Universe Scientific Publishing

DOI: 10.18686/ahe.v6i9.4726

Air Quality Prediction Based on Quadratic Prediction Model

Zijun Luo, Rui Zeng*Corresponding Author, Pan Wang

College of Urban Railway Transportation,Shanghai University of Engineering Science,Shanghai,201620,China

Abstract: In order to improve the performance of the prediction model of air quality prediction,a secondary prediction mathematical model is established in this paper.The first is to clean the data and find the potential model relationship between variables through data mining and correlation methods,so as to establish the limit learning machine model.The model needs to be able to explain the influence of meteorological index variables on pollutant concentration diffusion to a certain extent.Then,the EML model is optimized by genetic algorithm,rolling optimization and other methods to reduce noise and make the data as accurate as possible.
Keywords:Secondary Prediction Model;Genetic Algorithm;Air Quality Forecast

1. Introduction

In recent years,with the rapid increase of population and the excessive consumption of resources,environmental problems are becoming more and more serious,especially air pollution has become one of the most important environmental problems in the world. The practice of pollution prevention and control shows that by establishing the air quality prediction model,we can know the possible air pollution process in advance and take corresponding control measures to reduce the harm of air pollution to human health and the environment and improve the ambient air quality.At present,WRF-CMAQ simulation system is commonly used to predict air quality. However,the model is subject to the uncertainty of the simulated meteorological field and emission inventory,and the prediction results of the model are quite different from the actual air quality.Therefore,it is a key research problem to improve the quality of air quality prediction by combining the prediction data of WRF-CMAQ model with the measured data of actual observation points.

2. Model Establishment and Solution

2.1 Problem Analysis

In this paper,a quadratic prediction model which can be applied to three monitoring points A,B and C at the same time is established. According to the given data,the data of monitoring points A,B and C need to be preprocessed(data elimination and data interpolation). The input data of the secondary prediction model are primary prediction data and monitoring data,and the data from July 23,2020 to July 13,2021 are selected as the input data.Based on the limit learning machine model,the quadratic prediction mathematical model is constructed.The data is divided into training set and test set.The training set data is used to train the quadratic prediction model;The test set data is used to verify the applicability of the model.The fitness function is constructed to predict the effect of extreme learning machine.It is required that the maximum relative error of AQI prediction value in the prediction results of the secondary prediction model should be as small as possible,and the prediction accuracy of primary pollutants should be as high as possible.Therefore,we use the relative error of AQI of the prediction results of the secondary prediction model and the accuracy of primary pollutants as the fitness function.Genetic algorithm is used to optimize the quadratic prediction model,so that the fitness gradually decreases and tends to be stable.Finally,the optimal index parameters are obtained,and the optimized quadratic prediction model can be obtained.Based on the optimized secondary prediction model,the measured data at 7:00 on July 13,2021 and the predicted data at 8:00 on that day are selected as the input data,and the secondary prediction data at 8:00 on July 13,2021 are predicted by using the rolling learning method. By analogy,the single day concentration values of 6 conventional pollutants from July 13 to July 15,2021 can be solved,and then the corresponding AQI and primary pollutants can be calculated.

2.2 Model Establishment and Optimization

2.2.1 Establishment of quadratic prediction model based on limit learning machine

In the limit learning machine(ELM)algorithm,the connection weight between the input layer and the hidden layer and the threshold of neurons in the hidden layer are randomly generated,and there is no need to adjust in the training process.The only optimal solution can be obtained by setting the number of neurons in the hidden layer.Through comprehensive comparison,it is known that the limit learning machine has the following advantages:

(1) Limit learning machine(ELM)is a simple,fast and effective learning algorithm of feedforward neural network.Compared with the traditional learning algorithm based on gradient descent,limit learning machine has great advantages

(2) The calculation speed of elm is very fast.It randomly gives the connection weight of hidden layer,and the training process does not need iterative adjustment

(3) The traditional gradient descent algorithm is easy to fall into local minima,while the elm algorithm will not fall into local optima because its process of solving the least square solution of output weight is a convex optimization problem,so it has better

generalization than the traditional algorithm.The monitoring data at time t and the primary prediction data at time t+1 are input as input parameters into the secondary prediction model based on limit learning machine,and the secondary prediction results at time t+1 are obtained.

2.2.2 Optimization of quadratic prediction model based on genetic algorithm

It is required that the maximum relative error of AQI prediction value in the prediction results of the secondary prediction model should be as small as possible,and the prediction accuracy of primary pollutants should be as high as possible.Therefore,genetic algorithm is used to optimize the secondary prediction model.

For monitoring point A,354 hours of data from 0:00 on July 23,2020 to 24:00 on July 13,2021 are divided into 200 training sets and 154 test sets.A fitness function composed of the maximum relative error of AQI and the average relative error of primary pollutants is constructed,which is embedded in the quadratic prediction model optimized by genetic algorithm.A set of optimization parameters are randomly set,and the training set and randomly set parameters are used to train the model.When the fitness curve function value optimized by genetic algorithm gradually decreases and finally tends to be stable,the optimal parameters are obtained,The model achieves the best effect.The optimal parameters obtained from the test set and training are used as inputs to obtain the prediction results.The AQI is calculated using the prepared AQI calculation program and the primary pollutants are obtained.However,the monitoring data in the test set can only be used for subsequent prediction at 7:00 on July 23,2020(the closer the measured data is,the better the secondary prediction effect is,so the measured data of the previous hour,that is,7:00,is best used for the prediction of 8:00). It is necessary to use the rolling prediction method to obtain the secondary prediction data of 8:00 at the next time,and then use the secondary prediction data of 8:00 and the primary prediction data of 9:00 as the data source of 9:00 secondary data prediction.

The calculation of monitoring points B and C is the same as above.Through calculation,the single day concentration values of six pollutants,corresponding air quality index AQI and primary pollutants at monitoring points A,B and C from July 13 to July 15,2021 can be obtained.

Among them,after completing the steps of data analysis,preprocessing and feature dimensionality reduction,model training needs to be carried out repeatedly.Model training involves training set and test set.Training set is used to train and adjust model parameters;test set is used to verify the generalization ability of model.The following figure shows the training results of three monitoring points:

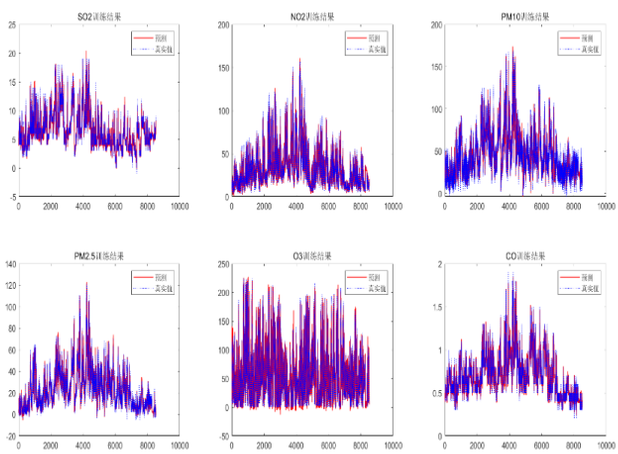

Figure 1 Training results of monitoring point A

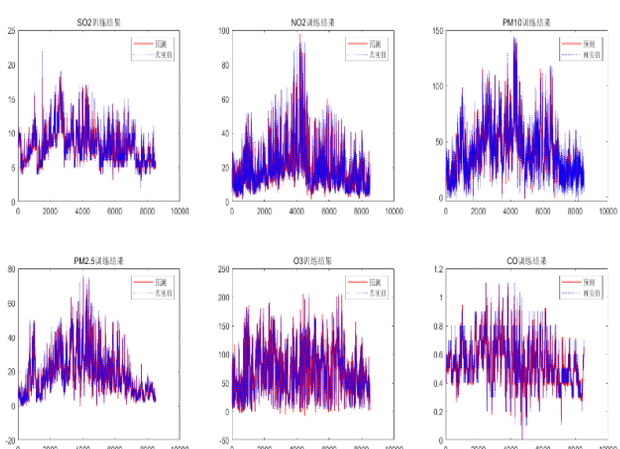

Figure 2 Training results of monitoring point B

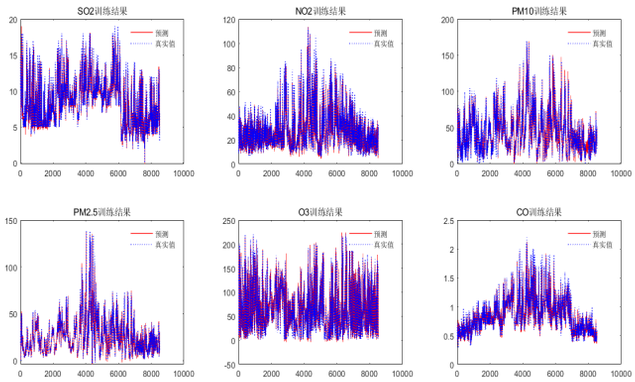

Figure 3 Training results of monitoring point C

3. Model Evaluation

Considering the different distances between A1,A2 and A3 monitoring points and main monitoring point A,we can build a collaborative prediction model.The closer the distance,the stronger the synergy.Therefore,when building the limit learning machine model,we need to assign different weights to the three auxiliary monitoring stations,so we build the weighted limit learning machine model.For the data input of the weighted limit learning machine model,we take A-A1-A2-A3 as the data source input,and the output result is the prediction result of monitoring point a predicted by A-A1-A2-A3.Through the performance analysis of the model,we find that the deviation between the predicted value of the secondary prediction model and the actual monitoring value of monitoring point A is large,and the deviation between the predicted value of the collaborative prediction model and monitoring point A is small,which shows that the prediction performance of the collaborative prediction model is better and the prediction is more accurate.

References:

[1] Liu Qian,Li CE,Yang Feng,Liu Libo,Deng Zhen.Research and application of weighted limit learning machine in pedestrian detection[J]. Computer engineering and design,2019,40(08):2366-2371

[2] He fahu,Liang Jiantao.Research on air quality prediction based on neural network[J].Modern computer,2021(18):64-67

[3] Wu Jianwei,Liu Fuyun,Li Qiao.Matlab genetic algorithm function GA optimization example[J].Mechanical engineering and automation,2017(02):61-63

[4] Xu Rui,Liang Xun,Qi Jinshan,Li Zhiyu,Zhang Shusen.Frontier progress and trend of extreme learning machine[J].Journal of computer science,2019,42(07):1640-1670

[5] Kuang Jia Qing.Classification and gene selection of breast cancer subtypes based on genetic algorithm and weighted extreme learning machine[D].Jilin University,2017.