IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 14, 2020, accepted January 23, 2020, date of publication February 3, 2020, date of current version February 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2971348

A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5)

TAOYING LI, MIAO HUA, AND XU WU
School of Shipping Economics and Management, Dalian Maritime University, Dalian 116026, China

Corresponding author: Taoying Li (litaoying@dlmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 71271034, Grant 51939001, and Grant 61976033, in part by the Liaoning Revitalization Talents Program under Grant XLYC1907084, in part by the Natural Science Foundation of Liaoning Province under Grant 20180550307, and in part by the Fundamental Research Funds for the Central Universities under Grant 3132019353.

ABSTRACT PM2.5 is one of the most important pollutants related to air quality, and the increase of its concentration will aggravate the threat to people's health. Therefore, the prediction of surface PM2.5 concentration is of great significance to human health protection. In this study, A hybrid CNN-LSTM model is developed by combining the convolutional neural network (CNN) with the long short-term memory (LSTM) neural network for forecasting the next 24h PM2.5 concentration in Beijing, which makes full use of their advantages that CNN can effectively extract the features related to air quality and the LSTM can reflect the long term historical process of input time series data. The air quality data of the last 7days and the PM2.5 concentration of the next day are first set as the input and output of the model due to the periodicity, respectively. Subsequently four models namely univariate LSTM model, multivariate LSTM model, univariate CNN-LSTM model and multivariate CNN-LSTM model, are established for PM2.5 concentration prediction. Finally, mean absolute error (MAE) and root mean square error (RMSE) are employed to evaluate the performance of these models and results show that the proposed multivariate CNN-LSTM model performs the best results due to low error and short training time.

INDEX TERMS Deep learning, CNN, LSTM, PM2.5 concentration prediction.

I. INTRODUCTION

In recent years, with the rapid development of China's economy and industrialization, the problem of environmental pollution is becoming increasingly serious [1]. Air pollution is particularly significant. In 2004, the first American Heart Association concluded that exposure to particulate matter (PM) air pollution contributes to cardiovascular morbidity and mortality [2]. Since then, air pollution has been widely concerned by governments and society. Nowadays, the air pollution is taken as the topmost important issue in our daily life because large scale haze attacks not only seriously affect people's normal transportation, but also seriously harm people's health. PM2.5 is one of the main components of haze [3] and increased daily mortality is specifically associated with particle mass constituents found in the aerodynamic diameter size range under 2.5 um [4], [5]. Therefore, monitoring and forecasting PM2.5 concentration are of great significance for human health and environmental management.

The associate editor coordinating the review of this manuscript and approving it for publication was Quan Zou.

The formation mechanism and process of PM2.5 is very complex [6] due to many complex properties, such as nonlinear characteristics in time and space, which will have a great impact on the prediction accuracy [7]. Therefore, it is essential to analyze it in detail. This air quality data is closely related to time, which means it belongs to time series data [8], and has obvious periodicity. Because of the timeliness of data, time series prediction has become a hot topic [9]. Time series analysis plays an important role in a large variety of application fields, such as economics, medicine, astronomy, geology, etc. [19]. There are many mature time series prediction methods, including ARMA [10], ARIMA [10], SARIMA [11], SVR [12], BP neural network [13], Bayesian network [14] and so on. However, with the increase of amount and complexity of obtained data, these methods can no longer meet the actual demand due to too much training time. With the development of deep learning, the time series model makes the predication of PM2.5 possible.

Recently, with the popularity of Artificial Intelligence, many deep learning algorithms have been developed, such as Deep Belief Network [15], Convolutional Neural

Network (CNN) [16] and Recurrent Neural Network (RNN) [17], etc., which are widely used in pattern recognition, object detection, natural language processing, image classification and other fields [18]. With the growth of data and the improvement of demand, the network built for data analysis is increasingly complex and it is no longer a single network model, but a more complex hybrid network. For example: Ding *et al.* [20] developed a new hybrid deep learning model that integrated CNN and LSTM that automatically recognized workers' unsafe actions. The model's accuracy exceeded the current state-of-the-art descriptor-based methods for detecting points of interest on images. By incorporating ConvLSTM into the encoding-forecasting structure, Shi *et al.* [21] built an end-to-end trainable model for precipitation nowcasting. Kanjo *et al.* [22] using a hybrid deep learning approach (CNN-LSTM) on large number of raw sensor data increased the accuracy levels of emotion models by more than 20% compared to a traditional MLP model. Duan *et al.* [23] proposed a deep hybrid neural network improved by greedy algorithm for urban traffic flow prediction with taxi GPS trace. These studies mentioned above show that deep learning is a promising approach and some researchers have already applied it to study the air quality. Zhao *et al.* [24] proposed a long short-term memory-fully connected (LSTM-FC) neural network, to predict PM2.5 contamination. Huang *et al.* [25] verified feasibility and practicability of the CNN-LSTM model to predict the PM2.5 concentration. Pak *et al.* [26] constructed a CNN-LSTM hybrid model that combines CNN to predict the next day's 8-h average ozone concentration in Beijing City.

But the lower accuracy and long predictable time of existing methods couldn't meet the demand for forecasting PM2.5 in daily life. Meanwhile, due to the complexity of PM2.5 formation, the high accuracy and efficiency demand for prediction, and the difficulty of deep learning network model in stability, it is essential to develop more effective model for forecasting PM2.5 concentration. Therefore, a hybrid CNN-LSTM model is proposed for forecasting the PM2.5 concentration of the next day (next 24 hours). In order to test alternative models which is best, four models incl. univariate LSTM model, multivariate LSTM model, univariate CNN-LSTM model, and multivariate CNN-LSTM model are compared and analyzed. Finally, two indicators are adopted to evaluate the models, which are mean absolute error (MAE) and root mean square error (RMSE).

The reminder of the paper is arranged as follows. In Section 2, the methodologies are described, in which CNN and LSTM are presented in detail and the hybrid CNN-LSTM model for forecasting PM2.5 concentration is proposed. In Section 3, the data preprocessing is completed for padding missing values with zeros and normalizing values of features to fall within the range of 0-1. In section 4, the results are shown and discussed. Finally, we conclude the paper in Section 5.

II. MATHODOLOGIES

Deep learning has become one of the most important methods in the field of machine learning, and has been used widely in image analysis, speech recognition and text understanding [27]. The most commonly used deep learning algorithms contain convolutional neural network (CNN), recurrent neural network(RNN), and self-encoders network, etc., among which, the advantage of CNN is the feature extraction [28], and RNN is good at mining the time series data [17]. Therefore, in order to make full use of their advantages, we combine these two models for obtaining a new effective model, which will be presented in three aspects as follows.

A. CNN MODEL

CNN is one of the most successful deep learning methods, and its network structures include 1D CNN, 2D CNN and 3D CNN [29]. 1D CNN is mainly used for sequence data processing [30], 2D CNN is often used for image and text recognition [28], and 3D CNN is mainly used for medical image and video data recognition [31]. So, 1D CNN is adopted in this paper.

1D CNN can be well applied on sensor data (gyroscope or accelerometer data) [32] for time series analysis, and can also be employed for analyzing periodical signal data (audio signal) [33]. The detailed process of the 1D CNN is described as following.

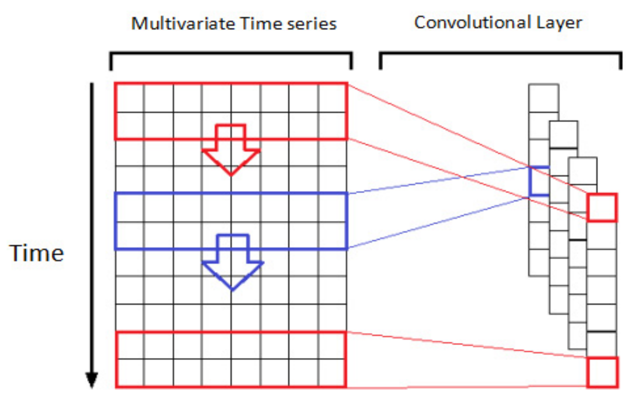


FIGURE 1. The process of 1D CNN.

The left of Figure 1 is the input time series data which is a multi-dimensional matrix, which is convoluted from top to bottom as shown by the arrow in Figure 1, and the red represents a filter. The number of the extracted feature dimensions is N*1 after convolution with a filter, where N is related to the number of input data dimensions, the size of filter and convolution step length. The blue indicates another filter, which can be followed by other filters. Suppose the number of filters is M, and the extracted feature dimension will be N * M.

Figure 2 shows the process of a one-dimensional convolution. For the PM2.5 dataset, the input is a 168*8 matrix, and

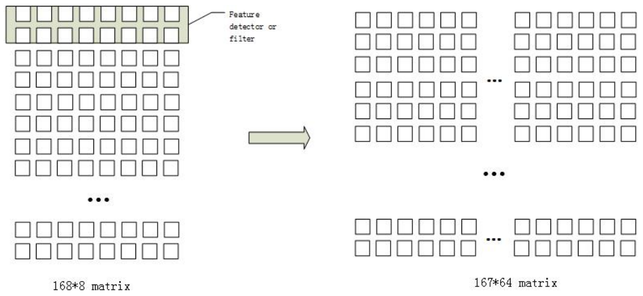


FIGURE 2. The results of air quality data after 1D convolution.

the output is a 167*64 matrix after convolution with 64 filters of size 2.

B. LSTM MODEL

Long short term memory (LSTM) [34] is a deformation structure of RNN by adding memory cell into hidden layer, so as to control the memory information of the time series data. Information is transmitted among different cells of hidden layer through several controllable gates (forget gate, input gate, output gate) [35], thus the memory and forgetting extent of the previous and current information can be controlled. Compared with traditional RNN, the LSTM has the long term memory function and its gradient disappearance problem can be avoided. Two gates of LSTM are designed for controlling the state of memory cell, one is forget gate which indicates how much "memory" of the last moment's cell can be saved, the other is input gate, which determines how much input of the current moment can be saved to the cell state, and controls the proportion of fusion of "historical" information and "current" stimulus. Finally, the output gate of LSTM is designed for controlling how much information is output for cell status. The structure of LSTM network is shown in Figure 3.

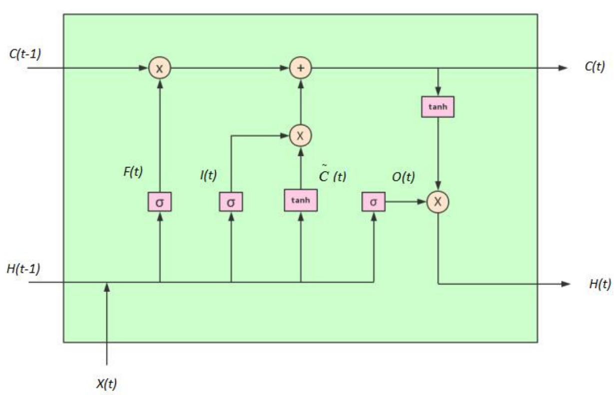


FIGURE 3. Network structure of the LSTM.

In Figure 3, σ is the sigmoid function shown in equation (7), whose output is a value between 0 and 1. Here, 0 means "let nothing pass" while 1 means "let everything pass". Then the hyperbolic tangent function illustrated in Equation (8), is used to overcome the problem of gradient disappearance. The input and output of the network structure of the LSTM in Figure 6 can be described as Eqs. (1) - (8).

$$F(t) = \sigma(W_f \cdot [H_{t-1}, X_t] + b_f) \quad (1)$$

$$I(t) = \sigma(W_i \cdot [H_{t-1}, X_t]) + b_i) \quad (2)$$

$$\tilde{C}(t) = tanh(W_c \cdot [H_{t-1}, X_t] + b_c) \quad (3)$$

$$C(t) = f_t * C_{t-1} + I_t * \tilde{C}_t \quad (4)$$

$$O(t) = \sigma(W_o \cdot [H_{t-1}, X_t] + b_o) \quad (5)$$

$$H(t) = O_t * tanh(C_t) \quad (6)$$

$$sigmoid(x) = \frac{1}{1+e^{-x}} \quad (7)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (8)$$

where W_f, W_i, W_c and W_o are input weights, b_f, b_i, b_c and b_o are bias weights, t represents the current time state, and $t-1$ is the previous time state, X represents input; H represents output and C is the cell status.

C. THE HYBRID CNN-LSTM MODEL

In this section, a hybrid CNN-LSTM model is constructed by combining CNN with LSTM for improving the accuracy of forecasting PM2.5 concentration. The proposed hybrid CNN-LSTM model is a prediction model with multivariate time series data as input and multi-step single time series data as output, whose process is given in Figure 4.

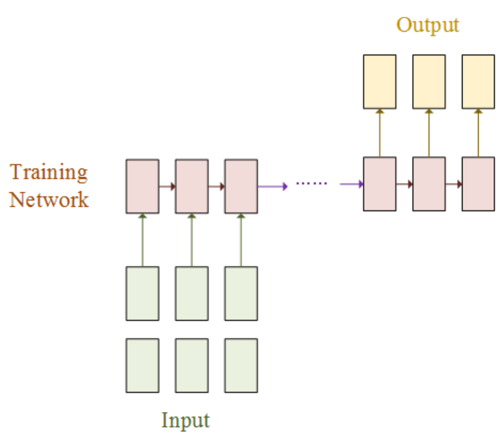


FIGURE 4. A simple architecture of the hybrid CNN-LSTM model.

In Figure 4, the light green represents the input, the light purple is the proposed hybrid CNN-LSTM model, and the yellow denotes the output. The network structure of the proposed hybrid CNN-LSTM model is developed as shown in Figure 5.

CNN is adopted for feature extraction, specifically, two one-dimensional convolutional layers and a MaxPooling layer are constructed. In order to process the data into the format required by the LSTM, a Flatten layer is connected. Overfitting is a common phenomenon in deep neural network (DNN) and there are many solutions, among all solutions, dropout is one of the simple ones and works well.

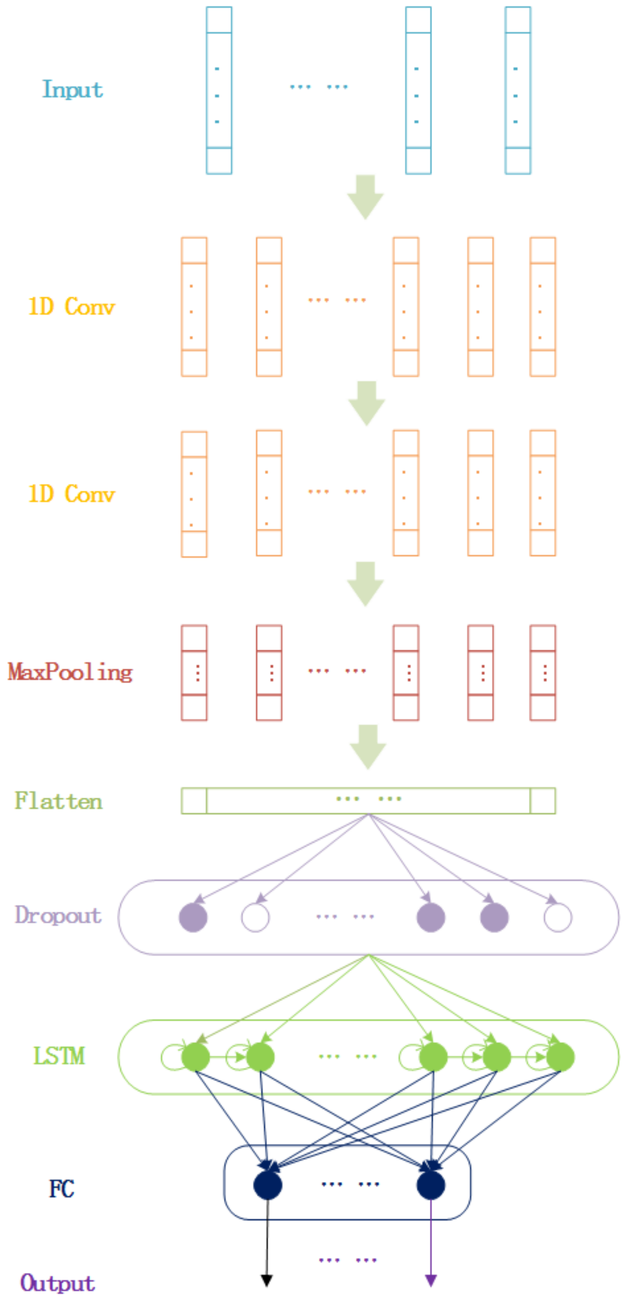


FIGURE 5. Network structure of the proposed hybrid CNN-LSTM model.

Dropout [36] refers to that during the training process of DNN, and the cell is temporarily dropped from the network according to a certain probability. In order to avoid overfitting, a Dropout layer is added, whose output is connected to the LSTM layer for prediction and finally connect to a FC layer. In Figure 5, in the dropout layer, the white cell is the temporarily discarded part. Please note that for stochastic gradient descent, each mini-batch will train a different network due to random drop.

The activation function puts the nonlinear factors into the neural network, which improves the expression ability of the

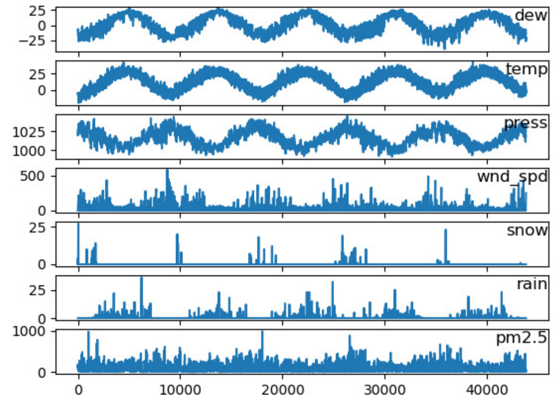


FIGURE 6. Time series chart of seven features of PM2.5 data.

neural network and solves the nonlinear problems that the linear model cannot solve [37]. There are many frequently used activation functions, including sigmoid function, *tanh* function (hyperbolic tangent function), *relu* function (Rectified Linear Unit), etc. The *relu* function can solve the problem of gradient disappearance, and its calculation speed and convergence speed are faster than sigmoid function and *tanh* function, which is defined as equation (9).

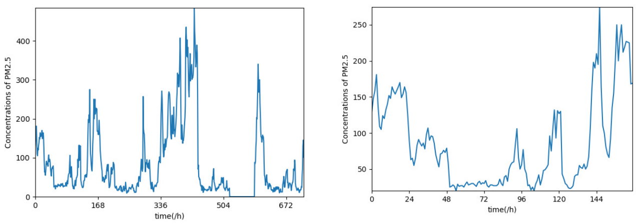
$$relu\,(x) = max(0, x) \tag{9}$$

III. DATA SOURCE AND PREPROCESSING

The dataset (https://archive.ics.uci.edu/ml/datasets.php) chosen in this paper contains the hourly values of PM2.5 concentration of US Embassy in Beijing and the meteorological data of Beijing Capital International Airport. This dataset totally contains 43800 records with multi-features, including date, PM2.5 concentration, dew point, temperature, atmospheric pressure, combined wind direction, cumulated wind speed, and cumulated hours of snow and rain. However, this dataset contains many missing values due to some uncontrollable reasons. Thus, in this paper, we first fill the missing values with zero. Next, the wind direction is encoded and converted into digital values. Then, we analyze the data and draw a line chart of each feature to observe its time characteristics.

Figure 6 shows that each feature of the dataset has certain periodicity, among all features PM2.5 concentration has the most complex characteristics, thus we analyze it in detail (see in Figure 7). In Figure 7, the two sub-graphs show the trends of PM2.5 concentration in one month (30 days) and one week (7 days), which indicates the periodic characteristic. The prediction of PM2.5 concentration for the next day can help people make decisions on travel and life. In order to forecast the PM2.5 concentration of the next day, the data of the last week (7 days) can be chosen as the input of the forecast model. Since then, the model takes the PM2.5 concentration of one week as input and that of next one day as output.

Due to the complexity of PM2.5 data, we draw the Figure 8 for probing into the distribution of PM2.5 concentration, where horizontal axis indicates the PM2.5 concentration



(a)The concentration of PM2.5 in one month

(b)The concentration of PM2.5 in one week

FIGURE 7. The trends of PM2.5 concentration.

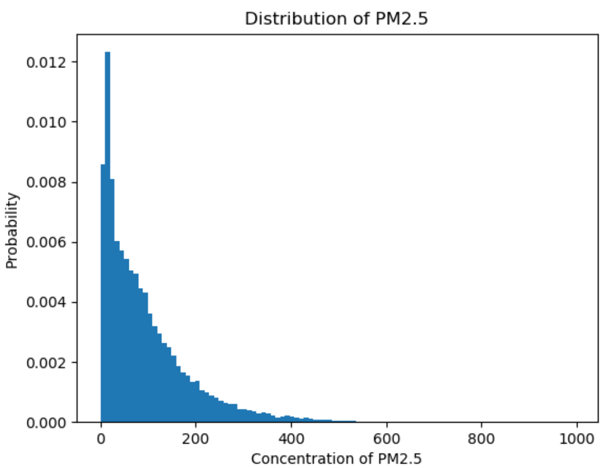


FIGURE 8. The histogram of PM2.5 concentration.

TABLE 1. The analysis of PM2.5 concentration.

Max	Min	Median	Average	Std
994	0	68	94.0135	92.2512

with the interval of 10 μg/m^3, and vertical axis represents frequency. Meanwhile, the values of the maximum, minimum, median, average, standard deviation of PM2.5 concentration are 994 ug/m^3, 0 ug/m^3, 68 ug/m^3, 94.0135 ug/m^3, 92.2512 ug/m^3 respectively (see in Table 1). The distribution of the PM2.5 concentration is very fragmented according to large value of the standard deviation of PM2.5 concentration, which increases the difficulty of prediction.

In order to improve the prediction accuracy, we normalize the value of PM2.5 concentration using Min-Max normalization method given in equation (10).

$$x = \frac{x - min}{max - min} \tag{10}$$

Considering that the PM2.5 concentration is time series data, we choose the first 80% of dataset as the training data and the remaining 20% as the test data.

IV. RESULTS AND FINDINGS

For the dataset mentioned above, the trained network structure of the proposed hybrid CNN-LSTM model, generated by

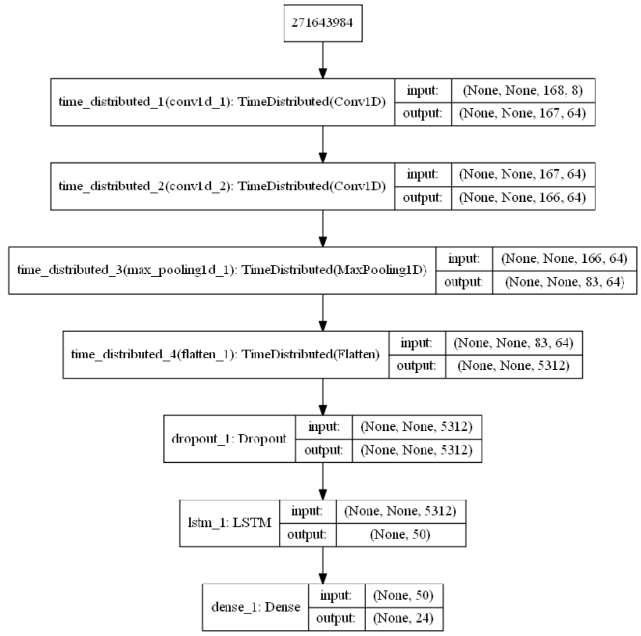


FIGURE 9. Internal network structure of the proposed hybrid CNN-LSTM model generated by anaconda platform.

anaconda platform, is obtained as shown in Figure 9, which contains 271,643,984 adjustable parameters and its input is a 168*8 matrix. After feature extraction of convolutional layer and time series prediction of LSTM layer, the size of output is a 24*1 matrix (vector). In other words, the data of the last week is used to predict the concentration of PM2.5 of the next day. The size of the input and output data of each layer can be easily expressed in Figure 9.

The prediction models for time series data can be divided into two categories: univariate prediction model and multivariate prediction model. Meanwhile, there are so many models for forecasting PM2.5 concentration, it is essential to test alternative models for identifying the best, so the univariate LSTM model, the univariate CNN-LSTM model, the multivariate LSTM model and the multivariate CNN-LSTM model are compared in this section. The following is the fitting effect chart of these four models.

Figure 10 shows the loss function chart of these four models, and the network is prone to the phenomenon of overfitting. We have adopted the Dropout method and set the number of iterations during the experiments for preventing overfitting and results show that all of these four models perform good fitting effect.

Figure 11 shows that the accuracy of these four models for forecasting PM2.5 concentration and PM2.5 concentration fluctuates widely due to the influence of temperature, wind direction, wind force, etc. After comparing these four models horizontally and vertically, we find that the prediction accuracies of CNN-LSTM models are higher than that of single LSTM models and the prediction accuracies of multivariate models are also higher than that of univariate models.

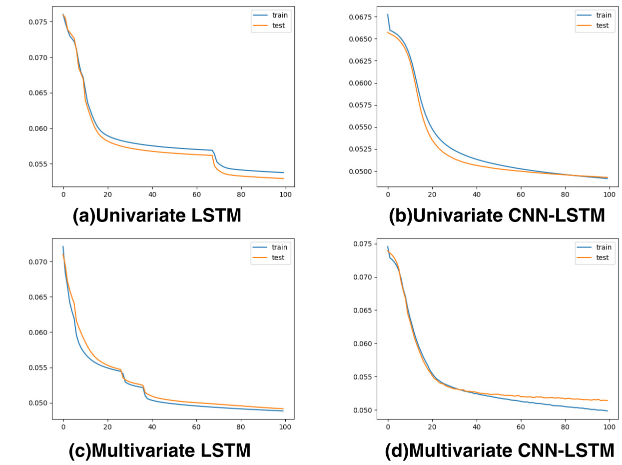


FIGURE 10. Figure of loss function.

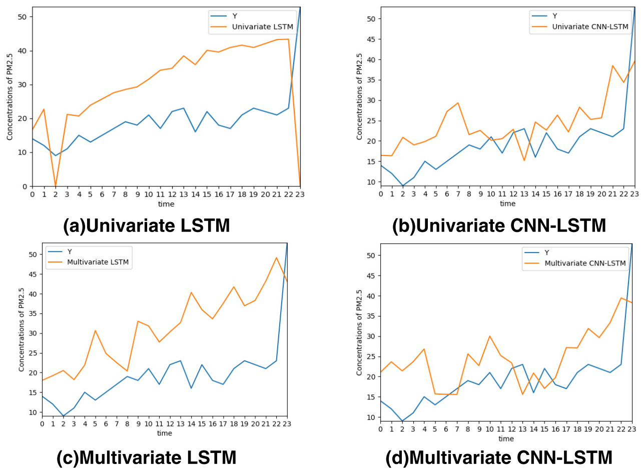


FIGURE 11. Comparison of true values and predicted values of these four models.

In order to evaluate the performance of these models, ten samples (ten 7-day time series data) were randomly selected from the test data for predicting the PM2.5 concentration of the next day respectively, and two indicators were employed, namely mean absolute error (MAE) and root mean square error (RMSE) defined by equations (11) and (12), and their values for these four models are given in Table 2 and Table 3.

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

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

where $\hat{y}_i$ and y_i indicate the predicted value and true value, respectively.

Table 2 shows the MAE of these four models on ten samples. 1-10 in the column of samples represent the number of these 10 different samples, and other values in are the MAE values between the predicted and true values of PM2.5 concentration. The values in the row of average are the average of MAE of 10 samples, which show that the MAE of multivariate CNN-LSTM model, 13.9697, is the

TABLE 2. The MAE of experimental results.

Samples	Multivariate CNN-LSTM	Multivariate LSTM	Univariate CNN-LSTM	Univariate LSTM
1	7.73	12.896	6.779	19.22
2	11.985	27.701	16.093	34.272
3	8.655	8.145	20.449	13.936
4	12.062	19.073	15.583	33.415
5	9.457	12.892	18.226	19.885
6	25.548	13.822	13.13	20.225
7	13.518	16.488	18.913	19.183
8	18.939	14.239	19.394	9.586
9	16.885	12.663	13.878	17.756
10	14.918	9.749	18.764	11.976
Average	**13.9697**	**15.3243**	**16.1209**	**19.9454**

TABLE 3. The RMSE of experimental results.

Samples	Multivariate CNN-LSTM	Multivariate LSTM	Univariate CNN-LSTM	Univariate LSTM
1	8.872	14.407	8.119	20.853
2	16.926	31.458	23.716	45.033
3	11.558	9.384	24.462	14.814
4	14.349	23.283	17.773	36.841
5	11.544	14.06	23.371	23.013
6	33.689	16.222	16.477	21.562
7	17.813	19.403	23.638	20.009
8	26.917	19.477	22.044	14.476
9	19.708	15.073	16.803	20.645
10	16.251	10.76	21.289	15.4
Average	**17.9306**	**18.0852**	**19.7692**	**23.2646**

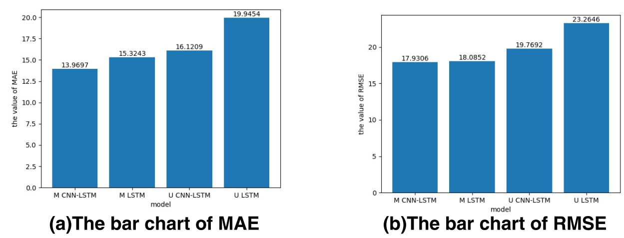


FIGURE 12. The bar chart of MAE and RMSE.

minimum. Moreover, the MAE values of multivariate models are obviously lower than that of univariate models, and the MAE values of CNN-LSTM models are greatly lower than that of LSTM models.

Similarly, Table 3 shows that the RMSE of multivariate CNN-LSTM model is the lowest with a value of 17.9306. The RMSE values of multivariate models are obviously lower than that of univariate models, and the RMSE values of both the univariate and multivariate CNN-LSTM models are slightly lower than that of LSTM models, which can be clearly shown in Figure 12.

Figure 12 (a) and (b) respectively represent the values of MAE and RMSE of these four models and show that the

TABLE 4. Results of different length of training data.

Samples	The length of the training data		
Results	1 day	7 days	14 days
MAE	16.810	11.985	22.053
RMSE	18.991	16.926	23.387
Training time(S/EPOCH)	10	55	165

accuracy of multivariate CNN-LSTM is higher than that of other models. Meanwhile, the experiments are based on a computer with Intel (R) core (TM) i7-3770 CPU @ 3.40GHz 3.40 GHz and the RAM 4.00 GB. In terms of training time, the multivariate CNN-LSTM took about 50-60 seconds for one epoch while other models consumed about 90-100 seconds for one epoch.

Therefore, based on the above results, we can summarize some findings at the technical aspects:

(1). All features of the air quality data of Beijing have certain periodicity, and after a detailed analysis of PM2.5 concentration, it has obviously daily and weekly time series periodicity. So it is necessary to choose the last week's (7 days') air quality data as the input for forecasting the PM2.5 concentration of the next day.

(2). There are many features related to air quality data which may influence the accuracy and efficiency of PM2.5 concentration prediction, so it is essential to adopt data driven methods for quickly identifying the key features and constructed univariate and multivariate models for comparison. In the univariate models, only one feature, PM2.5 concentration, is contained. In the multivariate models, some possible relevant features, including weather, wind speed, wind direction and atmospheric pressure, etc., are contained. The results show that the accuracies of multivariate models are higher than that of univariate models for predicting PM2.5 concentration. Therefore, if the data has multiple features, we should choose the multivariate model.

(3). Almost all models have their own advantages and disadvantages, so it is vital to propose a hybrid CNN-LSTM model for predicting the PM2.5 concentration, in which CNN is employed to extract related features from existing air quality features, and then LSTM is adopted to make predictions. The final results show that the proposed model improves the accuracy of prediction and reduces the training time. Moreover, the model has flexibility, its input and output can be adjusted according to demand. If we want to predict the PM2.5 concentration of the next 48 hours, we only need to adjust the length Y of the training data to 48.

Besides, we did some other studies and results indicate that no matter whether the training data is less or more, the results will get worse. As shown in Table 4, the data of one day, seven days, and 14 days were respectively employed as training data for predicting the PM2.5 concentration of the next day. Their results show that the prediction error with 7 days as training data is smallest and the training time will increase with the amount of the training data. Meanwhile, we also predicted the PM2.5 concentration of the next two days, but it showed the longer training time and lower accuracy.

V. CONCLUSION

It is obvious that the characteristics of air quality have periodicity by analyzing the data of air quality in Beijing. A hybrid CNN-LSTM deep learning network is proposed based on convolutional neural network and recurrent neural network for predicting the PM2.5 concentration of Beijing. The advantages of convolutional neural network for feature extraction and recurrent neural network for time series data processing are utilized for improving the accuracy of forecasting the air quality. Due to the periodicity of air quality data, the values of features related to air quality of one week are chosen as the input, and the PM2.5 concentration of the next day is chosen as the output. The process of the prediction contains following steps: First, the dataset was normalized and then divided into training data (the first 80% of data) and test data (remaining data). Then the proposed hybrid CNN-LSTM model was applied on the training data. Subsequently, the corresponding predicted values were compared with the true values. Finally, the performance of these models was evaluated by two indicators, namely MAE and RMSE. Ten samples were randomly selected from the test set for obtaining the mean of MAE and RMSE respectively. The univariate and multivariate models were compared and analyzed, followed by the traditional LSTM and the proposed hybrid CNN-LSTM model, and results show that multivariate models perform better results than that of univariate models, due to considering more air quality related features, and the MAE and RMSE of hybrid CNN-LSTM models are lower than that of LSTM model. The multivariate model should be selected if the amount of data is large and has multiple features while the univariate model can be considered if it has single feature. Meanwhile, the machine learning methods can be adopted if the amount of data is small. Moreover, the multivariate CNN-LSTM took about 50-60 seconds for one epoch, which was 40 seconds less than other models. So, the prediction effect of multivariate CNN-LSTM model is best in both low error and short training time. In the future, we will consider more relevant features for improving the accuracy of forecasting PM2.5 concentration.

REFERENCES

[1] J. H. Seinfeld, "Air pollution: A half century of progress," *AIChE J.*, vol. 50, pp. 1096–1108, Jun. 2004.

[2] R. D. Brook, S. Rajagopalan, C. A. Pope, J. R. Brook, A. Bhatnagar, A. V. Diez-Roux, F. Holguin, Y. Hong, R. V. Luepker, M. A. Mittleman, A. Peters, D. Siscovick, S. C. Smith, L. Whitsel, and J. D. Kaufman, "Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the american heart association," *Circulation*, vol. 121, no. 21, pp. 2331–2378, Jun. 2010, doi: 10.1161/cir.0b013e3181dbece1.

[3] R.-J. Huang *et al.*, "High secondary aerosol contribution to particulate pollution during haze events in China," *Nature*, vol. 514, no. 7521, pp. 218–222, Oct. 2014.

[4] J. Schwartz, D. W. Dockery, and L. M. Neas, "Is daily mortality associated specifically with fine particles?" *J. Air Waste Manage. Assoc.*, vol. 46, no. 10, pp. 927–939, Oct. 1996, doi: 10.1080/10473289.1996.10467528.

[5] J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer, "The contribution of outdoor air pollution sources to premature mortality on a global scale," *Nature*, vol. 525, no. 7569, pp. 367–371, Sep. 2015, doi: 10.1038/nature15371.
[6] J. L. Wang, "Chemical composition and quantitative relationship between meteorological condition and fine particles in Beijing," *J. Environ. Sci.*, vol. 16, no. 5, pp. 860–864, 2004.
[7] C. K. Chan and X. Yao, "Air pollution in mega cities in China," *Atmos. Environ.*, vol. 42, no. 1, pp. 1–42, Jan. 2008, doi: 10.1016/j.atmosenv.2007.09.003.
[8] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998, doi: 10.1098/rspa.1998.0193.
[9] Garima and S. Rani, "Review on time series databases and recent research trends in time series mining," in *Proc. 5th Int. Conf. Confluence Next Gener. Inf. Technol. Summit (Confluence)*, Sep. 2014, pp. 109–115.
[10] G. E. Box and G. M. Jenkins, "Time series analysis forecasting and control," *J. Time*, vol. 31, no. 4, pp. 238–242, 1976, doi: 10.1111/j.1467-9892.2009.00643.x.
[11] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transp. Res. Rec.*, vol. 1644, no. 1, pp. 132–141, Jan. 1998, doi: 10.3141/1644-14.
[12] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004, doi: 10.1023/b:stco.0000035301.49549.88.
[13] C. Haipeng, S. Ping, and H. Shengguo, "Study of aircraft hard landing diagnosis based on nerual network," *Comput. Meas. Control*, vol. 16, no. 7, pp. 906–908, 2008.
[14] S. Sun, C. Zhang, and G. Yu, "A Bayesian network approach to traffic flow forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 124–132, Mar. 2006, doi: 10.1109/tits.2006.869623.
[15] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: 10.1162/neco.2006.18.7.1527.
[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[17] A. Cleeremans, D. Servan-Schreiber, and J. L. Mcclelland, "Finite state automata and simple recurrent networks," *Neural Comput.*, vol. 1, no. 3, pp. 372–381, Sep. 1989, doi: 10.1162/neco.1989.1.3.372.
[18] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.
[19] L. Anghinoni, L. Zhao, D. Ji, and H. Pan, "Time series trend detection and forecasting using complex network topology analysis," *Neural Netw.*, vol. 117, pp. 295–306, Sep. 2019.
[20] L. Ding, W. Fang, H. Luo, P. E. Love, B. Zhong, and X. Ouyang, "A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory," *Automat. Construct.*, vol. 86, pp. 118–124, Feb. 2018.
[21] X. Shi, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Int. Conf. Algorithms Comput. Biol.*, 2015, pp. 68–80, doi: 10.1007/978-3-319-21233-3_6.
[22] E. Kanjo, E. M. Younis, and C. S. Ang, "Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection," *Inf. Fusion*, vol. 49, pp. 46–56, Sep. 2019, doi: 10.1016/j.inffus.2018.09.001.
[23] Z. Duan, Y. Yang, K. Zhang, Y. Ni, and S. Bajgain, "Improved deep hybrid networks for urban traffic flow prediction using trajectory data," *IEEE Access*, vol. 6, pp. 31820–31827, 2018, doi: 10.1109/access.2018.2845863.
[24] J. Zhao, F. Deng, Y. Cai, and J. Chen, "Long short-term memory–fully connected (LSTM-FC) neural network for PM2.5 concentration prediction," *Chemosphere*, vol. 220, pp. 486–492, Apr. 2019, doi: 10.1016/j.chemosphere.2018.12.128.
[25] C.-J. Huang and P.-H. Kuo, "A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, Jul. 2018, doi: 10.3390/s18072220.
[26] U. Pak, C. Kim, U. Ryu, K. Sok, and S. Pak, "A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction," *Air Qual., Atmos. Health*, vol. 11, no. 8, pp. 883–895, Oct. 2018, doi: 10.1007/s11869-018-0585-1.
[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 14539, pp. 436–444, 2015.
[28] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[29] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1D & 2D CNN LSTM networks," *Biomed. Signal Process. Control*, vol. 47, pp. 312–323, 2019, doi: 10.1016/j.bspc.2018.08.035.
[30] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks," *J. Sound Vib.*, vol. 388, pp. 154–170, Feb. 2017.
[31] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/tmi.2016.2528162.
[32] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, and D. J. Inman, "Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks," *J. Sound Vib.*, vol. 424, pp. 158–172, Jun. 2018, doi: 10.1016/j.jsv.2018.03.008.
[33] Ö. Yıldırım, P. Pławiak, R.-S. Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," *Comput. Biol. Med.*, vol. 102, pp. 411–420, Nov. 2018, doi: 10.1016/j.compbiomed.2018.09.009.
[34] S. Hochreiter and J. Schmidhuber, "Long short-time memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[35] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000, doi: 10.1162/089976600300015015.
[36] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[37] D. Yarotsky, "Error bounds for approximations with deep ReLU networks," *Neural Netw.*, vol. 94, pp. 103–114, Oct. 2017, doi: 10.1016/j.neunet.2017.07.002.

TAOYING LI received the B.S. and Ph.D. degrees from Dalian Maritime University, China. She is currently an Associate Professor with Dalian Maritime University. Her research interests are machine learning, big data, and data mining.

MIAO HUA received the B.S. degree from Liaocheng University, Shandong, China, in 2017. She is currently pursuing the M.S. degree with Dalian Maritime University, China. Her research focuses on deep learning.

XU WU received the B.S. degree from Liaocheng University, China. She is currently pursuing the M.S. degree with Dalian Maritime University, China. Her research focuses on machine learning.