Air Quality Index Prediction Model Integrating Multi-Head Self-Attention Mechanism

Lantao Yao [ab], Lizhi Liu *[ab]
[a] Hubei Key Laboratory of Intelligent Robot (Wuhan Institute of Technology), Wuhan, 430205, China;
[b] School of Computer Science & Engineering Artificial Intelligence, Wuhan Institute of Technology, Wuhan, 430205, China;
[*] Corresponding author: llz73@163.com

ABSTRACT

Aiming at the problem of insufficient accuracy of existing models in the process of air quality index prediction, an air quality index (AQI) prediction model (CALSTM) incorporating a multi-head self-attention mechanism is proposed based on Long Short-Term Memory network (LSTM). The model effectively extracts low-dimensional features of air pollutant concentrations and meteorological data related to the air quality index through Convolutional Neural Network (CNN) and uses LSTM to fully reflect the long-term historical process in the input time series. To improve the acquisition ability of the global information of the model context, the multi-head self-attention mechanism is used to extract the hidden information of the time series at different levels and improve the model's generalization ability. In addition, to further improve the prediction accuracy of the model, an AQI time series difference method is proposed based on the data correlation analysis. The experimental results show that the CALSTM using the AQI time series difference method has achieved an effect of 16.27% on MAPE, and on the indicators of MAE, MSE, RMSE, and R2, they are 6.99, 114.79, 10.71, and 0.9586, respectively. Compared with LSTM has achieved better prediction results.

Keywords: Air Quality Index Prediction, LSTM, CNN, Multi-Head Self-Attention Mechanism, Sequence Difference

1. INTRODUCTION

Related studies have shown that long-term exposure to air pollution can lead to an increased risk of cardiovascular disease[1]. Predicting air quality plays a vital role in protecting people's health. The existing air quality index (AQI) prediction methods mainly include numerical prediction model methods and statistical prediction model methods[2].

The numerical prediction model method mainly calculates the operating mechanism of the atmosphere through physical and chemical formulas and simulates the real atmospheric environment as much as possible. Falasca et al.[3] used the WRF-CHIMERE modeling system to study the influence of horizontal grid size, anthropogenic emissions, and the introduction of urban canopy models. Through experiments, it was found that a resolution of 4 km has obvious advantages.

The statistical prediction model method does not focus on chemical reactions but extracts useful information to fit the model according to the changes in historical pollutant concentration and the deep connection between meteorological data[4]. Statistical prediction model methods include traditional machine learning-based methods as well as deep learning-based methods. Chen et al.[5] proposed an autoregressive prediction model based on the perceived AQI value and fitted with an adaptive Kalman filter method, realizing effective prediction of AQI. In recent years, methods based on deep learning have received more attention. The mainstream methods include Convolutional Neural Network (CNN)[6], Recurrent Neural Network (RNN)[7], and Long Short-Term Memory network (LSTM)[8]. Fang W et al.[9] proposed a prediction model based on spatiotemporal similarity LSTM for the spatiotemporal similarity of data, and selected data with higher spatiotemporal similarity according to the proposed GCWDTW algorithm, and achieved good results. Gilik et al.[10] combined CNN with LSTM and used two methods of univariate model and a multivariate model to improve the prediction effect of the model. Existing studies have considered the temporal correlation of data, but have not paid enough attention to the low-dimensional features of the sequence and the correlation between parameters, resulting in room for improvement in the prediction accuracy of the model.

2. METHOD

To solve the above problems, this paper proposes an air quality prediction model CALSTM that incorporates a multi-head self-attention mechanism. The model uses convolution blocks to extract low-dimensional information in data, uses the LSTM network to extract time series information, and integrates a multi-head self-attention mechanism to extract hidden information of time series at different levels. For the problem that the correlation between the concentration of some pollutants and AQI is too high and the prediction accuracy is affected, the AQI time series difference method is proposed.

2.1 LSTM

LSTM is a variant of RNN. It is proposed to solve the problem of gradient explosion caused by subsequent nodes gradually forgetting the previous information as the number of network layers and iterations increase[11]. It can be used to capture time series and Long-term dependencies that exist.

2.2 Multi-head self-attention mechanism

The attention mechanism was originally used in the field of machine translation and later extended to time series problems with very good results. The multi-head attention mechanism was proposed by the Transformer model, and the full attention network was used to deal with machine translation problems. The attention weight is obtained by calculating the correlation between the target sequence and the dependent sequence. If the dependent sequence comes from the target sequence, the attention is self-attention. And each time the correlation operation is grouped (head), feature information can be extracted from multiple dimensions, which is multi-head self-attention.

To obtain the output corresponding to the input, splicing is performed after obtaining the operation results of each head, and then the fully connected layer without the activation layer is input for the linear operation to obtain the final output.

In the self-attention layer, the input sequence is calculated to obtain three vectors Q, K and V, and then these three vectors are calculated by the formula(1) to obtain the result, where d_K is the dimension of Q and K.

$$Attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d_K}})V \tag{1}$$

2.3 CALSTM

Based on CNN and LSTM, this paper proposes a CALSTM model that incorporates a multi-head self-attention mechanism for air quality index prediction. The model structure is shown in Figure 1.

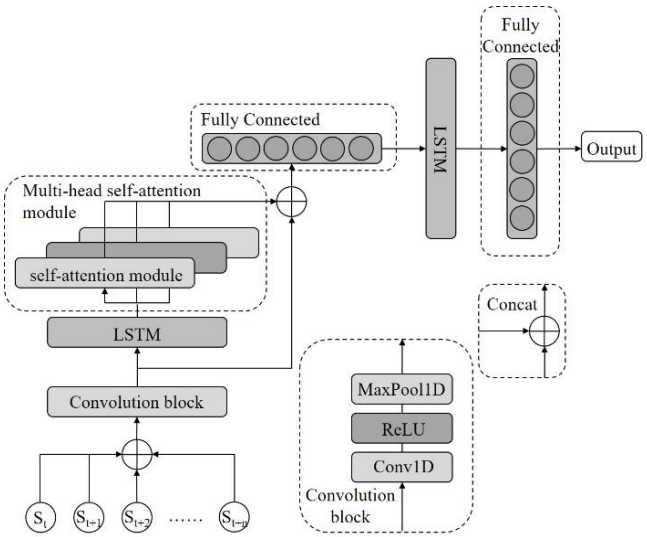

Figure.1 CALSTM Model Structure

As shown in Figure 1, the input of the model is a time series of S_t ,… , S_{t+n} with a step size of n, and the features of each time step are concatenated and used as the input of the convolution block. In this paper, a convolution block composed of one-dimensional convolution with ReLU nonlinear activation function and maximum pooling is designed in the model to extract low-dimensional features on time series. The output features of the convolutional block will be input to LSTM to further extract temporal features. In addition, this paper constructs a multi-head self-attention module for extracting hidden multi-dimensional information in time series. To reduce the impact of gradient disappearance caused by network deepening on model accuracy, this paper splices the multi-dimensional features extracted by the multi-head self-attention module and the low-dimensional features extracted by the convolution block according to the time dimension.

A fully connected layer is used to further extract the hidden features in the spliced feature sequence, and LSTM is used to obtain the sequence information in the hidden features. Finally, a fully connected layer is used to obtain the prediction results of the model.

2.4 AQI Time Series Difference Method

The Pearson correlation coefficient quantifies the linear correlation between two variables. The Pearson correlation coefficient between two variables is defined as the quotient of the covariance and standard deviation between the two variables, and its calculation formula is shown in formula (2):

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{2}$$

r in formula (2) is the Pearson correlation coefficient, X_i and Y_i are the data at position i in the sample, and $\overline{X}$ and $\overline{Y}$ are the average values of the two samples.

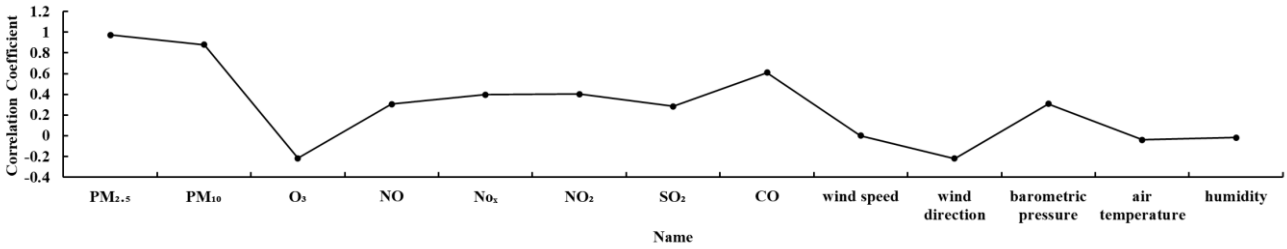

Figure.2 Pearson correlation coefficient between AQI and parameters

By calculating the Pearson correlation coefficient between AQI and other parameters in the data set, we can obtain Figure 2 after visualization. As shown in the figure, the absolute value of the correlation between $PM_{2.5}$, $PM_{10,}$ and AQI is greater than 0.8, the correlation is too high, and multiple correlations appear, that is, the high correlation between variables in the model affects the prediction accuracy of the model.

The difference method is used to correct the time series with multiple correlations to improve the prediction accuracy of the model. The difference method formula for time series is shown in formula (3), where $difference(t)$ represents the constructed difference variable at time t, and $observation(t)$ represents the observed value at time t.

$$difference(t) = observation(t) - observation(t-1) \tag{3}$$

The difference variable AQI_diff is constructed through the time series of difference AQI to reduce the correlation. Calculate the Pearson correlation coefficient between AQI_diff and other parameters, as shown in Figure 3.

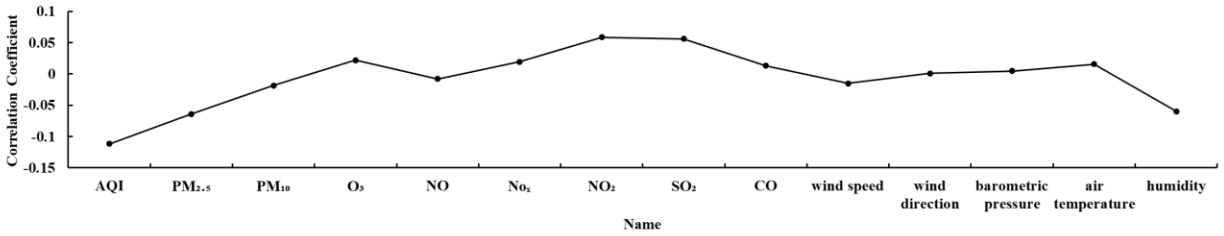

Figure.3 Pearson correlation coefficient between AQI_diff and parameters

As shown in Figure 3, the correlation between AQI_diff and other parameters is low, and the method of AQI time series difference is designed accordingly. The AQI time series difference method calculates the difference of the AQI sequence to obtain the sequence of the differential variable AQI_diff and takes AQI_diff as the prediction target of the model. After the model obtains the prediction result, use the formula (4) to obtain the predicted value of AQI, where $prediction(t)$ means The predicted value of AQI at time t, $pred\_diff(t)$ represents the predicted value of AQI_diff at time t, and $observation(t-1)$ represents the observed value of AQI at time $t-1$.

$$prediction(t) = pred\_diff(t) + observation(t-1) \tag{4}$$

3. RESULTS AND DISCUSSION

3.1 Dataset

The data in this article comes from 5 state-controlled monitoring stations in Yichang City. A total of 140,000 pieces of data were obtained from 5 stations from January 2019 to March 2022. The data items in the data set are shown in Table 1 except for time.

Table 1. Data items in the dataset.

Air Quality Evaluation Index	Concentration of Pollutants	Meteorological Data
AQI	$PM_{2.5}$/(ug·m^{-3}) PM_{10}/(ug·m^{-3}) O_3/(ug·m^{-3}) NO/(mg·m^{-3}) NO_2/(ug·m^{-3}) NO_x/(mg·m^{-3}) SO_2/(ug·m^{-3}) CO/(mg·m^{-3})	wind direction /deg wind speed /(m·s^{-1}) barometric pressure /Hpa air temperature /°C humidity /%

Due to the different dimensions in the original data, there are orders of magnitude differences between the data items, so use the formula (5) to normalize the data, where X is the normalized data, X is the original data, $X_{\max}$ and $X_{\min}$ are The maximum and minimum values of the data.

$$X = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{5}$$

Finally, the processed data are divided into chronological order, the first 80% of the data of each monitoring station is divided into the training set, and the last 20% is divided into the test set.

3.2 Ablation experiment

The experiment in this paper is based on the development of PyTorch, which is an open-source Python machine-learning framework based on Torch. It provides a variety of operations on tensors and a variety of neural network module interfaces, which facilitates the construction of dynamic neural network models and supports GPU/TPU accelerated computing.

In addition to implementing the CALSTM model and the LSTM model, the experiments in this paper also separate the convolutional layers of the CALSTM to form an independent ALSTM model as a comparison model for evaluating the performance of the CALSTM model. Through the analysis of data correlation in 1.4, it is found that the correlation between the concentration of some pollutants and AQI is too high, which has an impact on the prediction accuracy of the network. Therefore, the difference method of the AQI time series is used to change the prediction target to the AQI difference sequence AQI_diff based on the original input parameters Add the AQI index at the current moment, and CALSTM+ represents the CALSTM model using the AQI time series difference method.

During the experiment, the data of the Wujiagang monitoring station in the five stations of the data set were used as the experimental object to predict the AQI in the next hour. The loss function of the model is set to MSE, and the adaptive motion estimation algorithm (Adam) is used for optimization. The number of heads in the multi-head self-attention mechanism in the model is 8, the learning rate of the model is 0.00001, the number of samples included in each training batch_size is 100, and the number of training rounds epoch is 100.

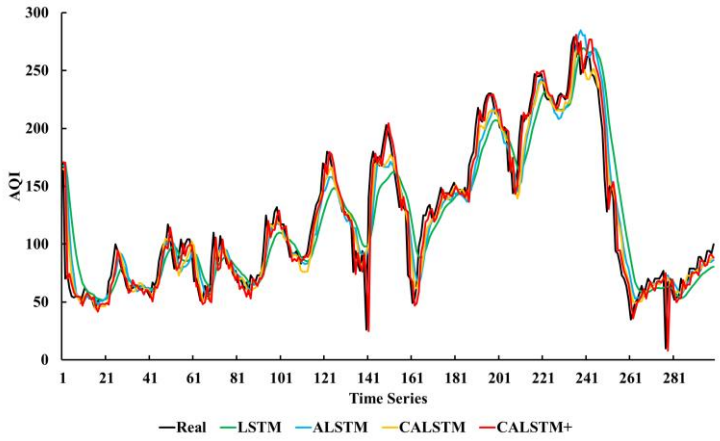

Figure.4 Prediction effect of each model

Figure 4 shows the prediction results of each model on the test set. CALSTM+ has the highest prediction accuracy and the best fit between the prediction and the true value.

As shown in Table 2, it is the evaluation index of each model on the test set. CALSTM+ using the improved AQI time series difference method achieved a result of 16.27% on MAPE, MAE was 6.99, MSE was 114.79, RMSE was 10.71, and R^2 was 0.9586, which are the best compared with other models.

Table 2. Evaluation indicators of each model

Model	MAPE (%)	MAE	MSE	RMSE	R2
LSTM	24.30%	11.30	315.70	17.77	0.8873
ALSTM	21.53%	9.39	216.51	14.71	0.9227
CALSTM	21.26%	8.55	171.51	13.10	0.9388
CALSTM+	16.27%	6.99	114.79	10.71	0.9586

In addition, the other 4 stations in the data set were selected for experiments, namely Bailonggang Monitoring Station, Dianjun District Monitoring Station, Yiling District Monitoring Station, and Xiaoting District Zhangjiawan Monitoring Station. These 4 stations are distributed in different locations in Yichang City. There is a certain difference from the Wujiagang monitoring station selected in the above experiment, to verify the generalization ability of the model.

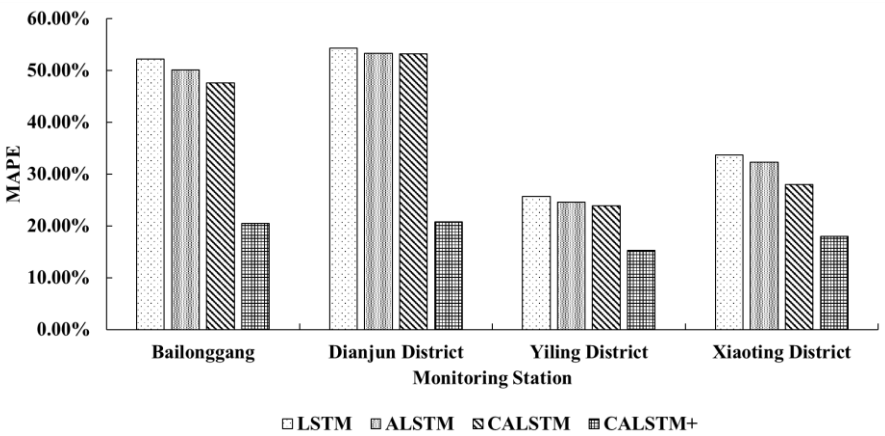

Figure.5 MAPE of each model on the dataset of other monitoring stations

As shown in Figure 5, on the MAPE index, the CALSTM model proposed in this paper performs the best. Using the time series difference method further improves the model's prediction accuracy, and the fitting effect is better.

4. CONCLUSIONS

Based on LSTM, this paper proposes an air quality index prediction model CALSTM with a fusion attention mechanism. In the CALSTM model, first pass the input through the convolution block to obtain a low-dimensional feature sequence, use LSTM to extract the temporal features in the low-dimensional feature sequence, and then use the multi-head self-attention mechanism module to extract multi-dimensional features in the temporal features. After splicing the multi-dimensional features and the low-dimensional feature sequence obtained by the convolution block, the fully connected layer is used to further extract the hidden features in the sequence, and the LSTM is used to extract the sequence information in the hidden features, and finally, the output is obtained through a fully connected layer.

Based on data correlation analysis, this paper proposes a different method of AQI time series, which further improves the prediction accuracy of the model.

The experimental results show that the CALSTM model using the time series difference method has achieved effects of 16.27%, 6.99, 114.79, 10.71, and 0.9586 on the indicators of MAPE, MAE, MSE, RMSE, and R^2, respectively. Comparison with the prediction effect of LSTM is better.

ACKNOWLEDGMENT

This work is supported by the Scientific Research Plan of Guidance Project of the Hubei Provincial Department of Education (B2017051), Innovation Foundation of Hubei Key Laboratory of Intelligent Robot (HBIRL202207), and the 14th Graduate Education Innovation Fund of Wuhan Institute of University (CX2022346).

REFERENCES

[1] Liang, F., Liu, F., Huang, K., et al. Long-Term Exposure to Fine Particulate Matter and Cardiovascular Disease in China [J] Journal of the American College of Cardiology, 75(7), 707–717. DOI: 10.1016/j.jacc.2019.12.031

[2] Zhu Y M, Xu A L, Sun Q. New progress for air quality forecasting methods based on deep learning[J]. Environmental Monitoring in China, 2020, 36(3):10-18. DOI: 10.19316/j.issn.1002-6002.2020.03.02.

[3] Falasca S, Curci G. High-resolution air quality modeling: Sensitivity tests to horizontal resolution and urban canopy with WRF-CHIMERE [J]. Atmospheric Environment, 2018, 187(AUG.): 241-254. DOI: 10.1016/j.atmosenv.2018.05.048

[4] Wang F. Research on Forecasting Method of Urban Ambient Air Quality [J]. Environment and Development, 2020, 32(09):176+178. DOI: 10.16647/j.cnki.cn15-1369/X.2020.09.097.

[5] Chen J, Chen K , Ding C , et al. An Adaptive Kalman Filtering Approach to Sensing and Predicting Air Quality Index Values[J]. IEEE Access, 2020, PP(99):1-1. DOI: 10.1109/ACCESS.2019.2963416
[6] Lecun Y, Bottou L . Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324. DOI: 10.1109/5.726791
[7] Liu J W, Song Z Y. A Survey of Research on Recurrent Neural Networks [J]. Journal of Computer Applications,2018,38(S2):1-6+26. DOI: 10.13195/j.kzyjc.2021.1241
[8] Hochreiter S , Schmidhuber J . Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780. DOI:10.1162/neco.1997.9.8.1735
[9] Fang W, Zhu R S. Air Quality Prediction Model Based on Spatial-Temporal Similarity LSTM [J]. Application Research of Computers,2021,38(09):2640-2645.DOI:10.19734/j.issn.1001-3695.2021.01.0011.
[10] Gilik A, Ogrenci A S, Ozmen A . Air quality prediction using CNN+LSTM-based hybrid deep learning architecture[J]. Environmental Science and Pollution Research, 2022, 29(8):11920-11938. DOI: 10.1007/s11356-021-16227-w
[11] Wang X, Wu J, Liu C, et al. Fault Time Series Prediction Based on LSTM Recurrent Neural Network [J]. Journal of Beijing University of Aeronautics and Astronautics, 2018,44(04):772-784. DOI: 10.13700/j.bh.1001-5965.2017.0285.