Time-series prediction model of PM2.5 concentration based on LSTM neural network

Yiyang Li*
School of Information and Communication Engineering, Communication University of China, Beijing, 100024, China
*lyy1026@cuc.edu.cn

Abstract. LSTM (long-short term memory neural network) is a neural network suitable for simulating and predicting time series changes. This paper designs a PM2.5 concentration prediction model based on the LSTM neural network. Use Python language programming implementation under the TensorFlow framework, training the model using real weather data from 2015-01-01~2019-01-01, and the obtained model can accurately predict the concentration change of future PM2.5. RMSE and DC were used as evaluation indexes to compare the LSTM neural network model with the traditional BP (Back Propagation) algorithm model. It turns out, LSTM model has the characteristics of high precision and stable prediction effect, proved that the time series prediction ability of the LSTM model is more superior. As an effective method for predicting PM2.5 concentration data, it can provide the basis for predicting pollutants.

1. Introduction
With the continuous development of industrial society, the degree of environmental pollution is aggravated, among which the air pollution is severe. PM (particulate matter) is an important index to measure the degree of air pollution. The higher the concentration in the air is, the more serious the air pollution is. PM2.5 are prone to attach toxic and harmful substances (e.g. heavy metals, microorganisms, etc.), and stay in the atmosphere for a long time, transport distance, not only affect the balance of ecosystems, but also cause harm to human health[1]. How to predict the concentration of PM2.5 so as to take reasonable prevention and control measures has become a hot topic.

Existing PM2.5 concentration prediction models mainly include the mechanism model, traditional time series model and machine learning model. The prediction method based on the mechanism model is to simulate the future air pollution by the physicochemical process of air pollutants. This kind of model process is not only complicated, but also challenging to obtain various parameters accurately, which makes the model accuracy not high[2]. The time series model can not construct the complicated nonlinear relationship in the PM2.5 data by using the historical time series information to predict the future PM2.5 concentration, so the prediction effect is not good. The shallow network based on machine learning, represented by artificial neural network and improved grey neural network model, can model complex nonlinear relationship prediction, but because the network structure is relatively simple, the prediction effect is generally[3]. Moreover, deep learning, as a new machine learning model, shows significant advantages in dealing with complex time series relationships. It effectively learns a large number of input data features and provides new research ideas and methods for PM2.5 time series prediction. Most of the research on prediction models only build concentration models with PM2.5 historical concentration or various meteorological elements as independent variables, but ignore that the
change of PM2.5 concentration is affected by its historical concentration and meteorological elements. The prediction of PM2.5 concentration needs to synthesize these two factors.

This paper proposes an LSTM neural network prediction model for PM2.5 prediction under the framework of TensorFlow depth learning, using the historical concentration and various meteorological elements as independent variables. Use real weather data training model and verify the prediction results. Furthermore, the results are compared with the constructed BP neural network prediction model.

The model has good nonlinear learning ability in PM2.5 data prediction, and the prediction results are more accurate than BP neural networks.

2. Model approach

2.1. LSTM Neural Networks

The neural network is a mathematical model that simulates the processing of information similar to the neural synaptic connection structure of the brain. It studies massive historical data to find the best way to solve the problem[4]. The prediction of PM2.5 pollutants needs to find the mapping relationship in the historical load data and then use it to predict the future data. Therefore, the neural network model can be used to solve the prediction problem. RNN (recurrent neural network) of neural network has a remarkable effect in processing time-series data. RNN neural network, different from the basic neural network, focuses on the cycle, not only establishing the weight connection between the layers, but also establishing the connection between the neurons in each layer. Use such sequential information to predict future changes (shown in Fig.1).

![Fig.1 Schematic diagram of RNN network model](image1)

The input $X_t$, output of the neural network unit is $Y_t$. Each operation of RNN neural network depends on the previous calculation results, in theory, the sequence information of any length can be used, but as the time interval of training data increases, the RNN learning ability will become weaker. The model will appear gradient descent or gradient explosion with the increase of time interval.

LSTM neural network, as a variant of RNN neural network, controls forgetting or adding information by adding cell state design similar to ‘highway’ and allowing information to be selectively designed through ‘gate’. In order to avoid RNN problems caused by the length of time series.

The gate structure of LSTM neural network includes three kinds: Input Gate, Output Gate and Forget Gate, which learn and predict the data by controlling the input, output and historical dependence of neurons (shown in Fig.2).

![Fig.2 Schematic diagram of LSTM network model](image2)
LSTM neurons include the following formula:

\[
\delta(x) = \frac{1}{1 + e^x}
\]

\[
f_t = \delta(W_f \cdot [y_{t-1}, x_t] + b_f)
\]

\[
i_t = \delta(W_i \cdot [y_{t-1}, x_t] + b_i)
\]

\[
C_t = e^{W_c \cdot [y_{t-1}, x_t] + b_c} - e^{-(W_c \cdot [y_{t-1}, x_t] + b_c)}
\]

\[
o_t = \delta(W_o \cdot [y_{t-1}, x_t] + b_o)
\]

\[
y_t = o_t \cdot \tanh(C_t)
\]

In the above formula, \( C \) is the cell state, the \( W \) is the corresponding weight coefficient matrix, and the \( b \) is the corresponding bias term.

According to the above formula, the principle of forgetting Gate is to read the output of the previous neuron and the current neuron input, and outputs a value between \([0,1]\) multiplied by the previous cell state. If output 1, it represents complete ‘memory’ of the last cell state information. If output 0, it represents complete ‘obliteration’ to the previous cell state. In this way, the long-term memory of the network model is guaranteed (formula (2)). The input gate receives \( x_t \) and \( y_{t-1} \) to obtain data and new candidate vectors (formula (3, 4)). Then, the output result \( o_t \) of the neuron is obtained by \( x_t \) and \( y_{t-1} \) through the output gate (formula (5)).

The final output \( y_t \) of the current neuron in formula (6) is obtained by multiplying the current cell state \( C_t \) processed by the \( \tanh \) layer with the results \( o_t \) obtained from the output gate.

2.2. BP Neural networks

BP neural network is one of the most widely used artificial neural network models. It is a multilayer feedforward network, which is trained by error inverse propagation algorithm. It can learn and store a large number of mathematical equations of input-output pattern mapping relation[5]. Its network model topology includes input layer, hidden layer and output layer. The input data first reaches the input layer and then goes to the output layer after the hidden layer training and learning. The hidden layer can be one or more layers, which are composed of multiple neurons. The state of each layer of neurons is only related to the state of the next layer of neurons. The topology of BP neural network is shown in Fig.3. If the output results obtained in the output layer do not meet the expected target, the error signal is returned along the original connection path, and the weight of each neuron is changed. Repeat the positive and negative calculations until the calculated results meet the requirements.

![Fig.3 Structure of BP neural network model](image-url)
3. Case analysis

3.1. Experimental data sets
The real data from 2015-01-01~2019-01-01 are selected as the experimental data set. The data set consists of 24 full-time weather data per day for the four years, with a total of 35040 (24×365×4) load data, each including the PM2.5 concentration, temperature, air pressure, wind direction, wind speed, accumulated hour snow, accumulated hour rainfall at that time point (shown in Fig.4).

![Fig.4 Original data presentation (first 2000)](image)

3.2. Data preprocessing
Due to the human operation or equipment failure, the missing data can be completed in data acquisition. By observing the data set, there is no large number of continuous missing and PM2.5 continuity of the data, so the average value of the missing value is used to complement the missing data in the experiment. All the above seven kinds of data are numerical data except wind direction data, but one-hot coding is used for wind direction data because of its unintentional value.
LSTM model is sensitive to the input data scale, and each index varies significantly in the order of magnitude. In order to reduce the influence of the order of magnitude difference on the prediction, the original data should be normalized. Normalize the completed data using the following formula:

\[
x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}
\]

The \( x \) is the original data, the \( x_{\text{min}} \) is the minimum value of the original data, the \( x_{\text{max}} \) represents the maximum value of the original data, and the normalized data (a value between \([0,1]\)) can reduce the data scale on the premise of ensuring the same trend of data.

### 3.3. Model Building

#### 3.3.1. LSTM model

In the LSTM model, the loss function is used to measure the proximity between the calculated value and the measured value, and the model parameters are determined according to the principle of the minimum loss function. Choose the average absolute error MAE as the loss function.

\[
MAE = \frac{1}{m} \sum_{i=1}^{m} |y_0(i) - y_f(i)|
\]

In the formula above, \( m \) is the number of samples; \( y_0 \) is the measured PM2.5 concentration; \( y_f \) is the concentration that the model calculates.

Divide the experimental data into two parts. The first three years are used as a model training set. Data for the fourth year as a test set, compare with the model prediction results to test the prediction effect of the model. Iterate the training data into the designed LSTM model. By continually tuning the model, finally, the model parameters are as follows: the number of hidden neurons is 60, 72 batches of training data, the initial learning rate is 0.001, and the number of iterations is set to 80.

In addition, the AdamOptimizer is selected to dynamically adjust the learning rate. The AdamOptimizer can adjust the learning rate adaptively from the two perspectives of gradient mean and gradient variance, and the effect is good.

From the diagram, it can be seen that the loss function decreases gradually and tends to be stable.
and there is no phenomenon of underfitting or overfitting, which indicates that the LSTM model rate is reasonable.

3.3.2. BP model
A total of 1 input layer, 3 hidden layers and 1 output layer structure are set up BP the neural network in this paper. The number of neurons in 3 hidden layers is 3, 5, 7, and the learning rate and iteration number are set to 0.01, 1000, respectively.

4. Experimental results and analysis

4.1. Training results
Using the trained LSTM, BP model generates the forecast data predict_steps. for the next 200 hours, the final experimental results are shown in the figures.

![Fig.7 Prediction results of LSTM model](image)

![Fig.8 Prediction results of BP model](image)

The graph shows 200 hours of data, the orange line is the trained prediction data (the last 200 hours of the data set), and the blue line is the real data of the part to be predicted.

4.2. Model evaluation indicators
Based on RMSE (Root Mean Square Error) and DC (Deterministic Coefficient), the accuracy of the model is evaluated.

(1)RMSE:

\[ Q = \sqrt{\frac{1}{m} \sum_{i=1}^{m} [y_o(i) - y_r(i)]^2} \]  

(10)

(2)DC:
In the formula above, \( m \) is the number of samples; \( y_0 \) is the measured PM2.5 concentration; \( y_f \) is the concentration that the model calculates, and \( y_a \) is the average value of the measured concentration. Obviously, \( Q \) measures the average proximity between the predicted results and the measured values. The smaller the value, the higher the prediction accuracy. Moreover, \( DC \) reflects the fitting effect of the prediction process and the measured process as a whole. The greater the value of \( DC \), the higher the model accuracy, and vice versa.

4.3. Comparison of evaluation indicators

Table 1. Comparison of evaluation indicators.

|   | Q       |   | DC       |   |
|---|---------|---|----------|---|
|   | LSTM    | BP| LSTM     | BP|
| 1 | 26.7860 | 27.7065 | 0.9228 | 0.9046 |
| 2 | 26.7060 | 27.5577 | 0.9224 | 0.9108 |
| 3 | 26.6440 | 27.4146 | 0.9227 | 0.9110 |
| 4 | 26.6760 | 27.2168 | 0.9225 | 0.9121 |
| 5 | 26.6840 | 30.9547 | 0.9225 | 0.8868 |
| 6 | 26.5730 | 27.9666 | 0.9221 | 0.9071 |
| 7 | 26.6950 | 27.2485 | 0.9224 | 0.9117 |
| 8 | 26.6160 | 28.5063 | 0.9229 | 0.9036 |
| 9 | 26.6330 | 27.3457 | 0.9228 | 0.9113 |
| 10| 26.6450 | 27.8058 | 0.9227 | 0.9089 |
| 11| 26.6560 | 27.3023 | 0.9227 | 0.9117 |
| 12| 26.7060 | 27.3500 | 0.9224 | 0.9112 |
| 13| 26.6890 | 27.6644 | 0.9225 | 0.9091 |
| 14| 26.6900 | 28.8934 | 0.9225 | 0.9008 |
| 15| 26.6200 | 27.7214 | 0.9229 | 0.9092 |
| 16| 26.6120 | 27.9790 | 0.9229 | 0.9077 |
| 17| 26.6750 | 31.4168 | 0.9226 | 0.8834 |
| 18| 26.5960 | 27.4874 | 0.9230 | 0.9102 |
| 19| 26.5490 | 27.2168 | 0.9233 | 0.9121 |
| 20| 26.5500 | 30.9547 | 0.9233 | 0.8867 |

According to RMSE, the table shows that the LSTM model is superior to the BP model in 20 training sessions. Moreover, LSTM model is better than the DC of BP model.

5. Conclusions

This paper selects the real pollution and weather data as the research object, constructs the future PM2.5 concentration forecast model based on the LSTM neural network and the BP neural network, and evaluates the model accuracy with the root mean square error and the deterministic coefficient as the index. The main conclusions are as follows:

On average, the root mean square error predicted by the LSTM neural network model can be maintained below 26.7, while the root mean square error of the BP neural network reaches more than
many times. According to the deterministic coefficient, LSTM can maintain the prediction of accuracy above 0.922, which is higher than the maximum accuracy of BP model. As can be seen, the LSTM model has high accuracy and good applicability for the studied weather data.

To sum up, LSTM neural network has long-term dependent learning ability compared with the BP neural network, which provides an effective solution for long period PM2.5 prediction.

References
[1] Xiao Han, Yuqin Liu, Hong Gao, et al. Forecasting PM 2.5 induced male lung cancer morbidity in China using satellite retrieved PM 2.5 and spatial analysis. 2017, 607-608:1009-1017.
[2] Mark D. Gibson, Soumita Kundu, Mysore Satish. Dispersion model evaluation of PM 2.5, NO x and SO 2 from point and major line sources in Nova Scotia, Canada using AERMOD Gaussian plume air dispersion model. 2013, 4(2):157-167.
[3] Qian Yin, Jinfeng Wang, Maogui Hu, Hoting Wong. Estimation of daily PM_{2.5} concentration and its relationship with meteorological conditions in Beijing[J]. Journal of Environmental Sciences, 2016, 48(10):161-168.
[4] Tao Liang, Gaofeng Xie, Dabin Mi, et al. PM2.5 Concentration Forecasting Based on Data Preprocessing Strategy and LSTM Neural Network. 2020, 12(6)
[5] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back propagating errors[J]. Nature, 1986, 323(6088):533-536.
[6] Xiang Li, Ling Peng, Xiaojing Yao, et al. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. 2017, 231:997-1004.
[7] Katja Petrowski, Christina Diane Bastianon, Stefan Bührer, et al. Air Quality and Chronic Stress: A Representative Study of Air Pollution (PM2.5, PM10) in Germany. 2019, 61(2):144-147.
[8] Weize Song, Haifeng Jia, Jingfeng Huang, et al. A satellite-based geographically weighted regression model for regional PM 2.5 estimation over the Pearl River Delta region in China. 2014, 154:1-7.
[9] Congbo Song, Lin Wu, Yaochen Xie, et al. Air pollution in China: Status and spatiotemporal variations. 2017, 227:334-347.
[10] Jianfeng Zhang, Yan Zhu, Xiaoping Zhang, et al. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. 2018, 561:918-929.