Atmospheric Pollutant Prediction Based on Wavelet Decomposition and Long Short-Term Memory Network

Lei Li, Zhexiang He*
Department of Environment Engineering, School of Metallurgy and Environment, Central South University, Changsha 410083, China

*Corresponding author e-mail: hezx@csu.edu.cn

Abstract. To solve the problem that current atmospheric pollutant prediction research only pays attention to one pollutant type, a Long Short-term Memory Network (LSTM) atmospheric pollutant prediction model based on Wavelet Decomposition (WD) is proposed, predicting daily average PM$_{2.5}$, PM$_{10}$, SO$_2$, NO$_2$ and O$_3$ concentration for the next day. Based on data collected from the national control station (NO.1338A) in Changsha from 2015 to 2018, the model was verified. The results show that for the prediction of atmospheric pollutants, compared with the LSTM prediction model established directly with the original atmospheric pollutants, the root mean square error and the absolute average error of the hybrid model are lower, and the accuracy of the level prediction is higher.

1. Introduction

Nowadays, neural networks are often used to predict atmospheric pollutants. In recent years, with the continuous improvement of computer performance, various deep neural networks have been developed based on early neural networks, such as the Convolutional Neural Network (CNN), the Recurrent Neural Network (RNN), and the Long Short-term Memory Network (LSTM) [1-3]. For example, ZHAO uses the fully connected layer LSTM to predict mid and long-term PM$_{2.5}$ hourly concentrations in Beijing [4]. Similarly, V. Yadav predicts daily average PM$_{10}$ concentrations using an ANN model combined with principal components analysis [5]. In these neural networks, LSTM was originally applied to time series such as text translation and speech recognition[6,7]. Since atmospheric pollutant prediction is also a time series, the use of LSTM to predict atmospheric pollutants has achieved better results than the application of other neural networks[8-10].However, through literature research, we also found that the current prediction of atmospheric pollutants based on neural networks has the following issues:

(1) Numerous studies have focused on PM$_{2.5}$, while less attention has been paid to other pollutants. In many countries, air quality monitoring not only focuses on PM$_{2.5}$, but also includes PM$_{10}$, SO$_2$, NO$_2$, CO and O$_3$. These pollutants are also harmful to humans and therefore need to be included in the forecasting category.

(2) In air pollutant forecasting, many studies focus on the short-term hourly prediction of pollutants, while daily average concentrations are largely neglected. However, in practical applications, the daily average concentration is the focus of the forecast and an important measure of future air quality.

(3) In the prediction of hourly concentration values, both the pollutant data and the meteorological data are often used as the input data of the model. However, using such an approach, the correlation coefficient (Person’s correlation test) between the meteorological data and the pollutant data is
relatively small. Because the weather is subject to rapid changes throughout the day, the addition of low-degree-of-correlation meteorological data can greatly increase the uncertainty and complexity of the system, potentially adversely affecting the prediction accuracy of the model.

To overcome these limitations and to fill this research gap, we combined wavelet decomposition and LSTM for atmospheric pollutant prediction. Wavelet decomposition is used to decompose low-dimensional atmospheric pollutant data into high-dimensional data and to fully exploit the information value of raw data. We then used LSTM to predict the decomposition data and compared it with the data obtained by directly using the atmospheric pollutants for LSTM prediction to verify the feasibility of the model.

2. Methods

2.1. Wavelet decomposition and reconstruction

Wavelet decomposition represents the use of finite length or fast attenuation of the oscillation waveform by scaling and translation to represent the signal, based on the local transformation of time and frequency, followed by effective information extraction from the signal (research data). Taking three-layer wavelet decomposition as an example, the decomposition diagram is shown in Figure 1.

![Figure 1 Three-layer wavelet decomposition](image)

Through the high-pass filter and the low-pass filter, the original sequence $S$ is decomposed into $D1$ and $A1$ sequences, and then the low-frequency signal $A1$ is decomposed in two steps to obtain the final decomposed sequences $A3$, $D1$, $D2$ and $D3$. The wavelet reconstruction is to restore the decomposed sequence to the original sequence, which can be completed by equation (1).

$$S = A3 + D1 + D2 + D3$$

2.2. LSTM

The basic unit of the LSTM model is a memory block, which includes a memory cell and three gate structures that control the cell state, namely a forget gate, an input gate, and an output gate. Forgetting the door decides to forget the useless history information from the memory unit state, the input gate determines the influence of the current input data on the state of the memory unit, and the output gate determines the output information. Compared to the RNN, the LSTM provides a more flexible choice of information to participate in the input. Currently, using the Keras platform developed by Google, it is only necessary to determine the number of neurons per layer, layer, optimizer, activation function, dropout, batch size, epoch and other parameters to build a complete LSTM prediction model.

2.3. Model prediction error evaluation index

To measure the prediction effect of the model, the root mean square error (RMSE) and the mean absolute percentage error (MAPE) were used. In addition, in China's daily air quality forecast, different air quality levels are often used to visually reflect future air conditions. According to the Chinese Environmental Protection Standard (HJ633-2012), air quality is divided into six grades (excellent, good, slightly polluted, moderately polluted, heavily polluted, and severely polluted). Therefore, to further verify the effect of the prediction model in practical applications, the pollution level prediction accuracy (ACC) is also used, calculated via the following equations:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right|$$

$$ACC = \frac{1}{n} \sum_{i=1}^{n} I(x_i = \hat{x}_i)$$

where $x_i$ is the actual value, $\hat{x}_i$ is the predicted value, $n$ is the number of samples, and $I(x_i = \hat{x}_i)$ is an indicator function that returns 1 if $x_i = \hat{x}_i$ and 0 otherwise.
\[ MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times \frac{100}{n} \]  

(3)

\[ ACC = \frac{n_i}{n} \times 100\% \]  

(4)

where \( n \) is the total number of predicted values, \( x_i \) and \( \hat{x}_i \) are the predicted values and the corresponding actual values, respectively, and \( n_i \) is the number of corresponding predicted values and actual values at the same pollution level.

3. The study

The data source is the daily average air pollutants (PM$_{2.5}$, PM$_{10}$, SO$_2$, NO$_2$, CO and O$_3$) concentration from 2015-2018 collected by the ground state control site (NO.1338A) of Changsha City, China. According to the data analysis, the CO concentration has been at the first level in recent years, and the predicted value is low, so it is eliminated.

3.1. LSTM model

We used data from 2015 to 2017 as training set and data from 2018 as test set. The daily average concentrations of the corresponding atmospheric pollutants on the previous day were used as inputs, and the concentrations of atmospheric pollutants in the next day were used as the output. The optimal parameters in the PM$_{2.5}$ one-step prediction of 1338A, obtained through experiments, are shown in Table 1. The parameters of other pollutants prediction models are the same as those in Table 1.

| pollutant | Neurons of every layer | Layer | optimizer | Activation function | Dropout | Batch size | epoch |
|-----------|------------------------|-------|-----------|---------------------|---------|------------|-------|
| PM2.5     | 200                    | 3     | Adam      | Tanh                | 0.25    | 4          | 500   |

3.2. LSTM based on wavelet decomposition (WD-LSTM) model

Taking PM$_{2.5}$ prediction as an example, first, we executed coif3 wavelet decomposition to the original PM$_{2.5}$ time series. Via the three-level coif3 wavelet decomposition, the series was decomposed into the low-frequency part A3 and the high-frequency parts D1, D2, and D3. Then use the decomposition sequence corresponding to the previous day as input and the value of the next day as output. Taking the PM$_{2.5}$ data of 1338A as an example, table 2 shows the optimal neural network parameters. The parameters of other pollutants prediction models are the same as those in Table 2.

| Decomposition sequence | Neurons of every layer | Layer | optimizer | Activation function | Dropout | Batch size | epoch |
|------------------------|------------------------|-------|-----------|---------------------|---------|------------|-------|
| A3                     | 200                    | 3     | Tanh      | Adam                | 0.25    | 4          | 500   |
| D1                     | 150                    | 2     | Tanh      | Adam                | 0.25    | 2          | 500   |
| D2                     | 150                    | 2     | Tanh      | Adam                | 0.25    | 2          | 500   |
| D3                     | 150                    | 2     | Tanh      | Adam                | 0.25    | 2          | 500   |

4. Results and Discussion

4.1. LSTM model

The one-step predicted results of each pollutant in 1338A are shown in Figure 2. Compared with other pollutants, the prediction effect for O$_3$ was poor, especially at high concentrations. For other pollutants, the fluctuation state was basically identical with the true value, although the fitting degree was insufficient. The RMSE, MAPE, and ACC are shown in Table 3.
According to Table 3, there were also significant gaps in the predictions when predicting different atmospheric pollutants. In contrast to RMSE, MAPE, and considering the original concentrations of atmospheric pollutants, we found that the predicted effects for PM$_{2.5}$, PM$_{10}$, SO$_2$ and NO$_2$ were considerably better than that for O$_3$. The MAPE of O$_3$ was significantly higher than that of other atmospheric pollutants. Most likely, this is because in the prediction of O$_3$, some true values were single digits, such as 9 $\mu$g/m$^3$, and the corresponding predicted value was 5 $\mu$g/m$^3$, which makes the MAPE being larger than that of other atmospheric pollutants.

The opposite results were observed for the ACC. The ACC of SO$_2$ and O$_3$ were considerably higher than those of other atmospheric pollutants; in particular, for SO$_2$, we observed an ACC of 100%. Most likely, the prediction accuracies for PM$_{2.5}$ and PM$_{10}$ were about 20% lower than those of other pollutants because, according to the National Environmental Protection Standard HJ 633-2012, the threshold of the air quality levels of PM$_{2.5}$ and PM$_{10}$ are lower than those of other pollutants such as SO$_2$ and O$_3$, and the true value and the predicted value are less likely to fall into the same interval. Also, because Changsha is not a heavy industrial city, air pollution control in recent years has resulted in low levels of sulfur oxides. Moreover, in recent years, atmospheric particulate matter has been the primary atmospheric pollutant. Although several measures have been proposed to decrease atmospheric particulate matter levels, they are still relatively high compared to those of other pollutants.
4.2. WD-LSTM

The predicted result is shown in Figure 3. True and predicted values have a large number of coincident regions, and when the true value fluctuates rapidly, the predicted value can also be well fitted. Especially for the prediction of extreme values, the hybrid model showed excellent performance, which is crucial for the prediction of extreme weather events. Compared with the LSTM, the prediction effect of WD-LSTM was significantly better. For example, for O3, which performs poorly in LSTM model predictions, the prediction effect of the hybrid model was considerably improved.

![Fig. 3 One-step prediction of air pollutant concentration in 1338A by the WD-LSTM](image)

Table 4 shows the RMSE, MAPE, and ACC values of the WD-LSTM. In one-step prediction, the RMSE, MAPE, and ACC of the WD-LSTM model were better than those of the LSTM model. The RMSE and MAPE accounted for about one-half of the ACC, indicating that the hybrid model fits the pollutant prediction better, with a more than 10% increased prediction accuracy for PM$_{2.5}$, PM$_{10}$ and NO$_2$.

| Station | Step | PM$_{2.5}$ | PM$_{10}$ | SO$_2$ | NO$_2$ | O$_3$ |
|---------|------|------------|------------|--------|--------|-------|
| code    | size | RMSE (μg/m$^3$) | MAPE (%) | AR (%) | RMSE (μg/m$^3$) | MAPE (%) | AR (%) | RMSE (μg/m$^3$) | MAPE (%) | AR (%) | RMSE (μg/m$^3$) | MAPE (%) | AR (%) |
| 1338A   | 1    | 9.943      | 18.1       | 83.1   | 13.276 | 21.8  | 12.376 | 21.8  | 89.5   | 1.861  | 15.1  | 100   | 4.484  | 16.3  | 95.2  | 12.950 | 35.2  | 95.5  |

In contrast to the models previously described [11-13], the WD-LSTM not only requires fewer input parameters (Single atmospheric pollutant time series), but also provides a solution for predicting multiple atmospheric pollutants.
5. Conclusions

In this paper, we established a one-step prediction model for the daily average concentration of atmospheric pollutants, based on wavelet decomposition and LSTM. Using wavelet decomposition, the information carried by atmospheric pollutant data can be effectively extracted. At the same time, the decomposition sequence has less noise and is therefore more conducive to model prediction.

To validate the effectiveness of the prediction model, a case study was conducted in Changsha, China. The results show that the WD-LSTM only requires the decomposition sequences of the corresponding pollutants for prediction, and the prediction results are better than those of the LSTM.

6. Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities of Central South University (Project (2019zzts513)

References
[1] Gregor K, Danihelka I, Graves A, et al. Draw: A recurrent neural network for image generation[J]. arXiv preprint arXiv:1502.04623, 2015.
[2] Sak H, Senior A, Beaufays F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition[J]. arXiv preprint arXiv:1402.1128, 2014.
[3] Kim P. Convolutional Neural Network[M]. 2017.
[4] Zhao J, Deng F, Cai Y, et al. Long short-term memory-Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction[J]. Chemosphere, 2019, 220: 486-492.
[5] Yadav V, Nath S. Novel hybrid model for daily prediction of PM 10 using principal component analysis and artificial neural network[J]. International Journal of Environmental Science and Technology, 2019, 16(6): 2839-2848.
[6] Wu Y, Schuster M, Chen Z, et al. Google's neural machine translation system: Bridging the gap between human and machine translation[J]. arXiv preprint arXiv:1609.08144, 2016.
[7] Graves A, Jaitly N, Mohamed A. Hybrid speech recognition with deep bidirectional LSTM[C]//2013 IEEE workshop on automatic speech recognition and understanding. IEEE, 2013: 273-278.
[8] Pak U, Kim C, Ryu U, et al. A hybrid model based on convoluted neural networks and long short-term memory for ozone concentration prediction[J]. Air Quality, Atmosphere & Health, 2018, 11(8): 883-895.
[9] Wang B, Kong W, Guan H, et al. Air Quality Forecasting based on Gated Recurrent Long Short Term Memory Model in Internet of Things[J]. IEEE Access, 2019.
[10] Qing T, Fang L, et al. Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU.IEEE access[J].2019:76692-76698
[11] Zamani Joharestani M, Cao C, Ni X, et al. PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data[J]. Atmosphere, 2019, 10(7): 373.
[12] Sun W, Sun J. Daily PM2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm[J]. Journal of environmental management, 2017, 188: 144-152.
[13] Zhu S, Lian X, Liu H, et al. Daily air quality index forecasting with hybrid models: A case in China[J]. Environmental pollution, 2017, 231: 1232-1244.