Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 5, Issue 10: 101-105, DOI: 10.25236/AJCIS.2022.051016

Air Quality Index Prediction Based on Multilayer Perceptron

Li Xiaogang[1], Lu Xin[1,2]*
[1.] *Xi'an highway research institute Co.,Ltd，Shaanxi Xian, China*
[2.] *Chang'an University, Shaanxi Xian, China*
[a] *66480737@qq.com* , [b] *405491870@qq.com*
**corresponding author*

Abstract: *Taking the comprehensive evaluation index AQI as the research goal, the data of Beijing is studied, and an AQI prediction model based on multilayer perceptron neural network (MLP neural network) is proposed. First, the Beijing air quality dataset is constructed, and the AQIs from 2014 to 2019 are extracted as input features. Then, the air quality index of Beijing in 2020 is predicted based on MLP. After that, through parameter adjustment and model comparison, the final prediction accuracy reached 0.861.*

Keywords: *Environmental protection; Air quality index; Multilayer perceptron; Neural network.*

1. Introduction

In recent years, with the rapid development of industrial production, air pollution has become more and more serious, which has had a great impact on the environment and human production and life [1]. AQI, a dimensionless index used to describe air quality conditions, is calculated from the concentrations of various air pollutants [2]. Air quality has received more and more attention from experts and scholars at home and abroad. An accurate and reliable AQI prediction model is of great significance to the control of air pollution [3].

Experts and scholars at home and abroad have carried out a lot of research [4-6]. Zhang L et al. compared the concentrations of PM2.5 with other pollutants, fused meteorological parameters, and used an autoregressive integrated moving average (ARIMA) model to predict PM2.5 concentrations. The results show that the fluctuation trend of PM2.5 concentration has obvious seasonality, which verifies the accuracy and applicability of the model [7]. Jeong J et al. took into account seasonal changes and other factors, and used a simple fitting method to predict the contributing factors of aerosol concentration at Korean observation sites. The model prediction results have good accuracy [8]. Li et al. used LSTM to automatically extract the effective features inherent in the historical data of air pollutants, and fused auxiliary data such as meteorological data and timestamp data into the model to improve the performance. The experimental results show that the method has higher accuracy in the prediction of atmospheric pollutant concentration [9]. Ji Chunlei et al. used the improved whale optimization algorithm to predict the AQI of 4 cities in China, which effectively improved the AQI prediction accuracy. It is of great significance to the sustainable urban development [10]. Yan R et al. combined CNN and LSTM deep spatiotemporal model to predict the AQI in Beijing, and the prediction accuracy is more generalized [11]. Skeher et al. constructed a new grey prediction multivariate model FHGM (1,N) to predict the air quality in Shijiazhuang. The experimental results show that the model has good prediction performance and generalization [12].

According to the above analysis of relevant papers in this field, this study takes the comprehensive evaluation index AQI as the research goal, and conducts research on the data of Beijing. First, this study constructs the Beijing air quality data set, and extracts the AQI from 2014 to 2019 as the input feature. Then, based on Multi-layer perceptron neural networks (MLP neural netwoks), this study predicts the AQI of Beijing in 2020. After that, by adjusting the parameters and comparing the model, the optimal prediction result is finally obtained.

ISSN 2616-5775 Vol. 5, Issue 10: 101-105, DOI: 10.25236/AJCIS.2022.051016

2. Marterial

2.1. Data sources

The experimental data selected Beijing's air quality data from 2014 to 2020, which came from the China Air Quality Online Monitoring and Analysis Platform (https://www.aqistudy.cn). Take the average of the monthly AQI data from 2014 to 2020, conduct visual statistics, and observe the monthly change trend of AQI. The lower the AQI index, the better the air quality, and the higher the index, the worse the air quality. As shown in Figure 1, it can be found that the AQI did not change much in the first half of the year. In June, the AQI reached its highest level of the year. Then the AQI began to decrease month by month. In September, the AQI reached the lowest value of the year, and then showed an upward trend.

The input characteristic index of this experimental model is the daily average AQI index of Beijing from 2014 to 2019, and the output index is the daily average AQI index of 2020. A total of 6*365+366=2556 pieces of data.

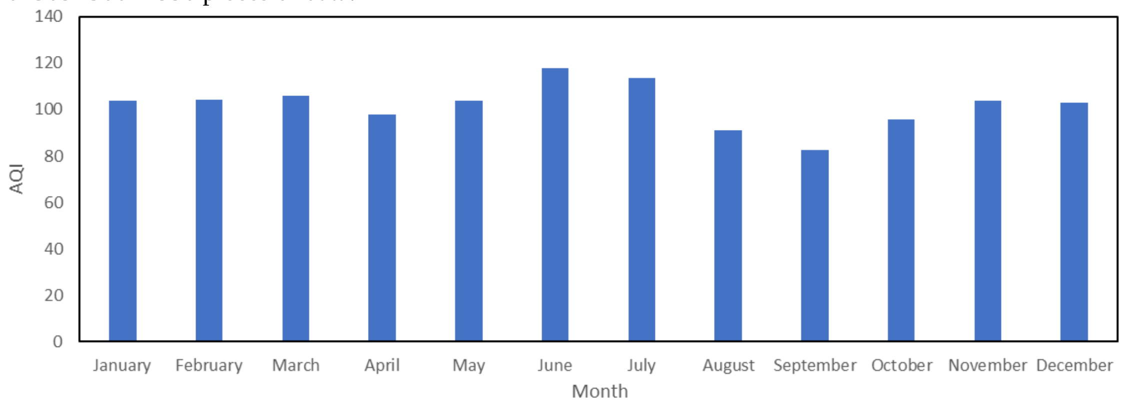

Figure 1: AQI monthly average change in 2014-2020

2.2. Analysis of air quality levels

The air quality in Beijing is classified according to the AQI value, and various air quality grade maps are drawn every year. As shown in Figure 2, the darker the color in the figure, the heavier the air pollution. It can be seen that the distribution of various air quality grades in each year is in a situation of steady fluctuation, the area of the light-colored (good) part gradually increases, and the area of the dark-colored (polluted) part gradually decreases. In order to reflect the changes in air quality levels more intuitively, the excellent rate and pollution rate over the years were calculated respectively, as shown in Figure 3.

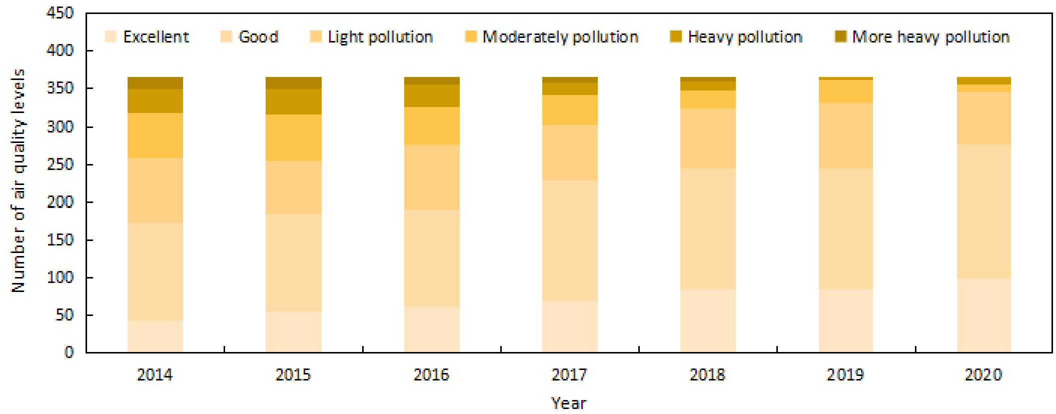

Figure 2: Distribution of air quality grades in Beijing from 2014 to 2020

ISSN 2616-5775 Vol. 5, Issue 10: 101-105, DOI: 10.25236/AJCIS.2022.051016

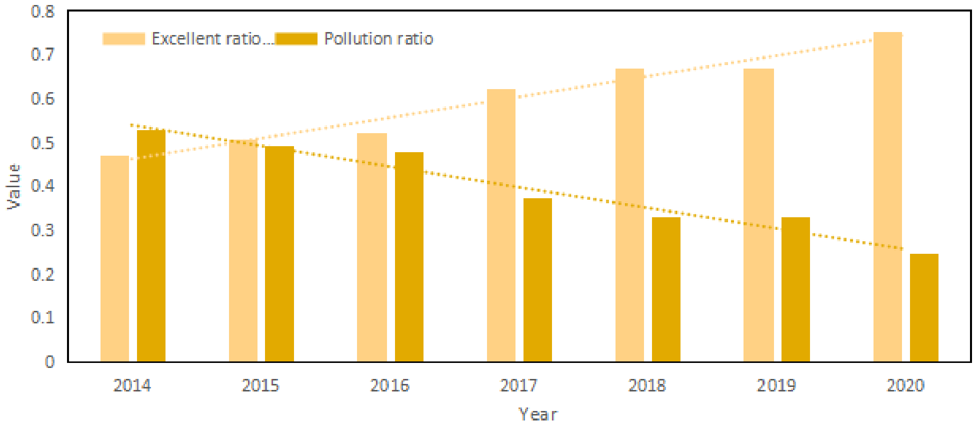

Figure 3: Changes in the excellent rate and pollution rate in Beijing from 2014 to 2020

3. Air quality prediction based on Multi-layer perceptron neural networks

3.1 MLP network parameter settings

MLP is based on reverse artificial neural network, which is a network with simple structure and strong plasticity [13]. The network structure consists of an input layer, a hidden layer and an output layer, each layer contains multiple nodes, and each layer node is fully connected to the next layer of the network. The most prominent feature of the MLP neural network is that the weights and thresholds are repeatedly corrected through the back-propagation of the error, so that the error function value can be minimized and the accuracy can reach the expected standard. In the process of backpropagation, the most commonly used error minimization method is the gradient descent direction, which adjusts the network parameters. In the network training process, the error is gradually reduced after each backpropagation iteration through the setting of the learning rate, and finally reaches the acceptable range of the system, that is, the optimal weight.

The parameters in the experiment are set as follows: the experimental training set and test set data are divided into 0.75:0.25. After more tests, it is found that when two hidden layers are set (the number of nodes in the first hidden layer is 21, and the number of nodes in the second hidden layer is 9), the accuracy is the highest. The final network structure diagram is shown in Figure 4.

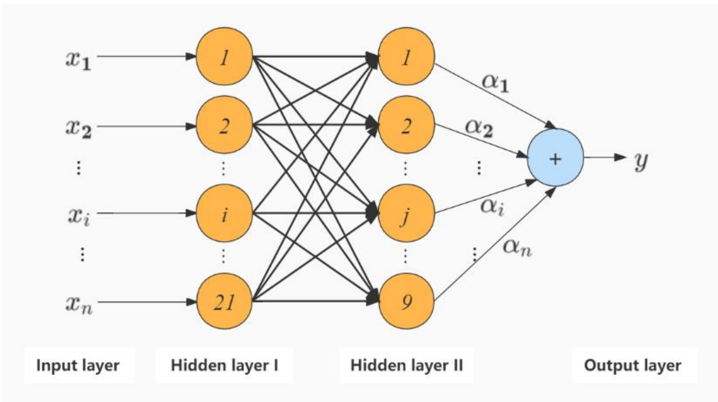

Figure 4: Network structure setting diagram

3.2 Result analysis

R-Square is used as the evaluation index of model prediction accuracy. Optimize the model by setting different number of network nodes. Taking 21 network nodes in the first hidden layer as an example, the parameter adjustment process of the network nodes in the second hidden layer is visualized, and the accuracy is shown in Figure 5. As can be seen from the figure, when the first hidden layer is 21 nodes and the second hidden layer is 9 nodes, the prediction accuracy is the highest, reaching 0.861.

ISSN 2616-5775 Vol. 5, Issue 10: 101-105, DOI: 10.25236/AJCIS.2022.051016

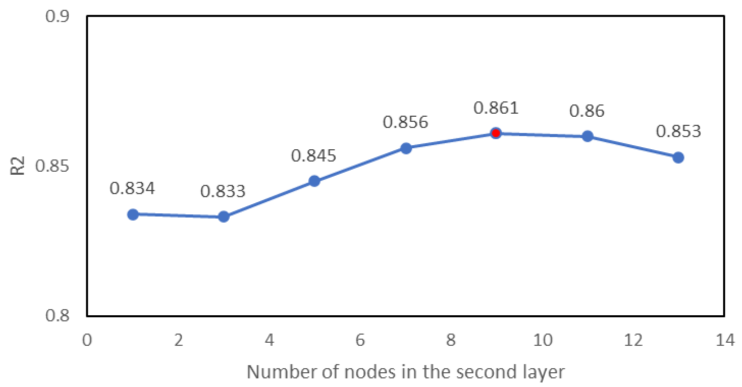

Figure 5: Model accuracy for different number of nodes of the second hidden layer

By comparing the actual and predicted values of Beijing AQI in 2020, a comparison curve is drawn, as shown in Figure 6.

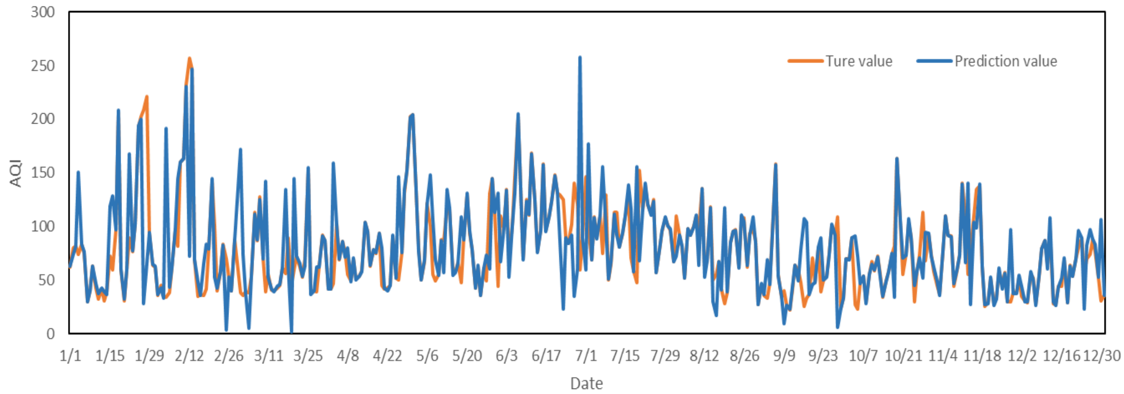

Figure 6: Comparison of Beijing AQI forecasting effects in 2020

4. Conclusions

Based on the air quality index prediction model of MLP, this study sets hidden layer nodes and parameters, and the prediction accuracy of AQI in Beijing in 2020 reaches 0.861, which can provide certain data support for air quality control.

References

[1] Liu H, Yin S, Chen C, et al. (2020) Data multi-scale decomposition strategies for air pollution forecasting: A comprehensive review [J]. Journal of Cleaner Production, 277:124023.
[2] Liu H, Yan G, Duan Z, et al. (2021) Intelligent modeling strategies for forecasting air quality time series: A review [J]. Applied Soft Computing, 102:106957.
[3] Gu Y, Zhao Y, Zhou J, et al. (2021) A fuzzy multiple linear regression model based on meteorological factors for air quality index forecast. Journal of Intelligent & Fuzzy Systems,6, 201222.
[4] Cong Zhao, Xuemei Li. (2019) Research on Air Quality Index Prediction Based on Neural Network Taking Beijing as an Example[C]. Proceedings of 5th International Conference on Economics,Management,Law and Education(EMLE 2019)(Advances in Economics, Business and Management Research, 110: 682-688.
[5] Huixiang Liu, Qing Li, Dongbing Yu, et al. (2019) Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms [J]. Applied Sciences, 19: 9194069.
[6] Suman. (2020) Air quality indices: A review of methods to interpret air quality status [J]. Materials Today: Proceedings (prepublish), 07: 141.
[7] Zhang L, Lin J, Qiu R, et al. (2018) Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model [J]. Ecological Indicators, 95:702-710.
[8] Jeong J I, Park R, Woo J H, et al. (2011) Source contributions to carbonaceous aerosol concentrations in Korea [J]. Atmospheric Environment, 45(5):1116-1125.

ISSN 2616-5775 Vol. 5, Issue 10: 101-105, DOI: 10.25236/AJCIS.2022.051016

[9] Li X, Peng L, Yao X, et al. (2017) Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation [J]. Environmental Pollution, 231:997–1004

[10] Ji C, Zhang C, Hua L, et al. (2022).A multi-scale evolutionary deep learning model based on CEEMDAN, improved whale optimization algorithm, regularized extreme learning machine and LSTM for AQI prediction.. Environmental research, Pt 1, 114228.

[11] Yan R, Liao J, Yang J, et al. (2020) Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering[J]. Expert Systems with Applications, 169(4):114513.

[12] Shi K, Ding R, Wu L, et al. (2023) Construction of a New Grey System Multivariate Model for Predicting Air Quality--A Case Study of Shijiazhuang City [J/OL]. Journal of Systems Science, 2, 75-81.

[13] PEI L, SUN Z, HU Y, et al. (2020) Neural Network Model for Road Aggregate Size Calculation Based on Multiple Features [J]. Journal of South China University of Technology (Natural Science Edition), 48(6): 77-86.