Air Quality Prediction Based on Neural Network Model of Long Short-term Memory

Zhehua Du¹*, Xin Lin²

¹Wuhan Second Ship Design and Research Institute, Wuhan, Hubei, 430205, China
²Hubei Province Engineering Consulting Co., LTD., Wuhan, Hubei, 430071, China
*Corresponding author’s e-mail: Jackydhz@163.com

Abstract. Air pollution is a serious environmental problem, which has caused more and more concern all over the world. At present, research focuses on air quality prediction, usually using the following two methods: deterministic method and statistical method. Due to the unreliability of pollutant emission data and incomplete theoretical basis, the prediction accuracy of simulation results of these two methods is low, and these methods do not take into account the problem of time loss. In order to solve the problems of low prediction accuracy, low efficiency, and missing time factors in current air quality prediction research, a simple air quality prediction method LSTM neural network model is proposed. According to the characteristics of time series, the problem of multiple input time variables can be well solved. The datasets in two cities respectively prove that LSTM neural network model can effectively improve the accuracy of air quality prediction. The model can also be used to predict other multivariable input time series. Although air quality prediction based on LSTM presented in this paper has achieved the expected goal, there is still room for improvement in considering both spatial correlation and temporal dimension.

1. Introduction
Air pollution is a serious environmental problem, which has caused more and more concern all over the world. At present, research focuses on air quality prediction, usually using the following two methods: deterministic method and statistical method[1]. Deterministic methods are performed in model-driven manner using theoretical meteorological emission and chemical model. Dynamic data from a limited number of monitoring stations are used to simulate pollutant discharge, migration and diffusion, as well as removal processes. The statistical method simply uses statistical modeling techniques to predict air quality in a data-driven manner. Due to the unreliability of pollutant emission data and incomplete theoretical basis, the prediction accuracy of simulation results of these two methods is low, and these methods do not take into account the problem of time loss.

In this paper, a simple air quality prediction method, LSTM neural network model, is proposed to solve the problems of low accuracy, low efficiency and missing time factors in current air quality prediction research.

2. Overview

2.1. Feature-based technology
Common methods for air quality prediction include gray system, multiple linear regression (MLR) model, hidden markov model (HMMs) and autoregressive moving average (ARMA) model. These
methods are all feature based technology. Bezuglov, A. et al. [2] predicted traffic parameters of short-
term expressway based on grey system theory model. Grey system theory model has the
characteristics of simple structure and low data requirement. Because this method cannot model non-
linear modes and usually can only obtain limited accuracy, they cannot predict extreme air pollutant
concentrations.

2.2. Deep learning based technology
Deep learning algorithm uses multi-layered architecture, which extracts inherent characteristics of data
layer by layer from bottom to top, and can identify representative structure in the data. Deep learning
algorithm can extract representative air quality characteristics without prior knowledge and has good
performance for air quality prediction. LSTM model is a technology based on deep learning.

3. Air quality prediction model

3.1. Recurrent neural network(RNN)
RNN is a very powerful algorithm that can classify, cluster, and predict data, especially time series
and text. RNN can be thought of as MLP network with loops added to the architecture. In Figure 1,
you can see that there are an input layer (containing nodes such as $x_1, x_2$), a hidden layer (containing
nodes such as $h_1, h_2$), and an output layer (containing nodes such as $y_1, y_2$), which is similar to MLP
system structure. The difference is that the nodes of hidden layer are interconnected. In ordinary RNN,
nodes are connected in one direction, which means that $h_2$ depends on $h_1$ and $h_3$ depends on $h_2$. The
nodes in hidden layer are determined by previous node in hidden layer[3]. RNN formula can be
expressed as follows. Among it, $\omega_{xh}$ is matrix parameter input to hidden layer, $\omega_{hh}$ is matrix parameter
input from hidden layer to hidden layer, $b_h$ is the bias parameter of hidden layer, and $\sigma$ can be the
function of sigmoid, tanh or ReLU.

$$h_t = \sigma(\omega_{xh} x_t + \omega_{hh} h_{t-1} + b_h)$$

Figure 1. RNN structure

3.2. Long short-term memory
Because RNN is prone to produce explosive gradient[4] when dealing with long time series, its
correctness is often poor. To solve this problem, LSTM was first introduced by Hochreiter, S. et al. [5]
and reemerged as a successful architecture. LSTM neural network is a variant of RNN structure. Its
main idea is to introduce an adaptive gating mechanism. It determines the extent to which LSTM unit
remains in its previous state. It can also remember the extracted features of current data input.
Although many variations of LSTM are proposed, standard architecture of LSTM is adopted in this
paper to predict air quality. Figure 2 shows a LSTM storage unit with gates.
Generally, LSTM recurrent neural network consists of 4 components. Input gate $i_t$ has corresponding weight matrix $w_{xi}, w_{hi}, w_{ci}, b_i$. Forget gate $f_t$ corresponding weight matrix $w_{xf}, w_{hf}, w_{cf}, b_f$. Output gate $o_t$ has corresponding weight matrix $w_{xo}, w_{ho}, w_{co}, b_o$. The function of input gate is to selectively record new information into cell state. The function of forget gate is to selectively forget information about cell state. The function of output gate is to output some information about the cell. All of these gates are set to use current input $x_t$ to generate a certain degree. The state $h_{t-1}$ generated in previous step and current state (peephole) of this unit $c_{t-1}$ are used to decide whether to accept the input, forget previously stored memory and output state generated later. The specific working process of LSTM is shown in equations. $\sigma$ is a Logistic Sigmoid function with output in $[0,1]$, and $\tanh$ represents a hyperbolic tangent function with output in $[-1,1]$.

$$
\begin{align*}
    i_t &= \sigma(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci}c_{t-1} + b_i) \\
    f_t &= \sigma(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf}c_{t-1} + b_f) \\
    o_t &= \sigma(w_{xo}x_t + w_{ho}h_{t-1} + w_{co}c_{t-1} + b_o) \\
    c_t &= f_t c_{t-1} + i_t \tanh(w_{xc}x_t + w_{hc}h_{t-1} + w_{cc}c_{t-1} + b_c) \\
    h_t &= o_t \tanh(c_t)
\end{align*}
$$

4. Experimental data and metrics

4.1. Data acquisition

Air quality data from PunjabiBagh monitoring station in Delhi from 2014 to 2016 and air quality data from HarrisCounty monitoring station in Houston from 2010 to 2016 were selected. The data are from central pollution control board of India (CPCB) (https://www3.epa.gov/airquality/cleanair.html) and environmental protection agency of United States of America (EPA) (http://cpcb.nic.in/). Delhi’s air pollutants include NO2, CO, O3 and PM10. Air pollutants in Houston include NO2, CO, O3, PM10 and SO2. Air quality (AQI) standards for India and United States of America are listed in Table 1 and Table 2.

| Range | AQI Category          | Range | AQI Category                  |
|-------|-----------------------|-------|-------------------------------|
| 0-50  | Good                  | 0-50  | Good                         |
| 51-100| Satisfactory          | 51-100| Moderate                     |
| 101-200| Moderately Polluted  | 101-200| Unhealthy for Sensitive Groups|
| 201-300| Poor                 | 201-300| Unhealthy                   |
| 301-400| Very Poor            | 301-400| Very Unhealthy              |
| 401-500| Sever                | 401-500| Hazardous                   |
Table 3. Delhi air quality data samples

| Date     | NO2  | CO   | O3  | PM10 | AQI   |
|----------|------|------|-----|------|-------|
| 07/04/2015 | 105.01 | 1.16 | 27.9 | 226.25 | 184.17 |
| 08/04/2015 | 73.37  | 0.94 | 68.31 | 212.23 | 174.82 |
| 09/04/2015 | 84.72  | 1.1  | 73.8  | 148.91 | 132.61 |
| 10/04/2015 | 72.32  | 1.15 | 70.68 | 201.3  | 167.53 |
| 11/04/2015 | 84.64  | 1.12 | 83.36 | 275.81 | 225.81 |

Table 4. Houston air quality data samples

| Date     | CO   | O3  | SO2  | PM2.5  | NO2  | AQI   |
|----------|------|-----|------|--------|------|-------|
| 07/04/2015 | 8    | 31  | 11  | 54    | 36   | 54    |
| 08/04/2015 | 10   | 40  | 21  | 57    | 35   | 57    |
| 09/04/2015 | 5    | 36  | 31  | 58    | 36   | 58    |
| 10/04/2015 | 5    | 31  | 47  | 45    | 25   | 47    |
| 11/04/2015 | 7    | 32  | 49  | 59    | 43   | 59    |

4.2. Correlation coefficient of air quality data

In this experiment, Pearson correlation coefficient (also known as PCCs or PPMCC, denoted by R) in statistics was used to analyze correlation between AQI and each parameter on Houston and Delhi datasets. The larger absolute value of correlation coefficient, the higher correlation between X and Y, as shown in Table 5 and Table 6.

Table 5. Houston correlation coefficient

| Pearson correlation coefficient | R    |
|--------------------------------|------|
| AQI                           | 1    |
| PM2.5                         | 0.768|
| O3                            | 0.608|
| CO                            | 0.180|
| SO2                           | 0.159|
| NO2                           | 0.210|

Table 6. Delhi correlation coefficient

| Pearson correlation coefficient | R    |
|--------------------------------|------|
| AQI                           | 1    |
| PM10                          | 0.991|
| O3                            | 0.185|
| CO                            | 0.413|
| NO2                           | 0.610|

In order to evaluate the performance of each regression model, statistical indicators such as mean absolute error (MAE), mean absolute percentage error (MAPE), correlation coefficient (R), root mean square error (RMSE), and index of agreement (IA) were selected for evaluation.

5. Experiment and result

5.1. Experimental design

The data was partitioned based on LSTM neural network. From 2014 to 2016, 460 data samples of Delhi air quality were selected as training samples, and 49 data samples were trained for 500 rounds. From 2010 to 2016, Houston air quality was selected with 1500 data as training samples and 500 data samples were trained for 500 rounds. In the experiment, the number of hidden layers was set as 2. When LSTM algorithm selected learning rate $l_r = 0.0009$ and $l_r = 0.0006$ on Delhi and Houston data sets, the effect was the best.

5.2. Performance comparison of different prediction models

In order to analyze the accuracy of LSTM in air quality prediction, LSTM algorithm would be compared with some existing methods on the data of Delhi and Houston. The comparison models included MLR (BGD), MLR (SGD), MLR (MBGD) and regression Model (SVR). In this study, $S_{MAPE}$, $S_{RMSE}$, $R$, $I_A$ and $S_{MAE}$ were used as metrics for model comparison. Table 7 and Table 8 show survey performance indicators for all models in Delhi and Houston. It can be seen from two tables that the
values of $S_{MAPE}$, $S_{RMSE}$ and $S_{MAE}$ of LSTM model were far lower than the corresponding values of comparison model.

### Table 7. Experimental results of different models on Delhi

| Model         | $S_{MAE}$ | $S_{MAPE}$ | $R$  | $S_{RMSE}$ | $I_2$ |
|---------------|-----------|------------|------|------------|-------|
| MLR (BGD)     | 10.89     | 5.85       | 0.982| 13.82      | 0.989 |
| MLR (SGD)     | 7.67      | 3.93       | 0.989| 10.68      | 0.991 |
| MLR (MBGD)    | 6.66      | 3.60       | 0.993| 8.80       | 0.994 |
| SVR           | 5.13      | 3.01       | 0.996| 6.20       | 0.998 |
| LSTM          | 3.68      | 1.94       | 0.967| 4.77       | 0.990 |

### Table 8. Experimental results of different models on Houston

| Model         | $S_{MAE}$ | $S_{MAPE}$ | $R$  | $S_{RMSE}$ | $I_2$ |
|---------------|-----------|------------|------|------------|-------|
| MLR (BGD)     | 10.46     | 12.34      | 0.929| 13.06      | 0.960 |
| MLR (SGD)     | 10.28     | 12.11      | 0.931| 12.92      | 0.962 |
| MLR (MBGD)    | 9.11      | 10.13      | 0.931| 10.90      | 0.963 |
| SVR           | 5.91      | 7.16       | 0.979| 7.25       | 0.988 |
| LSTM          | 1.65      | 3.19       | 0.980| 2.75       | 0.990 |

### 6. Conclusion

In order to solve the problems of low prediction accuracy, low efficiency, and missing time factors in current air quality prediction research, a simple air quality prediction method LSTM neural network model is proposed. LSTM neural network model can effectively improve the accuracy of air quality prediction. The other multivariable input time series can be predicted by the model.

Although the expected goal has been achieved, it still needs to be improved considering both spatial correlation and temporal dimension. In the future research, it is necessary to calculate the influence of spatial correlation and temporal dimension on air quality in a certain period of time.

### Acknowledgments

This paper was funded by the National Key R&D Program (Item Number:2017YFC0307800).

### References

[1] Perezv, A., Arasa, R., Codina, B. (2015) Enhancing air quality forecasts over catalonia (Spain) using model output statics. Journal of Geoscience and Environment Protection, 3:9-22.
[2] Bezuglov, A., Comert, G. (2016) Short-term freeway traffic parameter prediction: application of grey system theory models. Expert Systems with Applications, 62:284-292.
[3] Silibello, C., Dallura, A., Finardi, S. (2015) Application of bias adjustment techniques to improve air quality forecasts. Atmospheric Pollution Research, 6:928-938.
[4] Bengio, Y., Simard, P., Frasconi, P. (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5:157-166.
[5] Hochreiter, S., Schmidhuber, J. (1997) Long short-term memory. Neural Computation, 9:1735-1780.