

PAPER • OPEN ACCESS

Research of PM_{2.5} Prediction System Based on CNNs-GRU in Wuxi Urban Area

To cite this article: Haofei Xie *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **300** 032073

View the [article online](#) for updates and enhancements.

Research of PM2.5 Prediction System Based on CNNs-GRU in Wuxi Urban Area

Haofei Xie^{1, a}, Lin Ji^{1, b}, Quan Wang^{2, *} and Zhejian Jia^{1, c}

¹Chongqing University of Posts and Telecommunications, Chongqing 400065, China

²Wuxi Sensing Net Industrialization Research Institute, Wuxi 214000, China

*Corresponding author: seepwq@163.com, ^axiehf@cqupt.edu.cn,

^b183756002@qq.com, ^cjiazhejian@qq.com

Abstract. Aiming at the serious pollution situation and lack of effective prediction methods in Wuxi urban area, based on convolutional neural network (CNN) and gated recurrent unit (GRU), this paper proposes a PM2.5 prediction model that can automatically extract spatiotemporal features of multi-station and multimodal air quality data, and build a PM2.5 prediction system based on this model as well. The system model firstly takes multiple two-dimensional (2D) matrices constructed with time series of the air quality factors and weather factors from different monitoring stations in Wuxi urban area as input, automatically extracts and fuses the local variation trends and spatial correlation features of multi-station and multimodal data with CNNs structure. The results from the CNNs are input to the GRU network to further capture the long-term dependence feature of air quality data. Then, a fully connected network taking the spatiotemporal features as input is used to predict the PM2.5 concentration for the next 6 hours in Wuxi urban area. The PM2.5 prediction system based on CNNs-GRU model is tested on the real data set provided by Wuxi Environmental Protection Bureau. On the two test sets in January and June, the prediction accuracy of the PM2.5 prediction system reached 76.902% and 70.053% respectively, which is better than the comparative models. Finally, the prediction system based on the optimal CNNs-GRU model and real-time data obtained by crawlers, predicts the real-time PM2.5 concentration for the next 6 hours, and visualizes the prediction results on the Web through Echarts technology. It can provide valuable reference for citizens' travel, prevention and control of air pollution in Wuxi urban area.

1. Introduction

Air pollution has been a topic widely concerned by the whole society. Xing et al [1] suggested that long-term exposure to particulate pollutants increases the risk of cardiovascular disease, respiratory disease and cancer. The fine prediction of PM2.5 hourly concentration has become a hot research direction.

However, air quality changes are affected by a variety of complex factors [2-3], including climate, transportation, spatial distribution, physical and chemical processes, etc. Compared with the developed meteorological prediction [4-5], the accurate prediction of PM2.5 is more difficult. Currently, statistical models are frequently used in PM2.5 prediction, such as neural networks. Mahajan et al. [6] used shallow neural network to predict PM2.5 hourly concentration for the next hour Mahajan et al. [7] proposed an



exponential smoothing model based on historical data of single station PM2.5 to predict PM2.5 hourly concentration for the next hour and three hours. Biancofiore et al. [8], used the Recurrent Neural Network (RNN) to predict the PM2.5 daily concentration for the next 3 days. Although the above methods achieve certain effects in a specific scenario, there is a problem that time steps are too few or the time granularity is too large of the prediction, leading poor practical value. Moreover, the models rely too much on the artificial feature selection methods, which weakens the generalization ability. Neglecting the spatial correlation between adjacent stations in a region is not conducive to predicting the PM2.5 with strong diffusivity.

Related studies have shown that considering spatiotemporal relationships is critical for air quality analysis [9] [10]. Li et al. [11] proposed a spatiotemporal deep learning model STDL, which used a stacked automatic encoder to extract the features of air quality factors. Results showed that its prediction accuracy of PM2.5 hourly concentration for the next hour was higher than other traditional models that did not consider spatial features. It avoided artificial feature selection methods, but the input features and time steps of prediction were too few. Zheng et al. [12] proposed a hybrid model FFA, which used linear regression and ANN models to extract the spatiotemporal features of input data from different stations. This model could more accurately predict PM2.5 concentration for the next 6 hours than traditional models. Its shortcoming was it relied on artificial feature selection. And the spatial predictor only used the feature mean and median value of adjacent stations divided into regions as input, which reduced the sensitivity of the model to multi-station spatial features. Huang et al. [13] proposed a deep learning model APNet based on CNN and LSTM. CNN can automatically complete layer-by-layer abstraction of multiple feature sequences from a single station. The results showed that the prediction accuracy of PM2.5 hourly concentration through APNet was higher than CNN or LSTM alone. But this model did not consider the spatial relationship, and the prediction time steps were too few.

Based on the above analysis, taking Wuxi urban area as the object, this paper proposes CNNs structure based on CNNs and GRU. The CNNs, taking the multimodal data of 8 air quality monitoring stations in Wuxi urban area as input, automatically extracts and fuses the local variation trend and spatial correlation features of multi-station and multimodal data. It replaces the traditional artificial feature selection method, enhancing the feature extraction and generalization ability of the model. GRU can learn the long-term dependence features and complete the deep abstraction of spatiotemporal features. The fully connected network obtains the fine prediction result of the PM2.5 concentration for the next 6 hours. Finally, this paper evaluates the predictive performance of the CNNs-GRU and visualizes the real-time prediction results based on the web technology, constituting a complete prediction system.

2. Functional Design of PM2.5 Prediction System

Based on the above prediction process and modular design idea, functions of the PM2.5 prediction system in Wuxi urban area are mainly composed of four parts: acquisition of multi-station and multimodal air quality data, processing of multi-station and multimodal air quality data, PM2.5 prediction model based on CNNs-GRU, System model evaluation and visualization of prediction results. The main functions of each component are as shown in Fig.1.

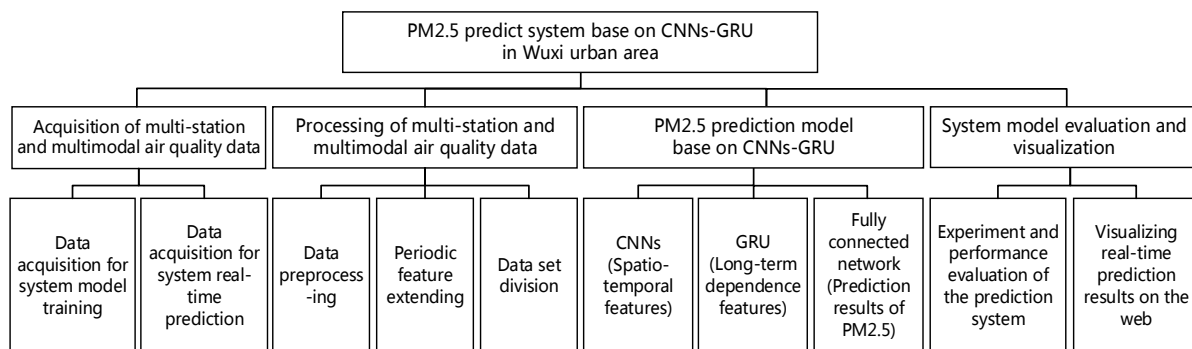


Figure 1. Functional composition of the PM2.5 prediction system in Wuxi urban area.

3. Acquisition of multi-station and multimodal air quality data

The PM2.5 concentration change is affected by many factors. The rich multimodal data contributes to the accurate prediction of PM2.5. Meanwhile, due to the complex network structure and numerous parameters, too little training data will lead to overfitting and decline in generalization of the model. Model training requires a large amount of historical data. This paper obtained historical data of air quality for 13 months from Wuxi Environmental Protection Bureau. The data is collected from 8 monitoring stations, including air quality factors and weather factors. A large.

In order to predict the PM2.5 concentration for the next 6 hours in Wuxi urban area, the system needs to obtain multi-station and multimodal air quality data from 8 monitoring stations in real time. This paper uses web crawler technology to crawl air quality data from PM2.5.in and China Weather Reuters in real time, stored in a local database.

4. Processing of multi-station and multimodal air quality data

In the process of collecting air quality data, there are data missing and abnormality. Part of the modal data uses a textual representation. It needs to preprocess the raw dataset, including data missing filling, abnormal data modifying, numerical encoding, and standardization. Air quality is often influenced by cyclical social activities and seasonal factors, showing periodic changes. Therefore, this paper extends a periodic feature according to different combinations of months and hours. Then, multi-station and multimodal air quality data needs to be divided according to a certain proportion into a training set, a verification set and a test set, used for the training and test of CNNs-GRU model.

5. CNNs-gru deep learning model

5.1. Overall architecture of CNNs-GRU model

As mentioned above, the PM2.5 prediction system adopts the deep learning model CNNs-GRU. The overall structure is shown in Fig.2, consisting of the following three parts:

(1) The CNNs structure consists of many one-dimensional CNN (CNN1D) units connected in parallel and cascaded. The parallel connected CNN1Ds correspond to an equal number of multi-station and single modal air quality data units. It automatically extracts local variation trends and spatial correlation features of the inputs. The cascaded CNN1D layers perform feature level fusion and deeper feature abstraction, and outputs the fusion spatiotemporal features.

(2) GRU network, composed of several GRU layers, captures the long-term dependence features of multi-station and multimodal air quality data, so that it can memory long-term dependence historical information.

(3) Fully connected network, including input vectorization layer (Flatten) and several full connected layers (Dense), outputs the prediction results of PM2.5 concentration for the next 6 hours at a monitoring station in Wuxi urban area.

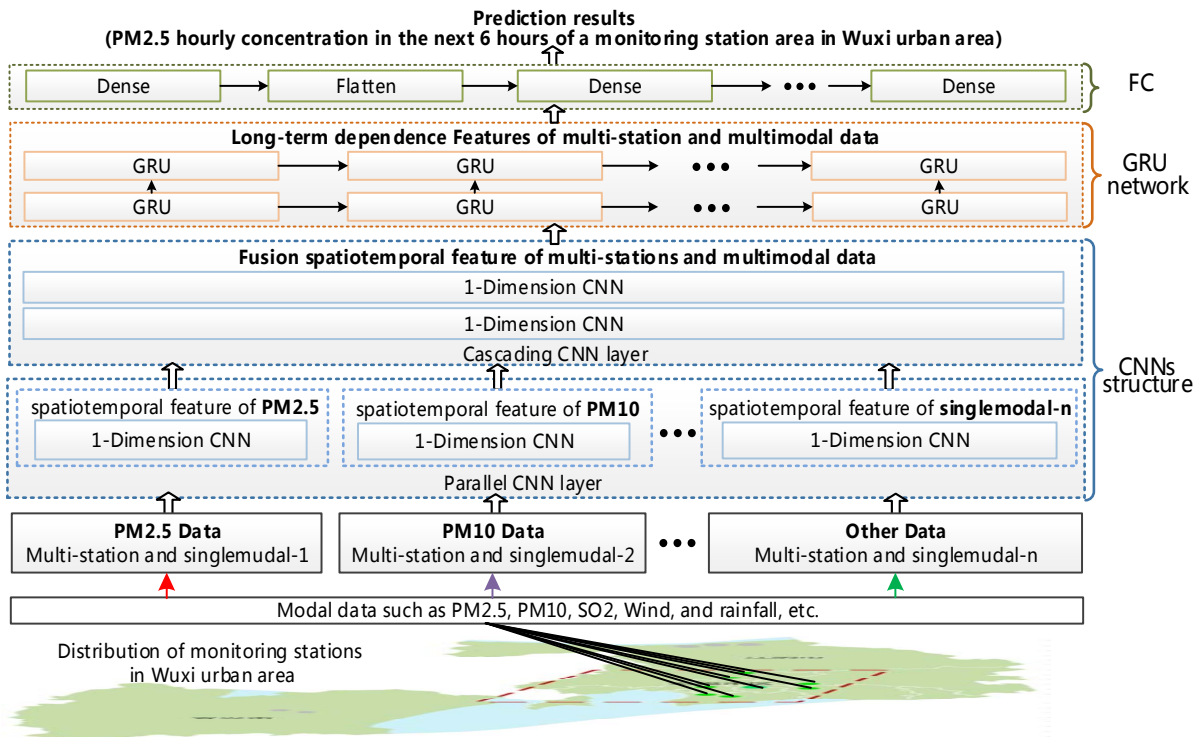


Figure 2. Overall architecture of CNNs-GRU model

5.2. CNNs structure

The CNNs structure consists of many CNN1D units connected in parallel and cascaded. The structure is input with multiple independent 2D matrices of multi-station and single modal air quality, each 2D matrix consists of time series of air quality attributes from 8 monitoring stations in Wuxi urban area. Each CNN1D connected in parallel can automatically extract local variation trends and spatial correlation features of multi-station and single modal air quality data through convolution operations of convolution kernels. Cascaded CNN1Ds taking the abstract spatiotemporal features of multiple independent multi-station and single modal data as input, realizes feature level fusion through linear merge operation, and further obtains deeper abstract fusion spatiotemporal features of air quality data. The operation flow of CNNs is shown in Fig.3.

The core component of the CNNs structure is the CNN1D, which contains several 1D convolution kernels. Taking PM2.5 data collected from 8 monitoring stations in Wuxi urban area as an example, Fig.4 shows how the CNN1D automatically extracts the spatiotemporal features of multi-station and multimodal data.

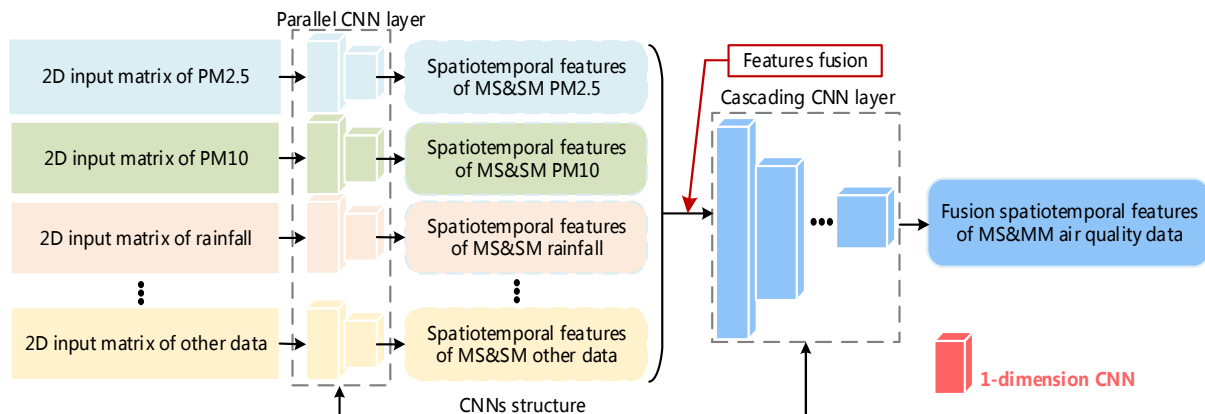


Figure 3. CNNs flow chart (MS&SM: multi-station and single modal, MS&MM: multi-station and multimodal)

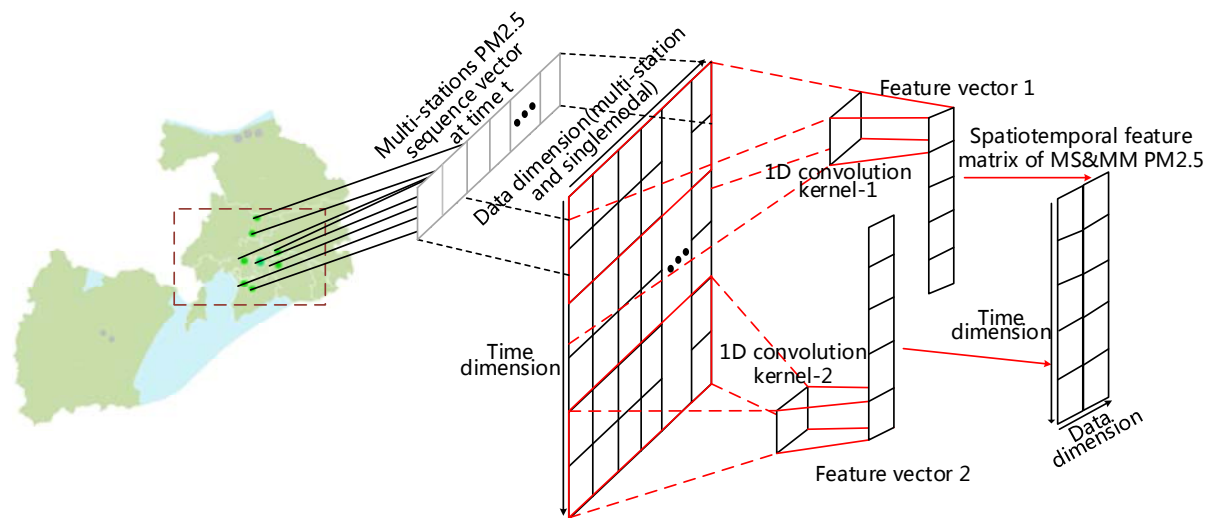


Figure 4. Schematic diagram of the CNN1D convolution principle (MS&MM represents multi-station and multimodal)

The length of the 1D convolution kernel indicates the size of the local receptive field, corresponding to the number of multi-station PM2.5 sequence vectors covered along the time axis. The width default covers a single modal PM2.5 sequence from 8 monitoring stations in Wuxi urban area, with a span equal to the number of station. The convolution kernel can learn the spatial correlation features between multi-station data by assigning different values to connections between the convolution kernel and different station inputs. By sliding along the time axis, it automatically extracts the local variation trends of multi-station and single modal PM2.5 in the time dimension. A plurality of 1D convolution kernels can learn multiple feature vector representations from different aspects. Multiple feature vectors are combined along the data dimension, so that it can obtain spatiotemporal features of multi-station and single modal inputs.

To provide an in-depth look at the complete workflow of the CNNs structure, the following is a formulation derivation process. Assuming that there are N sets of samples, M air quality attributes and P monitoring stations. The 2D input matrix unit composed of multi-station and single modal n can be expressed as:

$$\mathbf{X}_n = [\mathbf{x}_n^1, \mathbf{x}_n^2, \dots, \mathbf{x}_n^N]^T \quad (1)$$

$$\mathbf{X}_n^{t:t+T-1} = [\mathbf{x}_n^t, \mathbf{x}_n^{t+1}, \dots, \mathbf{x}_n^{t+T-1}]^T \quad (2)$$

Where $\mathbf{x}_n^t = [x_{n,1}^t, x_{n,2}^t, \dots, x_{n,p}^t] \in R^{1 \times P}$ represents the time series vector of multi-station and single modal air quality attribute n at time t . $\mathbf{X}_n^{t:t+T-1}$ represents T vectors of \mathbf{X}_n in the time interval of $[t, t+T-1]$. \mathbf{T} is the transposition.

If each CNN1D contains V 1D convolution kernels, the i_{th} convolution kernel k_i is denoted as $\mathbf{W}_i \in R^{T \times P}$. T Represents the length of processing time series of the convolution kernel along the time axis every time. The convolution process can be expressed as:

$$a_{t+T-1,n}^i = \sigma(\text{dot}(\mathbf{W}_i, \mathbf{X}_n^{t:t+T-1}) + \mathbf{b}) \quad (3)$$

Where σ is the activation function, this paper uses ELU. The convolution kernel k^i slides along the time axis, assuming stride is 1, obtaining the feature vector \mathbf{a}_n^i with $[N-T+1] \times 1$ size. The convolution results of the V convolution kernels are linearly merged along the feature direction into \mathbf{A}^n with $[t, t+T-1] \times V$ size. \mathbf{A}^n Represents the spatiotemporal feature matrix corresponding to multi-station and single modal air quality attribute n .

$$\mathbf{a}_n^i = [a_{t+T-1,n}^i, a_{t+T,n}^i, \dots, a_N^i]^T \quad (4)$$

$$\mathbf{A}^n = [\mathbf{a}_n^1, \mathbf{a}_n^2, \dots, \mathbf{a}_n^V], j \in [1, M] \quad (5)$$

The above formula completes the spatiotemporal feature abstraction of multi-station and single modal air quality input (such as PM2.5). Because there are M air quality attributes, CNNs structure firstly applies the above process to other modal through parallel CNN1D, in order to automatically extract spatiotemporal feature of multiple multi-station and single modal input. Then, feature level fusion is performed to obtain a fusion spatiotemporal feature matrix \mathbf{A} of multi-station and multimodal air quality data, as shown in equation (6). Finally, the cascaded CNN1Ds take \mathbf{A} as input and extract more abstract fusion spatiotemporal features according to equations (1)-(5).

$$\mathbf{A} = [\mathbf{A}^1, \mathbf{A}^2, \dots, \mathbf{A}^M] \quad (6)$$

5.3. GRU Network

GRU is optimized based on LSTM, including two gate structures, an update gate and a reset gate. It can learn the long-term dependence features of time series. In this paper, GRU is used to extracts the implicit long-term dependence features of multi-station and multimodal air quality data, which is helpful for PM2.5 prediction.

6. System model evaluation and visualization

The CNNs-GRU model of the PM_{2.5} prediction system is implemented using Keras based on the Tensor Flow backend. This paper estimates the accuracy of the system model with matplotlib and three evaluation metrics. Finally, the Echarts library is used to visualize the real-time prediction results of the PM_{2.5} prediction system.

6.1. Experiment setup of CNNs-GRU model

The proposed CNNs-GRU is an 8-layer deep learning model. The network structure settings are shown in Table 1. Conv1D, GRU, and Dense represent the 1D CNN layer, GRU layer, and fully connected layer in TensorFlow.

Table 1. Structure parameter setting of CNNs-GRU.

Layer number	Type	Number of neurons	Activation function
1,2,3	Conv1D	256,128,128	ELU
4,5	GRU	128,64	Tanh
6,7,8	Dense	256,64,6	Tanh*2,Linear

The prediction system model was trained using the Adam. In order to promote the rapid convergence and reduce the overfitting of the model, optimization methods of early stop (min_delta=0.001, patience=10), dropout (dropout rate=0.35), and batch normalization (default parameters) are used. Other training settings: epoch is 150, time step is 12, and batch size is 64. The proposed system model uses mean squared error (MSE) as loss function. In this paper, the three metrics are used to evaluate the comparative models, which are RMSE, MAE and MAPE. Finally, the proposed CNNs-GRU model is compared with five comparative models of ARIMA, shallow neural network, LSTM, APNet (LSTM) and APNet (GRU).

6.2. Experiment data

The system relies on an environmental protection research project in Jiangsu Province. The training data of the system model is provided by Wuxi Environmental Protection Bureau, including air quality factors (PM_{2.5}, PM₁₀, SO₂, NO₂, CO, O₃) and weather factors (temperature, humidity, wind speed, wind direction, rainfall, and air pressure). The data collection time is from May 2016 to June 2017, and the sampling period is hour. The data are collected from 8 monitoring stations in Wuxi urban area. The real-time air quality data is acquired by the crawler.

The preprocessed multi-station and multimodal data can be used to train and test the model. In order to fully verify the predictive performance of the CNNs-GRU model under different climatic conditions, the experiment divides the data in January and June 2017 into two separate test sets. 80% of the remaining data is used for model training and 20% is used as verification set. Taking Wang Zhuang Station as the prediction area, the prediction model uses 13 attributes of 8 monitoring stations for model training, then finely predict the PM_{2.5} concentration for the next 6 hours.

6.3. Experimental results and analysis

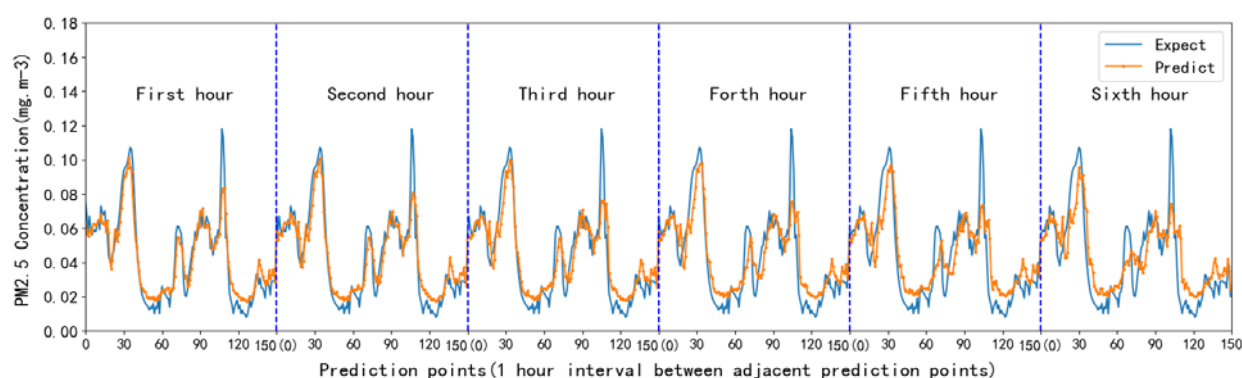
Table 2 shows the experimental results of the PM_{2.5} prediction system based on CNNs-GRU model and five comparative models. It can be found that in the two test sets, the accuracy of the prediction system based on CNNs-GRU model proposed in this paper is as high as 76.902% and 70.053% respectively, which is 8.8% and 5.1% higher than the comparative model APNet with the best performance.

Table 2. Comparison of PM2.5 predict performance metrics for the next 6 hours.

Model	RMSE		MAE		100-MAPE	
	January	June	January	June	January	June
ARIMA	29.983	15.977	20.933	11.740	47.616	48.963
Shallow Neural Network	23.655	13.099	16.912	9.679	59.061	58.656
LSTM	21.875	12.935	15.686	9.375	62.106	59.586
APNet (LSTM)	19.331	11.7184	13.993	8.680	66.665	64.840
APNet (GRU)	19.020	11.803	13.970	8.754	68.457	63.870
CNNs-GRU	15.570	9.577	10.720	7.068	76.902	70.053

From the analysis of Table 2, it can be seen the prediction system based on ARIMA has the worst effect. The system based on neural network can better mine the complex nonlinear relationship hidden in the PM2.5 variation trends, of which the overall performance is better. The prediction system based on the LSTM (GRU) is better than the prediction system based on the shallow neural network. It shows that the LSTM (GRU) with memory function can fully exploit the long-short term dependence features of the time series, which is helpful for prediction. The prediction system based on the fusion model APNet obtains higher accuracy than LSTM alone, proving that using CNN to replace the traditional artificial feature selection method can reduce the prediction error of PM2.5. APNet is modeled based on single-station and multimodal data, regardless of spatial correlation features of air quality data at different stations in a region. CNNs-GRU uses multi-station and multimodal air quality data from Wuxi Urban Area as input. The comparative results of the two models demonstrate that CNNs-GRU has a better prediction effect.

Figure 5 and 6 show the real prediction effects of the PM2.5 prediction system based on optimal CNNs-GRU on the two test sets. In order to better show the local details, only the prediction results of 150 sets of test data are shown. From the graph analysis, all of them have obtained ideal prediction results. The accuracy of the latter few hours was lower than that of the previous, but the overall performance is good. Among them, the following characteristics of PM2.5 predicted value are better on January test set, because the data has relatively stable distribution and higher average value. However, due to the characteristics of local violent oscillation and small mean value on the June test set, the prediction system can predict the sequence change trend relatively accurately, whose local performance in the rapid change is slightly lacking. In generally, its performance is within the acceptable range.

**Figure 5.** 6-hour prediction comparison of optimal CNNs-GRU on January test set

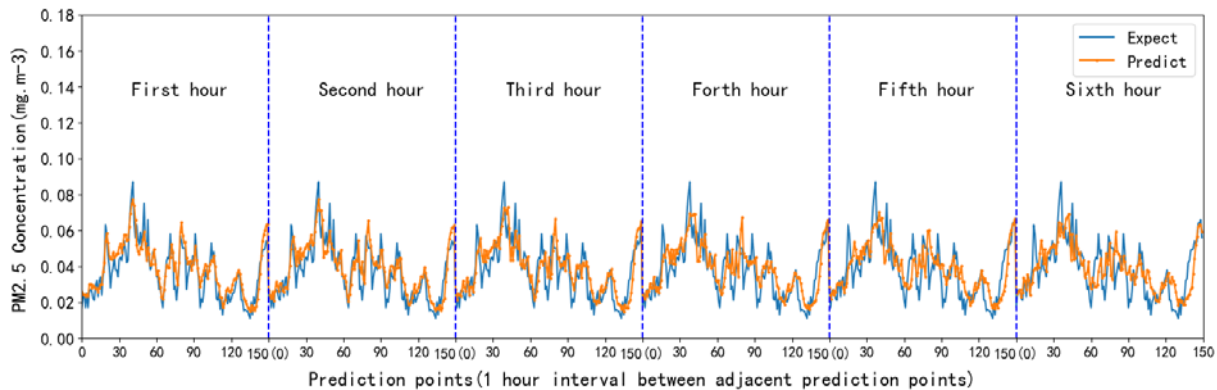


Figure 6. 6-hour prediction comparison of optimal CNNs-GRU on June test set

6.4. Real-time prediction visualization based on Echarts

According to the experimental results in Section 5.4, the PM2.5 prediction system based on CNNs-GRU model proposed in this paper can obtain the highest accuracy. Therefore, based on optimal CNNs-GRU model and real-time air quality data acquired by the web crawler, we can obtain the PM2.5 hourly concentration for the next 6 hours in Wuxi urban area. Finally, Echarts technology is used to visualize the prediction results on the Web, as shown in Fig.7.

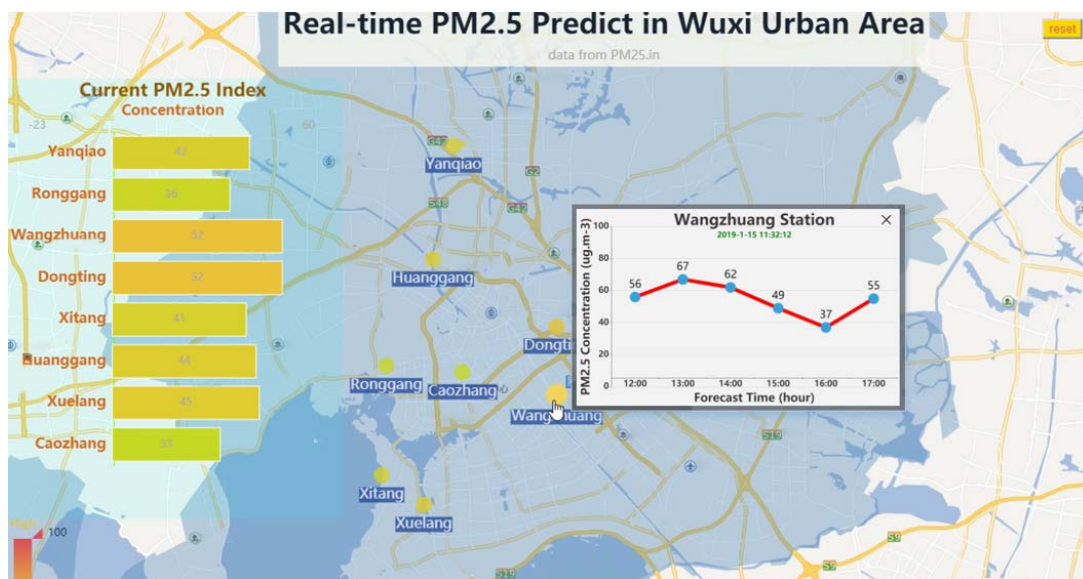


Figure 7. Visualization of prediction results based on Echarts

7. Conclusion

This paper proposes a deep learning model CNNs-GRU based on CNNs and GRU. The CNNs structure can automatically extract and fuse the local variation trend and spatial correlation features of multi-station and multimodal air quality data. GRU network is used to learn the long-term dependence of air quality data. Experimental results show the accuracy of the prediction system based on the CNNs-GRU model is as high as 76.902% and 70.053% on two test sets respectively, which is 8.8% and 5.1% higher than the comparative model APNet with the best performance. Meanwhile, the RMSE metric is 15.570 and 9.577 respectively, and the MAE metric is 10.720 and 7.068, which are better than all comparative models. It is proved that CNNs-GRU has superior performance in fine prediction of PM2.5 concentration.

Acknowledgments

This work was financially supported by National Key Research and Development of China. (No. 2017YFE0123000).

References

- [1] Y.F. Xing, Y.H. Xu, M.H. Shi, and Y.X. Lian, *Thoracic Disease*, 8, 49 - 74 (2016).
- [2] H.J. Chu, C.Y. Lin, C.J. Liao, and Y.M. Kuo, *Atmosphere Environment*, 60, 142 - 152 (2012).
- [3] A.P.K. Tai and D.J. Jacob, *Atmosphere Environment*, 44, 3976 - 3984 (2010).
- [4] G. Aditya, A. Kapoor, and E. Horvitz, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (ACM, USA, NY, 2015), pp. 379–386.
- [5] G.Q. Zhou, P. Li, F.H. Geng, J.M. Xu, F. Yang, and X.X. Tie, in *2012 IEEE Symposium on Robotics and Applications (ISRA)*, (IEEE, Malaysia, Kuala Lumpur, 2012), pp. 793 – 796.
- [6] S. Mahajan, L.J. Chen, and T.C. Tsai, in *2017 IEEE Smart World Congress*, (IEEE, USA, San Francisco, 2017), pp. 327 – 333.
- [7] J. Mahajan, L.J. Chen and T.C. Tsai, *Sensors* 2018, 18, 3223 - 3237 (2018).
- [8] F. Biancofiore, M. Busilacchio, M. Verdecchia, B. Tomassetti, E. Aruffo, S. Bianco, S.D. Tommaso, C. Colangeli, G. Rosatelli, and P.D. Carlo, *Atmosphere Pollution Research*, 8, 652 - 659 (2017).
- [9] B.S. Beckerman, M. Jerrett, M. Serre, R.V. Marting, S.J. Lee, D. Van, Z. Ross, J. Su, and R.T. Burnett, *Environment, Science & Technology*, 47, 7233 - 7241 (2013).
- [10] H.L. Yu and C.H. Wang, *Atmosphere Environment*, 44, 3053 - 3065 (2010).
- [11] X. Li, L. Peng, Y. Hu, J. Shao, and T.H. Chi, *Environmental Science and Pollution Research*, 44, 22108 - 22417 (2016).
- [12] Y. Zheng, X.W., M. Li, R.Y. Li, Z.Q. Shan, E. Chang, and T.R. Li, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (ACM, USA, NY, 2015), pp. 2267 – 2276.
- [13] C.J. Huang and P.H. Kuo, “A Deep CNN-LSTM Model for Particulate Matter (PM_{2.5}) Forecasting in Smart Cities”, *Sensors* 2018, 18, 2220 - 2241 (2018).