Environment, Development and Sustainability
https://doi.org/10.1007/s10668-021-02102-8

An air quality prediction model based on CNN-BiNLSTM-attention

Jingyang Wang[1] · Jiazheng Li[1] · Xiaoxiao Wang[1] · Tingting Wang[1] · Qiuhong Sun[1]

Received: 26 September 2021 / Accepted: 28 December 2021

Abstract
In recent years, the air pollution problem has been aggravated, which has brought some problems to people's production and life. A simple mathematical model cannot accurately predict air quality because of the characteristic of air quality volatility and obvious nonlinear characteristics. Therefore, this paper proposes a CNN-BiNLSTM-Attention air quality prediction model to forecast the AQI in the next hour. In the model, convolutional neural networks (CNN) are used to extract characteristics from the input air quality data and meteorological data. NLSTM is an improvement on long short-term memory (LSTM) to make output value range of forget gate more accurate, thus preserving more characteristics of the data. BiNLSTM is used to predict the time series data. Attention mechanism (Attention) is used to capture the effect of characteristic conditions at imparity times on AQI prediction. To prove the accuracy of the model, CNN-BiNLSTM-Attention, recurrent neural network (RNN), LSTM, ARIMA, BiLSTM, CNN-BiLSTM, and CNN-BiLSTM-Attention are used to forecast the hourly AQI from 00:00 on January 1, 2020, to 23:00 on September 30, 2020, in Shijiazhuang, Hebei Province in this paper. The results show that the MAE of CNN-BiNLSTM-Attention is 5.987 and the RMSE is 9.231. They are the smallest. R^2 is 0.9741, which is the closest to 1. The CNN-BiNLSTM-Attention air quality prediction model is more suitable to predict air quality, informs people in advance of air pollution who can take corresponding measures to reduce air pollution.

✉ Qiuhong Sun
sunqiuhong@hebust.edu.cn

Jingyang Wang
jingyangw@hebust.edu.cn

Jiazheng Li
69880353@qq.com

Xiaoxiao Wang
1772732517@qq.com

Tingting Wang
3509922875@qq.com

[1] School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China

Keywords Air quality · Prediction · CNN · BiNLSTM · Attention

1 Introduction

For the past few years, the process of urban industrialization is gradually accelerating, and the problem of air pollution has also followed (Jin et al., 2020). China's air pollution has been widely concerned, especially in the north of China, such as Hebei Province and Beijing, haze events often occur. The air quality in Shijiazhuang, Hebei province, is worrisomely polluted. From 2017 to 2020, Shijiazhuang had only 724 days with good air quality, accounting for 49.56 percent of the total. There were 437 days with mild pollution, 159 days with moderate pollution, and 141 days with severe pollution or above. The air pollution problem is very serious. Air pollutants mainly include gas pollutants such as CO, SO_2, O_3, NO_2, and particulate pollutants such as PM10 and PM2.5. These pollutants can have a potential impact on people's lives and even lead to physical illness (White, 1993).

Air quality index (AQI) is an indicator to evaluate air quality. It is a numerical index formed by calculation according to environmental quality standards and the concentration of pollutants. AQI can scientifically and intuitively reflect the degree of air pollution. Various pollution concentrations in the air are affected by meteorological factors and man-made factors. All these factors can cause the change of the pollutant concentration, thus affecting the change of AQI. The emergence of these factors increases the difficulty of air quality prediction.

Because of the increasingly serious air pollution problem, people have begun to gradually realize the harm of air pollution. The awareness of environmental protection has been gradually improved, and the call for improving air quality is also growing (Xing et al., 2021). Therefore, the establishment of an efficient and accurate air quality prediction model has a very important guiding significance for air pollution control and air quality improvement.

A CNN-BiNLSTM-Attention air quality prediction model is proposed in this paper to improve the accuracy of air quality prediction. The model uses air quality data and meteorological data to predict the AQI in the next hour. The model consists of three parts: CNN, BiNLSTM, and Attention. CNN is used to extract the characteristics of input data, and to extract the relevant information that affects AQI. The improved LSTM constitutes NLSTM, which makes the output value of the forget gate more accurate and retains more data characteristic information. BiNLSTM is used to predict the time series data. Attention is used to capture the effect of characteristic conditions at imparity times on AQI prediction, so as to predict the AQI of the next hour more accurately.

The contribution of this paper includes the following four points:

1. LSTM is improved by introducing 1-tanh function after the forget gate so that it can retain as many data characteristics as possible for the time-series data, and make the prediction more accurate.
2. By analyzing the time series data, a new model (CNN-BiNLSTM-Attention) for predicting air quality is proposed to predict the AQI in the next hour.
3. According to the analysis of time series, it is proposed that CNN is used as the characteristic extraction of time series data before time series prediction. Attention is used to

analyze the effect of the characteristic conditions of the past at imparity times on the AQI of the next hour. The more accurate AQI is obtained by weighting calculation.

4. This model is compared with five other models to predict air quality. The accuracy of this model is verified by predicting the hourly AQI, from January 1, 2020, to September 30, 2020, in Shijiazhuang City, Hebei Province, which shows that it is more suitable for air quality prediction.

The rest of this paper is organized as follows: Section II. introduces some previous air quality prediction models. Section III. describes the principle of CNN-BiNLSTM-Attention proposed in this paper, model training process and model prediction process; Section IV. introduces experimental data, dataset, data preprocessing, experimental parameter settings, experimental results, and discussion; Section V. summarizes the work of this paper and introduces further research.

2 Related works

Air quality prediction has always been a significant issue studied by scholars. At present, many scholars use traditional statistical methods (such as the method of mathematical statistics, autoregressive moving average (ARMA), grey system, hidden Markov model, multiple linear regression model, et al. (Wang & Wang, 2019; Gu et al., 2020; Zhao et al., 2019)) or machine learning methods (such as artificial neural network (ANN), support vector machine (SVM), et al. (Koo et al., 2020; Li et al., 2019)) to forecast air quality.

At the beginning of the study of air quality, scholars mostly used statistical models to predict air quality. Pagowski et al. (2006) used a dynamic linear regression model to predict O_3. Pai et al. (2010) used a multiple linear regression method to predict O_3, but the effect was not good. Sayegh et al. (2014) used statistical models to forecast PM10, indicating that statistical methods can predict air quality to a certain extent. Li et al. (2016) used wavelet decomposition and ARMA to predict the 2014 PM2.5 of Tianjin. Niu et al. (2016) used ARMA to forecast the air quality in Chengdu. However, for the obvious characteristics of non-linearity and uncertainty of air quality, the traditional statistical model is difficult to predict accurately.

For the past few years, increasingly scholars have found that air quality data has obvious nonlinear characteristics. They gradually begin to use machine learning to forecast air quality. The machine learning methods can effectively learn the characteristics of lots of input data, which provides a new research idea to forecast air quality. Caselli et al. (2009) used ANN to predict PM10. The experimental results indicated that the prediction accuracy of ANN was higher than that of the multivariate linear regression model. Mishra et al. (2015) used principal components analysis (PCA) and multi-layer perceptron (MLP) to predict NO_2. The experimental results indicated that PCA-MLP could better forecast the air pollution of the Agra Taj Mahal. Ong et al. (2016) used RNN to forecast PM2.5. The experimental results indicated that RNN could better deal with the problem of air quality prediction. Peng et al. (2017) used MLP to forecast hourly air quality in Canada. The experimental results indicated that nonlinear model was more accurate to forecast air quality. Nieto et al. (2018) used SVM to predict the PM10 of Oviedo (Northern Spain) and compared this method with MLP, autoregressive integrated moving average, vector autoregressive moving average. The experimental results indicated that the prediction of SVM was better. Krishan et al. (2019) used LSTM to predict the air quality in New Delhi, India. The experimental results indicated that LSTM could effectively forecast air quality. Ma et al. (2019) used BiLSTM to forecast PM2.5 of Guangdong Province. Compared with other models, it proved that this method had better performance for PM2.5 forecasting. Wang et al. (2019) used a double-layer RNN composed of LSTM and gated recurrent unit networks (GRU) to forecast PM2.5 in four cities. The experimental results indicated that the model could forecast PM2.5 better. Li et al. (2020) used CNN-LSTM to forecast the PM2.5. The experimental results indicated that the hybrid model was more accuracy in predicting PM2.5 and needed shorter training time. Wang, Yuan, et al. (2021) used BP neural network to predict PM2.5 in Chongqing. The experimental results showed that PM2.5 was related to meteorological factors. It was feasible to establish BP neural network to predict PM2.5 by using meteorological data.

3 Models

3.1 CNN

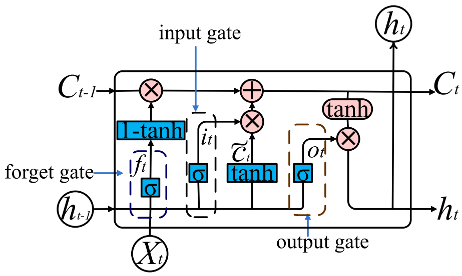
CNN is proposed by Lecun et al. (1998). CNN is first used to detect medical images. It has the characteristics of local perception, weight sharing, and ignoring useless information. CNN can reduce parameters and accelerate the learning efficiency in the neural network (Yang et al., 2020). The hidden layer of the classical CNN has convolution layer, pooling layer, and full connection layer (Eslami et al., 2020). CNN uses a special method to deal with the input data, and the main purpose of the convolution process is to extract the input eigenvalue points (Huang & Kuo, 2018). The convolution layer has multiple convolution cores. The main goal of the pooling layer is to sample the output vector of the convolution layer, reduce the vector dimension flowed in the network. It can also reduce the dimension of the flow vector of the model and parameters in the hidden layer, and reduce the computational complexity (Sayeed et al., 2020).

3.2 NLSTM

LSTM is proposed by Hochreiter et al. (1997). The problem of gradient explosion and disappearance exists in RNN. To resolve these problems, RNN is modified to maintain long-term information preservation by adding an internal gate control mechanism, resulting in LSTM (Liu et al., 2020; Xayasouk et al., 2020). Compared with RNN, LSTM can perform better in longer time series (Fong et al., 2020). The structure of LSTM is more complex. LSTM is composed of forget gate, input gate, and output gate. The forget gate is used to selectively forget cell state. The input gate is used to selectively save the new information to the cell state. The output door is used to output the final value according to the cell state (Freeman et al., 2018).

NLSTM is an improvement over LSTM. The output value of forget gate activated by the Sigmoid function in the LSTM ranges from 0 to 1. If the output value approach 0, the input information would be discarded. If the output value approach 1, the input information would be fully transmitted, and there is a situation of over-saturation of the output value. However, by introducing the 1-tanh function behind the forget gate, the output value can be changed to [0.25,1]. The output value of forget gate can be in a more obvious range. Therefore, NLSTM can preserve more characteristics of the input data, and improve the learning ability of NLSTM. The structure of NLSTM is shown in Fig. 1.

Fig. 1 Architecture of NLSTM memory cell



The calculation process of NLSTM has five steps:

1. The h_{t-1} and x_t are inputted into three gates. The corresponding output value is calculated by formula (1–3).

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{1}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{3}$$

2. The output value of forget gate is changed by formula (4).

$$n_t = 1 - \tanh\left(f_t\right) \tag{4}$$

3. The candidate cell state is calculated by formula (5).

$$\widetilde{C}_t = \sigma\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{5}$$

4. The cell status is updated by formula (6).

$$C_t = n_t * C_{t-1} + i_t * \widetilde{C}_t \tag{6}$$

5. The output of NLSTM is calculated by formula (7).

$$h_t = o_t * \tanh\left(C_t\right) \tag{7}$$

where W_f, W_i, W_c and W_o are the weight, b_f, b_i, b_c and b_o are the bias, f_t is the output of forget gate, i_t is the output of input gate, o_t is the output of output gate, $\widetilde{C}_t$ is the candidate cell state, C_t is the cell state.

3.3 Attention

Attention is proposed by Treisman et al. (1980). Attention mainly mimics human vision. Human vision can find important areas and add the focus of attention to the important areas to get the details of needed information. Attention can discover the characteristics of data at different moments in time series. Therefore, important information is selected from all the information by Attention. the important information is paid more attention.

The irrelevant information is ignored to avoid the interference of the irrelevant information to the final result, so as to optimize the traditional model.

The calculation process of Attention as follows: Firstly, the score s_t of the input vector k_t is calculated, the effect degree of input vector k_t to output value can be obtained. Secondly, the score s_t is normalized by the Softmax function, weight coefficient a_t is obtained. Finally, the weighted vector o_t is calculated by weight coefficient a_t and input vector k_t. The calculation formulas are shown in formula (8)-(10).

$$s_t = \tanh\left(W_h k_t + b_h\right) \tag{8}$$

$$a_t = softmax\left(s_t\right) \tag{9}$$

$$o_t = \sum_t a_t k_t \tag{10}$$

where W_h is the weight, b_h is the bias.

3.4 CNN-BiNLSTM-attention structure

CNN can find evident characteristics in the data, so CNN is used to extract characteristics. The basic idea of BiNLSTM is to input the time series into two layers of NLSTM in forward and reverse directions respectively, and finally get the output value according to the output of the two layers of NLSTM. BiNLSTM is used in time series due to its good characteristics of calculation based on the sequence. Attention can capture the importance of the characteristic conditions at imparity times on time series prediction, and it is often applied after RNN. Therefore, the use of attention after BiNLSTM calculation can focus on the characteristics that affect the results, to improve the prediction accuracy. Based on the characteristics of CNN, BiNLSTM, and Attention,

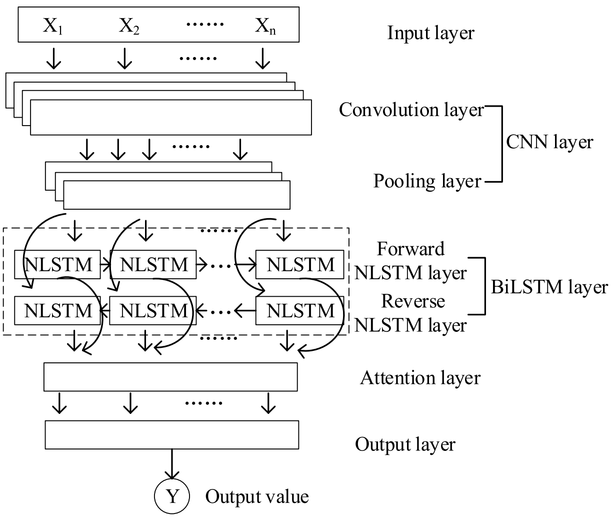


Fig. 2 CNN-BiNLSTM-Attention structure diagram

a CNN-BiNLSTM-Attention air quality prediction model is set up. CNN-BiNLSTM-Attention structure is shown in Fig. 2. It includes input layer, CNN layer, BiNLSTM layer, attention layer, and output layer.

4 Experiments

To prove the accuracy of CNN-BiNLSTM-Attention in predicting air quality, this paper uses the same data for experiments. All experiments are carried out on the same computer. A comparative experiment is carried out between CNN-BiNLSTM-Attention and RNN, LSTM, ARIMA, BiLSTM, CNN-BiLSTM, CNN-BiLSTM-Attention to predict the hourly AQI from 0:00 on January 1, 2020, to 23:00 on September 30, 2020, in Shijiazhuang City, Hebei Province. All models are written in Python 3.7.3 and Keras 2.1.0. To assess the prediction accuracy of CNN-BiNLSTM-Attention, MAE, RMSE, and R^2 are used as the evaluation indexes of the models. The smaller the MAE, RMSE, the better the model is, and the bigger R^2, the better the model is.

The MAE calculation formula is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{y}_i - y_i\right| \tag{11}$$

where $\widehat{y}_i$ is the predicted value and y_i is the real value.

The RMSE calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y}_i - y_i\right)^2} \tag{12}$$

where $\widehat{y}_i$ is the predicted value and y_i is the real value.

The R^2 calculation formula is as follows:

$$\mathrm{R}^2 = 1 - \frac{\left(\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2\right)/n}{\left(\sum_{i=1}^{n}\left(\overline{y_i} - \widehat{y}_i\right)^2\right)/n} \tag{13}$$

where $\widehat{y}_i$ is the predicted value, y_i is the real value, and $\overline{y_i}$ is the average value.

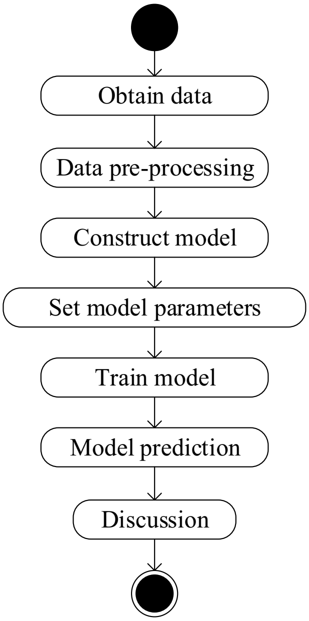
4.1 Experimental process

The experimental process is shown in Fig. 3.

The experimental process has seven steps:

6. Obtain data. Air quality data and weather data required by the experiment are obtained from corresponding websites.
7. Data pre-processing. The final experimental data is obtained by data preprocessing.
8. Construct model. The models are constructed using Python based on the model structure.

Fig. 3 The experimental process diagram



9. Set model parameters. According to the experimental data, the parameter values of models are set.
10. Train model. After setting parameters, the models are trained using training set data. The optimal model of each model is saved after the training.
11. Model prediction. The test set data is predicted using the optimal models. According to the prediction results, the evaluation indexes of each model are obtained.
12. Discussion. According to the prediction results of each model, each model is discussed and analyzed.

4.2 Dataset

The data in this experiment include air quality data and meteorological data. Air quality data is acquired from http://data.epmap.org/, and meteorological data is acquired through API interface (https://www.nowapi.com/api/weather.history). All data is merged and stored in CSV in chronological order. Each data contains 12 items, which are date, AQI, NO_2, O_3, CO, SO_2, PM10, PM2.5, humidity, temperature, weather, and wind level. Where,

Table 1 Partial sample data

Date	AQI	CO	NO_2	O_3	PM10	PM2.5	SO_2	Temperature	Humidity	Weather	winp
2017/1/1 0:00	57	5.7	94	7	524	267	86	−2	97	Foggy	1
2017/1/1 1:00	424	6.5	103	7	524	267	91	−2	97	Foggy	1
2017/1/1 2:00	500	6.8	125	7	696	363	49	−2	97	Foggy	1
2017/1/1 3:00	500	7	128	7	666	373	40	−2	97	Foggy	1
2017/1/1 4:00	500	7	127	8	634	371	40	−2	97	Foggy	1

AQI, NO_2, O_3, CO, SO_2, PM10, PM2.5, humidity, temperature, weather and wind level are inputs to the models. The AQI of the next hour is the output of the models. Some of the data is shown in Table 1. The experimental data are 32,856 pieces of data hourly from 00:00 on January 1, 2017, to 23:00 on September 30, 2020, in Shijiazhuang City, Hebei Province. Among them, 26,282 pieces of data from 00:0 on January 1, 2017 to 23:00 on December 31, 2019 are training sets, and 6574 pieces of data from 00:00 on January 1, 2020, to 23:00 on September 30, 2020, are test sets.

4.3 Data pre-processing

Data pre-processing includes three parts: data transformation, data cleansing, and data standardization. Data transformation is to convert some non-numerical data into numerical data. Data cleansing is to delete duplicate data and fill missing data and unreasonable data. Data standardization is to reduce the difference between data items.

4.3.1 Data transformation

Because there is some non-numerical data in meteorological data, such as sunny and cloudy in weather conditions. The neural network can only calculate the numerical data, so it is necessary to transform the weather conditions in the meteorological data into the corresponding numerical data. The relationship between weather conditions and corresponding values is shown in Table 2. According to the experiment, the weather conditions with the same influence of AQI are converted into the same value (Wang, Li, et al., 2021).

4.3.2 Data cleansing

Data cleansing is used to delete duplicate data and fill in missing data and unreasonable data.

Table 2 Weather conditions quantification

Weather	Value	Weather	Value
Haze	1	Heavy rain	8
Foggy	2	Torrential rain	8
Sunny	3	Heavy torrential rain	8
Cloudy	4	Snow shower	8
Overcast	5	Heavy snow	8
Sleet	6	Blizzard	8
Light rain	6	Extremely torrential downpours	9
Light snow	6	Freezing rain	10
Moderate rain	7	Sandstorm	10
Moderate snow	7	Floating dust	10
Shower	8	Dusty weather	10
Thundery shower	8	Severe sandstorm	10
Hail	8		

Table 3 Partial final data

Date	AQI	CO	NO_2	O_3	PM10	PM2.5	SO_2	Temperature	Humidity	Weather	Winp
2017/1/1 0:00	57	5.7	94	7	524	267	86	−2	97	2	1
2017/1/1 1:00	424	6.5	103	7	524	267	91	−2	97	2	1
2017/1/1 2:00	500	6.8	125	7	696	363	49	−2	97	2	1
2017/1/1 3:00	500	7	128	7	666	373	40	−2	97	2	1
2017/1/1 4:00	500	7	127	8	634	371	40	−2	97	2	1

Table 4 Parameters' Setting of CNN-BiNLSTM-Attention

	Parameters	Value
Input layer	Items	11
	Filters	16
	Kernel size	1
Convolution layer	Activation function	Tanh
	Padding	Valid
Pooling layer	Pool size	1
	Padding	Valid
BiNLSTM layer	Number of hidden units	8
	Activation function	Sigmoid
Output layer	Item	1

There are duplicate dates in the data of this experiment. For duplicate data, the first occurrence of data is saved in this experiment.

When the data is recorded on the corresponding website, there are data that has not been recorded in the time period, so there is missing data in the acquired data. The method to cope with the missing data is to select the average of data for the previous hour and the next hour as the missing value of the missing data.

When the data is recorded on the corresponding website, there is a few data with recording errors. Some data that are not zero are recorded as 0, so it is necessary to turn these data into non-zero data. The processing method for such data is to select the average of data for the previous hour and the next hour as the replacement value of the data.

4.3.3 Data standardization

Due to the large difference in the input data, the data needs to be standardized. The processed data can be better used for model training. In this paper, the data is standardized using z-score standardization. The calculation formula is shown in formula (14).

$$y_i = \frac{x_i - \overline{x}}{s} \tag{14}$$

where x_i is original value, $\overline{x}$ is average value, s is standard deviation.

After data transformation, data cleansing,` and data standardization, they can be input into the model for calculation. The part of final data is shown in Table 3.

Table 5 Comparison of different time step evaluation indexes

Time step	MAE	RMSE	R^2
20	6.049	9.324	0.9736
24	5.987	9.231	0.9741
28	6.086	9.342	0.9734

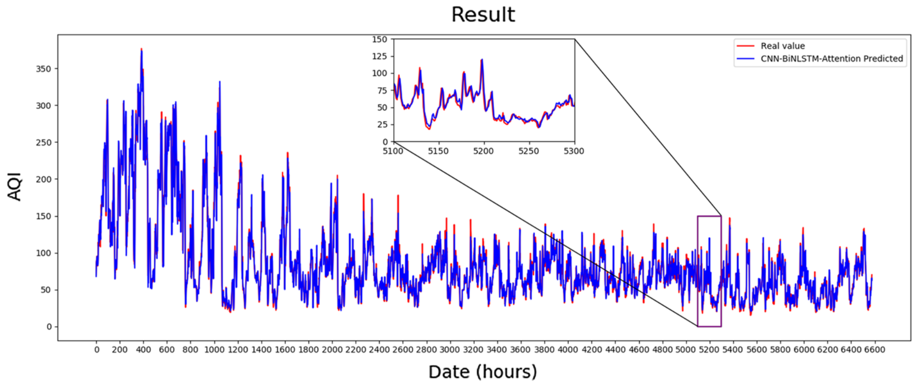


Fig. 4 CNN-BiNLSTM-Attention predicted value and real value

4.4 Parameters setting

The parameters setting of the CNN-BiNLSTM-Attention are shown in Table 4.

Other training parameters are also set. The time step is 24, the learning rate is 0.001, the loss function is MAE, epochs are 128, and the batch size is 128, the optimizer is Adam. All the models are cyclically trained 100 times, and the optimal model is preserved.

4.5 Results

To prove the time step, which is 24, is best, different time steps (20, 24, 28) are used for CNN-BiNLSTM-Attention model training, and predictions are made for the same test set in this paper. The MAE, RMSE, R^2 results of different time steps of CNN-BiNLSTM-Attention model are shown in Table 5.

In Table 5, when time step is 24, the MAE, RMSE is the smallest and the R^2 is the biggest. Therefore, the time step which is 24 is best.

The training set data composed of processed air quality data and meteorological data are used to train all models 100 times. The test set data is predicted using the optimal model.

The test set data is predicted using the trained CNN-BiNLSTM-Attention, and the predicted value and the real value are shown in Fig. 4.

Because a large amount of test set data is too chaotic to demonstrate in one picture, this paper selects part of the test set data from 0:00 on September 21, 2020, to 23:00 on

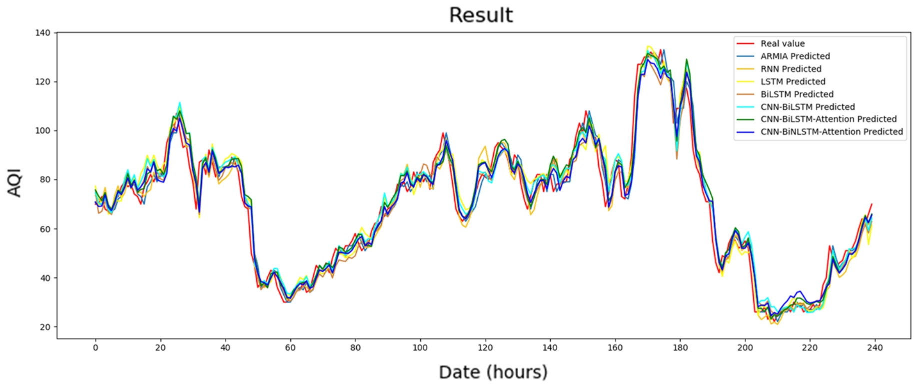


Fig. 5 Seven models predicted value and real value

Table 6 Seven methods evaluation indexes

Model	MAE	RMSE	R^2
RNN	6.819	10.239	0.9681
LSTM	6.589	9.795	0.9708
ARMIA	6.026	9.650	0.9716
BiLSTM	6.325	9.595	0.9720
CNN- BiLSTM	6.284	9.497	0.9726
CNN-BiLSTM-Attention	6.024	9.254	0.9739
CNN-BiNLSTM-Attention	5.987	9.231	0.9741

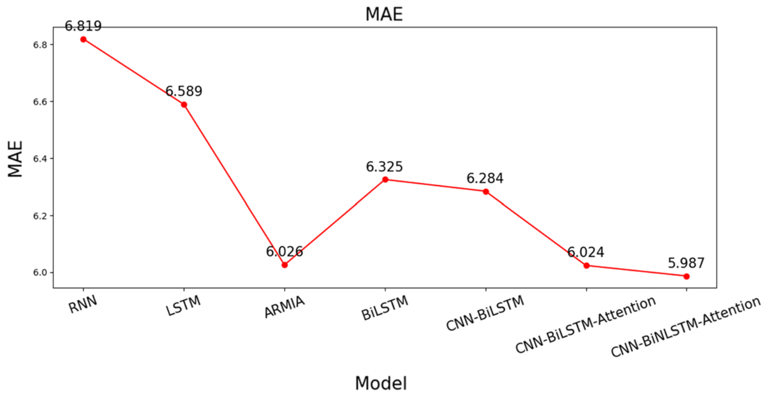


Fig. 6 The MAE of seven models

September 30, 2020, to show the predicted value and real value of the seven models. As shown in Fig. 5.

The MAE, RMSE, R^2 of seven models are shown in Table 6 and Figs. 6, 7, 8.

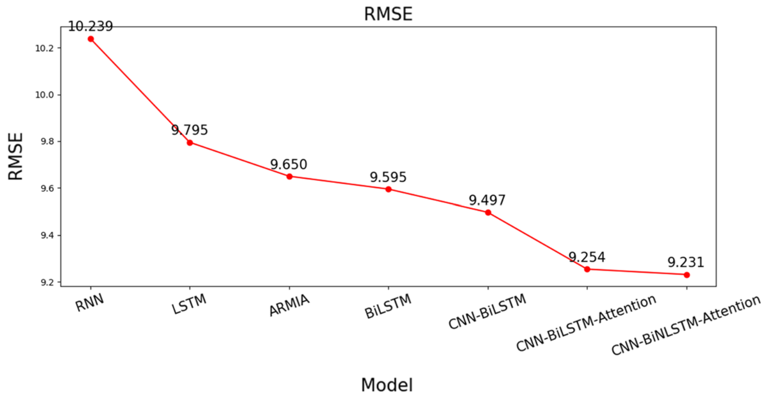


Fig. 7 The RMSE of seven models

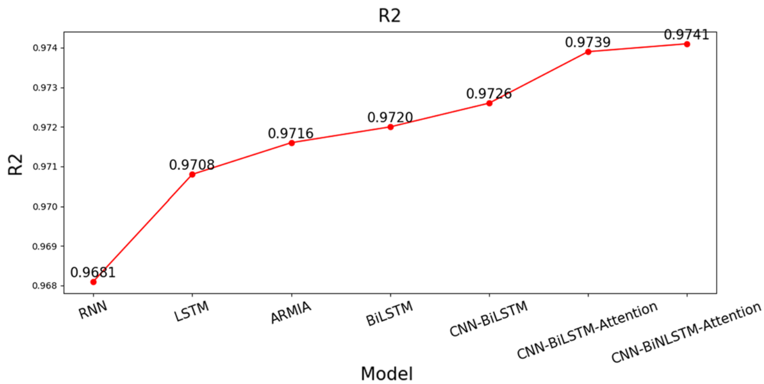


Fig. 8 The R^2 of seven models

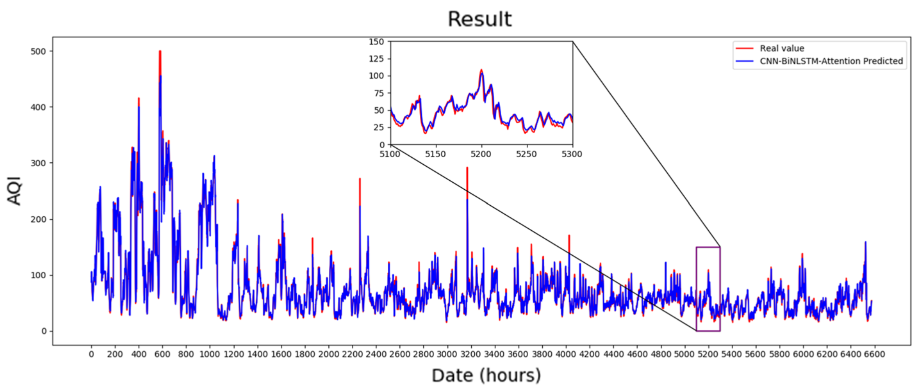


Fig. 9 Comparison of CNN-BiNLSTM-Attention predicted value and real value

To prove the generalization ability of CNN-BiNLSTM-Attention model, 6574 pieces of data hourly from 00:00 on January 1, 2020, to 23:00 on September 30, 2020, in Baoding City, Hebei Province is predicted in this paper. The predicted value is compared with the real value as shown in Fig. 9. The MAE is 5.673, the RMSE is 10.055 and R^2 is 0.9711. Therefore, the model has generalization ability.

4.6 Discussion

In Table 6 and Figs. 6, 7, 8, the MAE, RMSE of CNN-BiNLSTM-Attention are the smallest, and R^2 is the biggest. Among the seven models, CNN-BiNLSTM-Attention has the highest prediction accuracy and RNN has the lowest prediction accuracy. Comparing LSTM with RNN, MAE decreases from 6.819 to 6.589, RMSE decreases from 10.239 to 9.795 and R^2 increases from 0.9681 to 0.9708. The results show that LSTM is better than RNN in prediction accuracy. Comparing BiLSTM with LSTM, MAE and RMSE of BiLSTM decrease by 0.264 and 0.200, respectively, and R^2 increases by 0.0027. The results show that BiLSTM is better than LSTM. Comparing CNN-BiLSTM with BiLSTM, MAE decreases from 6.325 to 6.284, decreases by 0.041. RMSE decreases from 9.595 to 9.497, decreases by 0.098; R^2 increases by 0.9726 from 0.9720, increases by 0.0006. The results present that compared with a single network, the prediction accuracy of the composite network is improved. Comparing CNN-BiLSTM-Attention with CNN-BiLSTM, MAE reduces 4.14%, RMSE reduces by 2.56%, and R^2 adds 0.1%, from 0.9726 to 0.9739. Comparing CNN-BiNLSTM-Attention with CNN-BiLSTM-Attention, MAE decreases from 6.024 to 5.987, RMSE decreases from 9.254 to 9.231, and R^2 increases from 0.9739 to 0.9741, which is closer to 1. The results p`resent the prediction accuracy can be improved by using Attention. The comparative experiments show that all the evaluation indexes of CNN-BiNLSTM-Attention are the best. MAE, RMSE are 5.987 and 9.231 respectively. R^2 is closest to 1, with a value of 0.9741.

Therefore, the CNN-BiNLSTM-Attention air quality prediction model can well predict the AQI in the next hour. It also can provide a reference to help people to take measures in advance to reduce air pollution. However, this model can't predict AQI over a longer period of time, but can only predict the AQI for one hour in the future.

5 Conclusions

By analyzing the data, this paper concludes that the data have the characteristics of time series, and a CNN-BiNLSTM-Attention is proposed air quality prediction model to predict the AQI for the next hour. The model includes CNN, NLSTM, and attention. CNN is used to extract characteristics of the input air quality data and meteorological data. Then NLSTM is an improvement on LSTM. By introducing the 1-tanh function, the output value of forget gate can be in a more obvious range. Therefore, NLSTM can preserve more characteristics of the input data. BiNLSTM is used to predict time series data. Finally, attention is used to capture the effect of the characteristic conditions at imparity times on AQI prediction, and the more accurate AQI is obtained by weighting calculation. The model is compared with five other models. All models use the same data to train. And the hourly AQI of Shijiazhuang City, Hebei Province from 00:00 on January 1, 2020, to 23:00 on September 30, 2020, is predicted. The experimental results indicate that CNN-BiNLSTM-Attention is an optimal model to predict air quality. MAE, RMSE are 5.987 and 9.231 respectively and R^2 is the closest to 1. Improving the complexity of the model is conducive to improve the accuracy of prediction. CNN-BiNLSTM-Attention is suitable for predicting AQI, which is beneficial to improve air pollution and provide early warning of air quality for people. The proposal of this model also provides a reference for people to research air quality prediction.

Future research will focus on improving the structure of NLSTM and Attention model to make it have better learning ability, and applying the model to other time series prediction problems to make it universal.

Acknowledgments This research was funded by Innovation Foundation for Postgraduate of Hebei Province under Grant CXZZSS2021104, the Special Project for Cultivating College and Middle School Students' Scientific and Technological Innovation Ability under Grant 2021H011410 and the fund of Hebei University of Science and Technology under Grant 2019-ZDB02.

References

Caselli, M., Trizio, L., Gennaro, G., & Lelpo, P. (2009). A simple feedforward neural network for the PM 10 forecasting: Comparison with a radial basis function network and a multivariate linear regression model. *Water, Air, and Soil Pollution, 201*(1–4), 365–377.

Eslami, E., Choi, Y., Lops, Y., & Sayeed, A. (2020). A real-time hourly ozone prediction system using deep convolutional neural network. *Neural Computing and Applications, 32*, 8783–8797.

Fong, I., Li, T., Fong, S., et al. (2020). Predicting concentration levels of air pollutants by transfer learning and recurrent neural network. *Knowledge-Based Systems, 192*(15), 1–10.

Freeman, B., Taylor, G., Gharabaghi, B., & The, J. (2018). Forecasting air quality time series using deep learning. *Journal of the Air & Waste Management Association, 68*(8), 866–886.

Gu, K., Zhou, Y., Sun, H., Zhao, L., & Liu, S. (2020). Prediction of air quality in Shenzhen based on neural network algorithm. *Neural Computing and Applications, 32*, 1879–1892.

Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory*. MIT Press.

Huang, C., & Kuo, P. (2018). A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. *Sensors, 18*(7), 2220–2241.

Jin, X., Yang, N., Wang, X., Bai, Y., Su, T., & Kong, J. (2020). Deep hybrid model based on EMD with classification by frequency characteristics for long-term air quality prediction. *Mathematics, 8*(2), 214–230.

Koo, J., Wong, S., Selvachandran, G., Long, H., & Son, L. (2020). Prediction of air pollution index in Kuala Lumpur using fuzzy time series and statistical models. *Air Quality, Atmosphere & Health, 13*(1), 77–88.

Krishan, M., Jha, S., Das, J., et al. (2019). Air quality modelling using long short-term memory (LSTM) over NCT-Delhi India. Air quality. *Atmosphere & Health, 12*, 899–908.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE., 86*(11), 2278–2324.

Li, T., Hua, M., & Wu, X. (2020). A hybrid CNN-LSTM model for forecasting particulate matter (PM2.5). *IEEE Access, 8*, 26933–26940.

Li, X., Luo, A., Li, J., & Li, Y. (2019). Air pollutant concentration forecast based on support vector regression and quantum-behaved particle swarm optimization. *Environmental Modeling & Assessment, 29*(2), 205–222.

Li, X., Peng, L., Shao, J., Cui, S., & Tian, H. (2016). Air pollution forecast based on wavelet decomposition and ARMA model. *Environmental Engineering, 34*(8), 110-113+134.

Liu, D., Hsu, Y., Chen, H., et al. (2020). Air pollution prediction based on factory-aware attentional LSTM neural network. *Computing, 103*, 75–98.

Ma, J., Ding, Y., Gan, V., Lin, C., & Wan, Z. (2019). Spatiotemporal prediction of PM2.5 concentrations at different time granularities using IDW-BLSTM. *IEEE Access, 7*, 107897–107907.

Mishra, D., & Goyal, P. (2015). Development of artificial intelligence based NO2 forecasting models at Taj Mahal Agra. *Atmospheric Pollution Research, 6*(1), 99–106.

Nieto, P., Lasheras, F., García-Gonzalo, E., & Juez, F. (2018). PM 10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: A case study. *Science of The Total Environment, 621*(15), 753–761.

Niu, B., & Yin, Y. (2016). The prediction and research of air quality in chengdu based on ARMA model. *Statistics and Application, 5*(4), 365–372.

Ong, B., Sugiura, K., & Zettsu, K. (2016). Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM2.5. *Neural Computing & Applications, 27*, 1553–1566.

Pagowski, M., Grell, G. A., Devenyi, D., et al. (2006). Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts. *Atmospheric Environment, 40*(18), 3240–3250.

Pai, T., Sung, P., Lin, C., et al. (2010). Predicting hourly ozone concentration in dali area of taichung county based on multiple linear regression method. *International Journal of Applied Science and Engineering, 7*(2), 127–131.

Peng, H., Aranildo, L., Teakles, A., et al. (2017). Evaluating hourly air quality forecasting in Canada with nonlinear updatable machine learning methods. *Air Quality, Atmosphere and Health, 10*, 195–211.

Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A., & Jia, J. (2020). Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Networks, 121*, 396–408.

Sayegh, A., Munir, S., & Habeebullah, T. (2014). Comparing the performance of statistical models for predicting PM10 concentrations. *Aerosol & Air Quality Research, 14*(3), 653–665.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.

Wang, B., Kong, W., Guan, H., & Xiong, N. (2019). Air quality forecasting based on gated recurrent long short-term memory model in internet of things. *IEEE Access, 7*, 69524–69534.

Wang, J., Li, J., Wang, X., Wang, J., & Huang, M. (2021a). Air quality prediction using CT-LSTM. *Neural Computing and Applications, 33*, 4779–4792.

Wang, X., & Wang, B. (2019). Research on prediction of environmental aerosol and PM2.5 based on artificial neural network. *Neural Computing and Applications, 31*, 8217–8227.

Wang, X., Yuan, J., & Wang, B. (2021b). Prediction and analysis of PM2.5 in fulling district of Chongqing by artificial neural network. *Neural Computing and Applications, 33*, 517–524.

White, H. (1993). Economic perdication using neural networks: The case of ibm daily stock returns. *IEEE 1988 International Conference on Neural Networks*, *2,* 451–458.

Xayasouk, T., Lee, H., & Lee, G. (2020). Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models. *Sustainability, 12*(6), 2570–2586.

Xing, H., Wang, G., Liu, C., & Suo, M. (2021). PM2.5 concentration modeling and prediction by using temperature-based deep belief network. *Neural Networks, 133*, 157–165.

Yang, G., Lee, H., & Lee, G. (2020). A hybrid deep learning model to forecast particulate matter concentration levels in seoul South Korea. *Atmosphere, 11*(4), 348–367.

Zhao, J., Deng, F., Cai, Y., & Chen, J. (2019). Long short-term memory—Fully connected (LSTM-FC) neural network for PM 2.5 concentration prediction. *Chemosphere, 220*, 486–492.