Study of LSTM Air Quality Index Prediction Based on Forecasting Timeliness

Hongna HE¹* and Fei LUO²

¹Software Engineering, Chengdu University of Information Engineering, Chengdu, Sichuan, 610225, China
²Software Engineering, Chengdu University of Information Engineering, Chengdu, Sichuan, 610225, China
*Corresponding author’s e-mail: 1226028456@qq.com

Abstract: The change of urban air quality is affected by pollutant emission, meteorological conditions and other factors, so air quality prediction is a multi-variable, nonlinear and time-series problem, which is difficult to be predicted by traditional methods. To solve this problem, a circular neural network based on Long Short Time Memory (LSTM) is proposed to predict Air Quality Index (AQI) by considering the pollution sources, meteorological conditions and time series. The transfer entropy is used to select the meteorological factors that affect the strong change of AQI. Combined with the prediction time, the prediction accuracy of this algorithm in the future 0~48 hours within different forecast time is studied and compared with the traditional BP neural network and the Gated Recurrent Unit (GRU), and then the Root Mean Square Error (RMSE) was used for evaluation. Taking the measured data of hourly air quality index of Chengdu from January 1, 2018 to September 15, 2019 and the measured data of meteorological factors in the same period as experimental examples, the experimental results show that LSTM has better prediction accuracy and robustness than traditional neural networks in the aging of 0~48h forecasting, and has the advantages of temporary forecasting and short-term forecasting. At the same time, it is verified that GRU has no obvious advantage in air quality index prediction application compared with LSTM.

1. Introduction

In recent years, with the rapid development of the economy, the addition in energy consumption has increased the pressure on urban air quality prevention and control. On the one hand, the number of motor vehicles has been increasing substantially, and the discharge of various pollutants has been increasing year by year, which has result in more serious urban air pollution; on the other hand, under the highly nonlinear chaotic system, the interaction between air quality changes and meteorological conditions is becoming more and more complicated [1]. From time to time, the tragic situation of ten face crouching is staged throughout the country, as shown in Figure 1. Studies have shown that fine particles in smog will increase the risk of cardiovascular disease and lung cancer, which has seriously threatened human health, and efforts to address air pollution is imminent [2]. The Air Quality Index (AQI) is a dimensionless index that quantitatively describes the state of air quality. The higher the value is, the higher the level is, indicating that the air is contaminated more seriously. Therefore, predicting accurately AQI has become one of the important research directions for improving and controlling air quality.
At present, statistical analysis models and numerical prediction models are commonly used in air quality prediction research at domestic and foreign. The numerical forecasting method has higher requirements on time and space distribution data and meteorological patterns, the calculation is complicated, and the applicable range is small. Statistical forecasting methods are simple, economical and easy to implement. Time series analysis, multiple regression models and traditional neural network models are commonly used. Mou J F. et al [3] achieved good results in predicting AQI in Shenzhen in the next three to five days by establishing an Autoregressive Integrated Moving Average model (ARIMA(3,0,1)), but once the forecasting days was increased, the prediction accuracy will be reduced and the model overemphasizes the role of time factors. Yin Q. et al [4] combined Genetic Algorithms (GA) with Support Vector Machines (SVM) to predict AQI in Taiyuan, the results proved that the GA-SVM model had advantages in the accuracy and the training speed, but the model only was applicable to small sample size, which had limitations. Zhang P D. et al [5] established a BP air network based urban air quality prediction model, the results showed that the model had higher prediction accuracy, but BP neural network was prone to over-fitting and local minimization. After that, Yang Y. et al [6] proposed a prediction method based on Genetic Algorithm optimization BP neural network (GA-BP), the results showed that GA-BP model was better than BP neural network in generalization ability, and solved BP neural network localization minimized problem. Sun M [7] verified that the deep learning models had better predictive ability than the traditional neural network models and statistical models in the context of big data. Zhang L C. et al [8] used Tensorflow as the development platform and used the AQI daily average data of Taiyuan City to establish the LSTM prediction model, which verified the feasibility of the method and provided a new idea for AQI research, but did not consider wind speed, rainfall, etc., just used historical data to simply simulate predictions.

This paper considers that the same model has different forecasting abilities in different forecasting aging. For meteorology, the forecast of 0~2h is called temporary forecast, the forecast of 0~12h is called short-term forecast, the forecast within three days is called short-term forecast. Therefore, based on forecasting aging and meteorological factors, this paper proposes a LSTM-based AQI prediction model, using Chengdu hourly air quality and the same measured meteorological factors for simulation experiments, aiming to study the AQI prediction ability of LSTM model in different aging in the future 0~48h.

2. Transfer entropy

According to the new ambient of air quality standard (GB3095-2012) issued in the second half of 2012, the pollutants involved in the assessment are carbon monoxide (CO), respirable particulate matter (PM\(_{10}\)), sulfur dioxide (SO\(_2\)), and nitrogen dioxide (NO\(_2\)), ozone (O\(_3\)), fine particles (PM\(_{2.5}\)). In the case of stable urban sources, meteorological factors are the main factors which affect AQI. So the selection of meteorological factors is very necessary. Schreiber T. et al [9] proposed information transfer entropy in 2000, which not only quantified the degree of correlation between factors, but also described the direction of information transmission. Transfer entropy is a common and effective method for dealing with relationships between nonlinear systems in information theory. Therefore, this paper proposes transfer entropy to select meteorological factors that have a strong impact on AQI changes. Assuming that \(x_n = \{x_n, x_{n-1}, x_{n-2}, \ldots, x_{n-k+1}\} \) and \(y_n = \{y_n, y_{n-1}, y_{n-2}, \ldots, y_{n-t+1}\} \) are the random time series of the samples x, y respectively, Schreiber defined the formula as follows:
\[ T_{X \rightarrow Y} = \sum_{y_{n+1}} p(y_{n+1}, x_n^{k}, y_n^{l}) \log \left( \frac{p(y_{n+1} \mid x_n^{k}, y_n^{l})}{p(y_{n+1} \mid y_n^{l})} \right) \]

\[ T_{Y \rightarrow X} = \sum_{x_{n+1}} p(x_{n+1}, x_n^{k}, y_n^{l}) \log \left( \frac{p(x_{n+1} \mid x_n^{k}, y_n^{l})}{p(x_{n+1} \mid x_n^{k})} \right) \] (1)

\[ T_{X \rightarrow Y} \] in Equation (1) is the information that transfer entropy from state \( n \) to state \( n+1 \) under the influence of time series values of \( x_n \). The larger the value is, the greater the amount of information transmitted is. If the value is zero, it means that no information from \( x_n \) is transmitted to \( y_n \), so \( x_n \) is not a factor that affects the change of \( y_n \). \( p(y_{n+1}, x_n^{k}, y_n^{l}) \) is the transition probability from the state \((x_n^{k}, y_n^{l})\) to the state \( y_{n+1} \). \( p(y_{n+1} \mid x_n^{k}, y_n^{l}) \) is the conditional probability to the state \( y_{n+1} \) in the state \((x_n^{k}, y_n^{l})\). \( p(y_{n+1} \mid y_n^{l}) \) is the conditional probability to the state \( y_{n+1} \) in the state of \( y_n^{l} \).

It is worth noting that the delay time selected in Equation (1) is one, but the interaction between them may be longer. Therefore, Equation (1) can be written as Equation (2):

\[ T_{X \rightarrow Y} = \sum_{y_{n+u}} p(y_{n+u}, x_n^{k}, y_n^{l}) \log \left( \frac{p(y_{n+u} \mid x_n^{k}, y_n^{l})}{p(y_{n+u} \mid y_n^{l})} \right) \]

\[ T_{Y \rightarrow X} = \sum_{x_{n+u}} p(x_{n+u}, x_n^{k}, y_n^{l}) \log \left( \frac{p(x_{n+u} \mid x_n^{k}, y_n^{l})}{p(x_{n+u} \mid x_n^{k})} \right) \] (2)

\( u \) represents the delay time of the interaction between \( x_n = \{x_n, x_{n-1}, x_{n-2}, \ldots, x_{n-k+1}\} \) and \( y_n^{l} = \{y_n, y_{n-1}, y_{n-2}, \ldots, y_{n-u}\} \).

In this paper, \( x_n \) is the time series value of a certain meteorological factor, \( y_n \) is the time series value of the AQI, and AQI is the reference indicator. If the information transfer entropy value is negative, it means negative correlation, and if it is positive, which means positive correlation. The selection criterion is that if the absolute value of the information transfer entropy value is more than 0.01, the factor is regarded as the key meteorological factor, otherwise, the meteorological factor is weakly related to AQI[10], as shown in Table 1:

| Reference indicator | Key meteorological factor | Information transfer entropy |
|---------------------|---------------------------|------------------------------|
| AQI | air pressure | 0.027 |
| | temperature | 0.033 |
| | wind speed | 0.060 |
| | cloud volume | 0.042 |
| | relative humidity | 0.038 |
| | precipitation | 0.030 |

As shown in Table 1, wind speed is the most important correlation factor that affects AQI, followed by five factors of cloud volume, air pressure, temperature, relative humidity and precipitation. Therefore, this paper selects 12 predictive factors of CO, PM₁₀, SO₂, NO₂, O₃, PM₂.₅, wind speed, cloud volume, air pressure, temperature, relative humidity and precipitation as the input characteristics of AQI prediction model.
3. AQI prediction model based on LSTM

3.1 LSTM introduction

The traditional Recurrent Neural Network (RNN) tends to disappear or even explodes after multiple cycles. And when the depth of the model increases, the information of the subsequent nodes will become weak or even forgotten [11]. To solve the problem of traditional RNN, Hochreiter S. et al [12] proposed the Long Short Term Memory (LSTM) algorithm in 1997, as shown in Figure 2, there are three types of control gates in the LSTM: forget gate, input gate, and output gate. In recent years, the main application of LSTM is in text classification, speech recognition [13], machine translation [14], information retrieval and other fields. Subsequently, Cho K. et al [15] proposed a Gated Recurrent Unit (GRU), which is simpler than the LSTM model after the combination of two control gates and other modifications, which is a popular variant of LSTM. Lei T. et al [16] proposed a Simple Recurrent Unit (SRU) to improve the training speed of the model, which is another variant of LSTM.

$$\sigma = \frac{1}{1 + e^{-z}}$$

Figure 2. LSTM memory cell structure

The forget gate ($f_t$), the input gate ($i_t$), the output gate ($o_t$), the unit state ($c_t$), and the current unit output ($h_t$) calculation formulas are as follows (3) to (8):

$$f_t = \sigma(W_f h_{t-1} + W_f x_t + b_f)$$  \hspace{1cm} (3)
$$i_t = \sigma(W_i h_{t-1} + W_i x_t + b_i)$$  \hspace{1cm} (4)
$$c_t^\prime = \tanh(W_c h_{t-1} + W_c x_t + b_c)$$  \hspace{1cm} (5)
$$o_t = \sigma(W_o h_{t-1} + W_o x_t + b_o)$$  \hspace{1cm} (6)
$$c_t = f_t * c_{t-1} + i_t * c_t^\prime$$  \hspace{1cm} (7)
$$h_t = o_t * \tanh(c_t)$$  \hspace{1cm} (8)

In Equations, symbol $\sigma$ denotes the sigmoid function, which is a commonly used nonlinear activation function. There are three sigmoid functions in the figure to control the forgetting gate, input gate and output gate respectively, and the output value is zero to one in the interval, describing how much information is passed, zero means no information is passed, one means all information passes; tanh means activation function, used in the state and output of the data, the output value is negative one to one in the interval; $x_t$ represents the input at time $t$; $W_f$, $W_i$, $W_c$, $W_o$ are the weight matrix; $b_f$, $b_i$, $b_c$, $b_o$ are the corresponding offset vectors; $h_t$ stores the useful information at the current time and each time before, it is a state vector.

3.2 Building of model
Forecast results of different aging in Chengdu within 0-48 hours in the future

Realizing the input of the main influence factor

Data input layer

Filtering before the data entry model

Data preprocessing

Null value filling

Data correction

Data alignment

Data normalization

Result output layer

Converting model output to actual air quality index

Air quality index prediction model

Model for achieving air quality index prediction

Dividing different timeliness

Generating time series

LSTM

CO, PM_{10}, SO_{2}, NO_{2}, O_{3}, PM_{2.5}

Wind speed, cloud volume, air pressure, temperature, relative humidity, precipitation

Figure 3. LSTM-based AQI prediction model overall architecture

Input layer, hidden layer and output layer are the parts of LSTM which is based on AQI prediction model. Contaminants may be diluted or precipitated over time under the influence of atmospheric conditions. Therefore, this paper is based on the total of 12 features as the input of the current contaminant concentrations (CO, PM_{10}, SO_{2}, NO_{2}, O_{3}, PM_{2.5}) and the next meteorological conditions (precipitation, relative humidity, temperature, cloud volume, wind speed, air pressure); the AQI at the following moment is the output; hidden layer is the point. Useful information will be transferred by the memory unit of the hidden layer all the time like the conveyor belt, and it will integrate the new information, update the memory, and form a new memory at the same time, as shown in Figure 3.

3.3 Preprocessing of date
In the data preprocessing stage, the outliers are eliminated, and the forward replenishment method is used for the missing value processing[17], and then, the data is normalized, the value range is zero to one in the interval. According to different forecasting timeliness, the air quality index data and meteorological data with time series properties are translated correspondingly, so that the original data set is converted into a new one, which takes the pollutant concentration and the meteorological factor as input, and the AQI value as output.

4. Analysis of experiment

4.1 Sources of experimental data
The experimental data mainly includes air quality index data and meteorological factor data measured in the same period. The air quality index data is measured hourly from January 1, 2018 to September 15, 2019 in Chengdu, and it is from the Chengdu Environmental Protection Department. The measured meteorological data at the same period is derived from meteorological center of Leshan City.

4.2 Design of experiment
This paper sets air quality index data hourly from January 1, 2018 to September 15, 2019 as the analysis object, and is based on LSTM for the deep learning model. Developing deep neural network prediction model and making prediction of AQI are based on the Keras platform as a tool and Python3.5 as the development environment.
After data preprocessing, there are 9100 valid data, about 1800 data behind the data set is selected as the test set, and the rest of the data is used as the training set. In the design experiment process, considering the different forecasting ability of the same model in different aging, eight different forecasting aging will be set to the future 3h, 6h, 9h, 12h, 15h, 18h, 21h, 24h within 0~24h; LSTM, GRU and BP neural network were used to establish training models for different forecasting aging; RMSE was used as error evaluation index, and multi-angle analysis was used to compare the three models with forecasting aging and RMSE relationship. The following uses the LSTM model to predict the AQI value as an example. The parameters are configured as follows:

- Using the simple data model sequential in the Keras series, and adding the add() function to linearly stack multiple network layers;
- Setting hidden layer parameters. The hidden layer parameter units is 64, the activation function is relu, establishing the full connection layer between the hidden layer and the output layer, and setting the activation function activation is relu, and discard rate of the network node of each layer is 0.5 to avoid over fitting. The loss function used in this paper is the Mean Square Error (MSE), and the optimizer uses the Root Mean Square Propagation (RMSProp);
- Training model. The samples number of batch size is 32 for each training and the number of training epochs is 60;
- Conducting the prediction. It needs to be verified on the test set after the model training. In this paper, the evaluation index of the prediction error is the Root Mean Square Error (RMSE).

### 4.3 Results and analysis of experiment

In this paper, the error evaluation index of the prediction model is RMSE. The range of RMSE is zero to positive infinity, and the smaller the value is, the more accurate the prediction result is. After repeated experiments, multiple iterations and updated parameters, the prediction error evaluation index values of the three models in the future 0~24h are shown in Table 2:

| Aging (h) | BP neural network | LSTM | GRU |
|----------|------------------|------|-----|
|          | RMSE             | RMSE | RMSE |
| Future 3 | 4.006 661 91     | 3.494 456 3 | 3.566 449 |
| Future 6 | 6.827 347 65     | 3.634 891 5 | 4.119 993 |
| Future 9 | 7.155 493 37     | 4.147 745 3 | 4.398 331 6 |
| Future 12| 6.745 614 12     | 3.972 398 | 4.295 193 7 |
| Future 15| 6.078 987 44     | 4.014 598 | 4.114 598 |
| Future 18| 5.911 444 1      | 4.570 075 7 | 3.570 075 7 |
| Future 21| 6.194 001 22     | 4.579 78 | 4.066 78 |
| Future 24| 6.126 808 78     | 4.659 00 | 4.579 00 |

As shown in Table 2, firstly, in the next 0~24 hours, regardless of any aging, the RMSE of the BP neural network model is higher than the LSTM and GRU models. Therefore, the AQI prediction values of LSTM and GRU are closer to the actual value than the AQI prediction of the BP neural network model, which have higher accuracy; secondly, in the next 0~24 hours, regardless of any aging time, the performance of the GRU model is comparable to that of LSTM, indicating that although GRU is an improved model of LSTM, there is no obvious advantage in air quality index prediction applications; thirdly, in the next 0~12h, the RMSE of the BP neural network model is large and unstable, and tends to be stable within 12~24h, while RMSE of the LSTM and GRU is smaller and stable in the next 0~24 hours, it is verified that the traditional BP neural network is not suitable for temporary forecasting and short-term forecasting of AQI, while LSTM and GRU have obvious advantages in temporary forecasting and short-term forecasting of AQI; finally, It is calculated that within 12~24h, the accuracy of GRU is 32.12% higher than that of BP neural network model, while the accuracy of LSTM is 33.89% higher than that of BP neural network.
In order to further study the trend of forecasting ability of each model with the increase of forecasting time, this paper will gradually increase the forecasting time to 48. As shown in Table 3, eight different forecasting aging will be set as the future 27h, 30h, 33h, 36h, 39h, 42h, 45h, and 48h.

| Aging (h) | BP neural network | LSTM | GRU |
|----------|------------------|------|-----|
| Future 27 | 5.937 666 73     | 4.180 180 5 | 4.331 722 |
| Future 30 | 5.662 315 41     | 3.612 169 3 | 4.541 446 |
| Future 33 | 5.865 657 54     | 3.854 232    | 3.931 296 |
| Future 36 | 6.297 727 95     | 3.804 231 6 | 3.705 08 |
| Future 39 | 6.198 381 86     | 4.136 973    | 4.022 720 3 |
| Future 42 | 5.822 703 59     | 3.689 630 5 | 4.263 976 |
| Future 45 | 5.751 815 96     | 4.187 246    | 4.143 066 4 |
| Future 48 | 5.818 404 13     | 4.003 877 5 | 4.180 079 |

As shown in Table 3, within 24~48h, regardless of any aging, the RMSE of the BP neural network model is also higher than the LSTM and GRU models, but the forecasting ability tends to be more stable, and the forecasting ability of LSTM and GRU have not changed significantly. Similarly, with the increase in forecasting timeliness, GRU still has no clear advantage in air quality index prediction. In summary, in the next 0~48h, LSTM has better forecasting accuracy and robustness than traditional neural networks with increasing aging, and has the advantages of temporary forecasting and short-term forecasting. At the same time, it is verified that GRU has no obvious advantage compared with LSTM in air quality index prediction.

5. Conclusion
This paper adapts air quality data hourly for that it is more accurate to reflect the true state of urban air quality. First of all, selecting six meteorological factors that have significant effects on air quality changes through the transfer entropy. Then, combined with forecasting aging, based on LSTM, GRU and BP neural network model, the prediction ability of three models in different forecasting aging in the next 0~48h is studied, and RMSE is used as the error evaluation index for comparative analysis. The results show that with the increasing aging, LSTM has better prediction accuracy and robustness than traditional neural network, and has the advantages of temporary forecasting and short-term forecasting. At the same time, it is verified that GRU has no obvious advantage in air quality index prediction application compared with LSTM. Air quality monitoring started late in China, and the air quality was affected by various factors such as pollution source, traffic flow, underlying surface and meteorological environment, which made the AQI prediction inaccurate and the prediction difficulty increased. The next research direction can combine this method with numerical simulation prediction methods to make air quality prediction more accurate, which may provide reference for urban air pollution assessment and treatment. There are two shortcomings in this paper. Firstly, the paper only makes the experimental results within the forecast period of 0~48h in the future. If the forecasting time is increased or the step size of the forecasting time is increased, the forecasting ability of each model may be different. Secondly, this paper only considers meteorological factors and pollutant factors, and future researches can take into account other factors such as urban economy and traffic flow.

Acknowledgments
This work was supported by Sichuan science and technology plan project (No. 2018GZDZX0049). In the whole research process of the thesis, firstly, I would like to thank my teacher, Luo Fei, for helping me, she has helped me a lot in the direction, ideas and technology of the thesis, and has been tracking the progress of the research. Secondly, I would like to thank my classmate Liu Yan for giving me a lot of suggestions, which make me spend less time, and she always gives me encouragement and support in my life. Finally, I would like to thank another classmate, Chen Jun, who often recommends me some
advanced technologies, which indirectly open up my horizons and greatly help my accumulation of professional knowledge. Thanks again.

References
[1] Yang, J.Y. (2018) Analysis of air pollution in China [J]. Resource Conservation and Environmental Protection. 203(10):76.
[2] Qi, H.D. (2013) Fine particulate matter pollution and residents health under smog weather [J]. Chinese Journal of Preventive Medicine. 47(6): 491-493.
[3] Mou, J.F., Zhao, X., Fan, J.J., et al. (2017) Prediction of air quality index time series in Shenzhen based on ARIMA model [J]. Journal of Environmental Hygiene. 2:102-107.
[4] Yin, Q., Hu, H.P., Bai, Y.P., et al. (2017) Prediction of Tai Yuan air quality index based on GA-SVM [J]. Mathematical Practice and Understanding.
[5] Zhang, P.D. (2014) Urban environmental air quality prediction model based on BP neural network [J]. Automation Technology and Application. 33(1): 9-11.
[6] Yang, Y., Yang, Y. (2016) Prediction and analysis of air quality index in heating period based on GA-BP neural network [J]. Journal of Shaanxi University of Science and Technology. 34(4): 171-176.
[7] Sun, M. (2018) Research on air quality index prediction model based on optimized deep confidence network [D].
[8] Zhang, C.L., Bai, Y.P. (2018) Application of tensorflow based LSTM model to air quality AQI index prediction in Taiyuan [J]. Journal of Chongqing University of Technology (Natural Science). v.32; No. 386 (8): 143-147.
[9] Schreiber, T. (2000) Measuring Information Transfer [J]. Physical Review Letters. 85(2):461.
[10] Wen, Q., Luo, F. (2017) Application of entropy in air quality index (AQI) prediction [J]. Journal of Sichuan Institute of Technology (Natural Science Edition).
[11] Bai, S.N., Shen, X.L. (2019) Prediction of PM (2.5) based on LSTM circular neural network [J]. Computer Applications and Software. 36(01):73-76+110.
[12] Hochreiter, S., Schmidhuber, J. (1997) Long short-term memory [J]. Neural Computation. 9(8) : 1735－1780.
[13] Graves, A., Jaitly, N. (2014) Towards end-to-end speech recognition with recurrent neural networks [D]/Proceedings of the 31st International Conference on Machine Learning. Beijing: JMLR. 1764－1772.
[14] Wu, Y., Schuster, M., Chen, Z. et al. (2016) Google's neural machine translation system: Bridging the gap between human and machine translation [J]. arXiv Preprint. 2016: arXiv: 1609.08144.
[15] Cho, K., Merrienboer, B., Gulcehre, C., et al. (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. arXiv Preprint, 2014: arXiv: 1406.1078.
[16] Lei, T., Zhang, Y., Airtzi, Y. (2017) Training RNNs as fast as CNNs [J]. arXiv Preprint, 2017: arXiv: 1709.02755.
[17] Fan, J.X., Li, Q., Zhu, Y.J., et al. (2017) The space-time air pollution forecast model based on RNN study [J]. Journal of Surveying and Mapping Science. 7: 80-87 + 124.