3442. https://doi.org/10.1007/s13369-020-05109-x

[5] Lee K, Yu J, Lee S, et al. Development of Korean AQ Prediction System version 1 (KAQPS v1) with focuses on practical issues. Geoscientific Model Development, 2020, 13(3): 1055-1073. https://doi.org/10.5194/gmd-13-1055-2020

[6] Khan A W, Kumar P. Impact of chemical initial and lateral boundary conditions on AQ prediction. Advances in Space Research, 2019, 64(6): 1331-1342. https://doi.org/10.1016/j.asr.2019.06.028

[7] Gryech I, Ghogho M, Elhammouti H, et al. Machine learning for AQ prediction using meteorological and traffic related features. Journal of Ambient Intelligence and Smart Environments, 2020, 12(5): 379-391. https://doi.org/10.3233/AIS-200572

[8] Brasseur G P, Xie Y, Petersen A K, et al. Ensemble forecasts of AQ in eastern China-Part 1: Model description and implementation of the MarcoPolo-Panda prediction system, version 1. Geoscientific Model Development, 2019, 12(1): 33-67. https://doi.org/10.5194/gmd-12-33-2019

[9] Maleki H, Sorooshian A, Goudarzi G, et al. Air pollution prediction by using an artificial neural network model. Clean Technologies and Environmental Policy, 2019, 21(6): 1341-1352. https://doi.org/10.1007/s10098-019-01709-w

[10] Hansun S, Wicaksana A, Kristanda M B. Prediction of Jakarta City AQI: modified double exponential smoothing approaches. Int J Innov Comput Inf Control, 2021, 17(4): 1363-71.

[11] Olga L, Vladimir L, Leonidas N. Estimation and prediction of the effect of alternative engine technologies and policy measures on the AQ in St. Petersburg in 2010-2030. Architecture and Engineering, 2018, 3(4): 31-35. https://doi.org/10.23968/2500-0055-2018-3-4-31-35

[12] Lepioufle J M, Marsteen L, Johnsrud M. Error Prediction of AQ at Monitoring Stations Using Random Forest in a Total Error Framework. Sensors, 2021, 21(6): 2160. https://doi.org/10.3390/s21062160

[13] Sanjeev D. Implementation of machine learning algorithms for analysis and prediction of AQ. International Journal of Engineering Research & Technology (IJERT), 2021, 10(3): 533-538.
[14] Masih A. Machine learning algorithms in AQ modeling. Global Journal of Environmental Science and Management, 2019, 5(4): 515-534.
[15] Ahmed M M, Hoque M, Rahman S, et al. Prediction of COVID-19 Cases from the Nexus of AQ and Meteorological Phenomena: Bangladesh Perspective. Earth systems and environment, 2022, 6(1): 307-325. https://doi.org/10.1007/s41748-021-00278-7
[16] Sethi J K, Mittal M. Ambient AQ estimation using supervised learning techniques. EAI Endorsed Transactions on Scalable Information Systems, 2019, 6(22): e8-e8. https://doi.org/10.4108/eai.29-7-2019.159628
[17] Freeman B S, Taylor G, Gharabaghi B, et al. Forecasting AQ time series using deep learning. Journal of the Air & Waste Management Association, 2018, 68(8): 866-886. https://doi.org/10.1080/10962247.2018.1459956
[18] Zayed R, Abbod M. Big Data AI System for AQ Prediction. Data Science and Applications, 2021, 4(2): 5-10.