



An air quality forecasting model based on improved convnet and RNN

Baowei Wang^{1,2,3} · Weiwen Kong¹ · Peng Zhao¹

Accepted: 26 April 2021 / Published online: 10 May 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

With the development of quality of life, people pay more and more attention to the surrounding environmental factors, especially air pollution. The problem of air pollution in China is becoming more and more serious, which poses a great threat to people's health. Therefore, the prediction of air quality concentration is very important. PM_{2.5} is the primary indicator for evaluating the concentration of smog. Currently, studies have been proposed to predict the concentration of PM_{2.5}. Before, most methods use traditional machine learning or real-time monitoring to predict pollution of PM_{2.5} value. However, the previous prediction methods cannot meet the requirement of the accuracy. For this end, this paper uses Convnet and Dense-based Bidirectional Gated Recurrent Unit to predict PM_{2.5} value which combined Convnet, Dense and Bi-GRU. The feature in air quality data was extracted from convnets without max-pooling instead another convolutional layer and Bi-GRU with additional Dense could provide a more accuracy result. Experiments show that the effectiveness of our method PM_{2.5} mass concentration prediction model provides a more superior method for PM_{2.5} mass concentration prediction.

Keywords Air pollution forecasting · Recurrent neural network · Convolutional neural networks · Bidirectional gated recurrent unit · Dense

1 Introduction

Today, PM_{2.5} is becoming more and more harmful to the human body. It causes a drop in air quality. The air contains a large amount of harmful substances that are most likely to affect the health of the respiratory system (Feng et al. 2015). Besides, the body's respiratory system is exposed to the air in a long time. Therefore, air quality can inhale more particulate matter, high virus concentration, and easily cause many diseases (Wang et al. 2019a). In recent years, smog weather has frequently affected people's health in many provinces

and cities across the country. Haze pollutants are prone to cardiovascular disease. The smog reduces the visibility and increases the probability of traffic accidents, which affects people's travel. Haze affects people's production and life. Haze pollution has become one of the most serious problems in the world. It causes serious air quality problems.

The air quality prediction is to describe and analyze the future air pollution status and environmental quality trends and the dynamic changes of major pollutants and pollution sources (Tao et al. 2019). Before, the data and prediction all come from the sensor and internet of things (Al-Janabi 2020; Al-Janabi et al. 2020b). It provides a basis for proposing countermeasures to prevent further deterioration of the environment and improve the environment. Therefore, it is particularly important to predict the smog index in a timely manner so that people can take preventive measures in a timely manner to minimize losses. Today, air quality issues are becoming more and more important, and providing accurate environmental predictions can be a very important issue (Wang et al. 2019d). Therefore, the prediction of air quality data is very necessary.

Therefore, it is particularly important to predict the air quality in a timely manner so that people can take preventive measures in a timely manner to minimize losses (Qian-Rao 2016; Al-Janabi and Alkaim 2020). Due to the

✉ Baowei Wang
wang@nuist.edu.cn

Weiwen Kong
kongweiwen@nuist.edu.cn

Peng Zhao
zhaopeng_159@163.com

¹ School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China

² Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), NUIST, Nanjing 210044, China

³ Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing 210044, China

variety of sources of air pollution, i.e., automobile exhaust, volcanic eruptions, industrial emissions, air pollution is usually formed by a mixture of various sources of pollution. Moreover, in the smoggy weather in various regions, the extent of the effects of different sources of pollution varies. The atmospheric system is very complex, and various factors in the atmosphere have a certain correlation (Lonati et al. 2005). It is found by factor analysis to find out which factors in the atmosphere are related to the smog index. Therefore, smog is difficult to predict.

At present, the methods of air quality prediction can be roughly divided into three categories (Yan et al. 2018). The first is statistical methods. The second is the traditional forecasting method. The third is the deep learning algorithm (Li and Xu 2018; Al-Janabi et al. 2020a). Statistical methods include linear regression, grey prediction, Markov prediction, and so on. Most statistical models have certain requirements for data, and the models also have relatively clear mathematical forms. It is difficult to describe data with complex components with limited mathematical formulas. The traditional method is characterized by simple method, mature theory and easy implementation. It only needs to search historical statistics of various indicators. Traditional analysis (Al-Janabi and Abaid Mahdi 2019) makes it easy to analyze the future development of data, especially short-term trends. The deep learning algorithm can establish the intrinsic mapping relationship between environmental indicators and influencing factors, and its prediction accuracy is significantly higher than the traditional analysis method and can be accurate to a specific date.

To this end, a Convnet and Dense-based Bidirectional Gated Recurrent Unit (CDBGRU) is proposed, which got a better performance on air quality data than other methods. It used a rebuilt convnet to replace traditional convnet in which another convolutional layer is used instead of the max-pooling layer. To sum up, the main contributions of this paper are as follows:

1. This paper proposed a Convnet and Dense-based Bidirectional Gated Recurrent Unit (CDBGRU). In our model, a rebuilt convnet, Bi-GRU and additional Dense were used to predict a more accuracy air quality value.
2. We verify the effectiveness and performance advantages of the proposed method through a large number of comparative experiments on the data in Beijing air quality data from 2018-01-01 to 2019-07-01.

This article is introduced from these sections: Sect. 2 introduces the relevant related to this article. Section 3 introduces the research methods and principles of this paper. Section 4 shows experimental results and forecast accuracy. Finally, we give the article summary and future prospects in Sect. 5.

2 Related work

Air quality study has a long history. In the past, the existing method forecasting the air quality always focused on statistical method. Afterward, shallow machine learning methods were proposed to solve the forecasting problem, such as SVR (Drucker et al. 1997), DTR (Xu et al. 2005) and GBR (Huang and Oosterlee 2011). As a common classifier, SVR achieves superior performance on many scenarios. Generally, SVR can be divided into three classes, i.e., linear-SVR, poly-SVR and rbf kernel SVR. Specially, rbf kernel SVR often has the best performance on complex data, for example, air quality data. DTR is based on Decision Tree, and usually we use CART Decision Tree for regression due to the inconvenience on ID3 Trees and C4.5 Trees. GBR is a boosting model which is a technique for learning from its mistakes. In essence, it is to gather ideas and integrate a bunch of poor learning algorithms for learning. Additionally, many researches consider the link of each feature in air quality data, for example, (Deleawe et al. 2010) expanded machine learning and put it into predicting CO₂ level, which gives the inspiration of prediction method.

However, with the study on deep learning (Hao et al. 2018) and big data analyze (Wang et al. 2019c), more and more researches considered how to combine the deep learning (Singh and Srivastava 2016) model with air quality data. Because air quality data have dynamic, nonlinear and, especially, time series-related characteristics, more and more researches focus on data-driven models. So far, a large number of air quality forecasting methods based on the big data analyze (Najafabadi et al. 2015) have been proposed to predicted each feature and pollution value. For example, (Zheng et al. 2013) proposed a semi-supervised learning approach which is based on a co-training framework consisting of ANN and CRF to predict PM_{2.5}.

In recent years, deep learning model especially convolution neuron networks and recurrent neuron networks has changed great work. As a time series predicted model, RNN (Zaremba et al. 2014) has a good performance on air quality forecasting for air quality data is a typical time series data. So far, a large number of derivatives of RNN have appeared. In order to deal with gradient disappearance in RNN, LSTM (Gers et al. 1999) and GRU was proposed which has been widely used in industry. More than this, the derivatives of LSTM and GRU also have appeared, such as Bi-directional GRU (Schuster and Paliwal 1997), which uses the backward of the data to smooth the nonlinear characteristics. CNN (Zhang et al. 2018; Long and Zeng 2019; Pan et al. 2019) is a good appliance to extract features and (Tao et al. 2019) combine CNN and Bi-GRU to predict the value of PM_{2.5} called CBGRU which has a good performance. However, the methods above still have space to improve. So, CDBGRU was proposed.

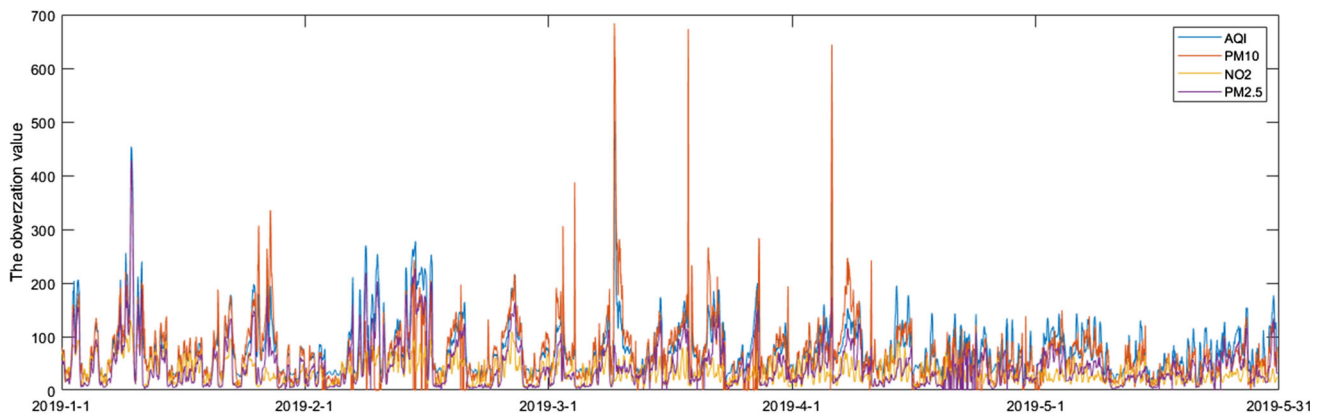


Fig. 1 Some typical air pollution features

3 Proposed method

In this section, we will propose a method to measure the value of PM_{2.5} called Convnet and Dense-based Bidirectional Gated Recurrent Unit (CDBGRU) and introduce each part of the method.

3.1 Problems and motivations

Nowadays, air quality forecasting has become a popular research topic in control of urban and rural air pollution. The goal of air quality forecasting is to predict the current time change of PM_{2.5} value at the observation point. The observation time interval is usually set for one or two hour, which is decided by the ground-based monitoring station (Wang et al. 2019b). Figure 1 shows the typical air pollution data such as PM_{2.5}, PM₁₀, AQI, NO₂.

PM_{2.5} prediction problem is usually illustrated as follows. Assuming time T and the air quality index P_T , the goal of prediction is to predict the PM_{2.5} concentration value P_{T+1} at time $T+1$ or P_{T+n} at time $T+n$. Usually, the goal achieved accurate result by modeling the history air quality-related time series dataset $\{P_t | t = 1, 2, \dots, T\}$, where P represents the history air quality index including PM_{2.5} and other air quality-related time series data such as NO₂, temperature, SO₂, etc. The difficulty of prediction is how to combine the trend of historical data with the relationship of time series. To this end, CDBGRU was proposed.

3.2 Our proposed method

3.2.1 Overview of CDBGRU

To solve the above problems, in this paper, an air quality forecasting method called Convnet and Dense-based Bidirectional Gated Recurrent Unit (CDBGRU) predicting PM_{2.5} is proposed. In general, due to the different representations

and data structures in different time, the statistical characteristics of air quality related time series are discrepant, in which shallow machine learning models cannot have a good performance to deal with complex scenarios. In recent years, hybrid deep learning model has also been used on predicting PM_{2.5} value, which achieved more effective results than those of classic machine learning models.

CDBGRU is a combination of CNN, additional full connecting layer and Bi-directional GRU that considers the spatial temporal dependence of air quality-related time series data. The model consists of three part. In the first part, the one-dimensional convnet without max-pooling performs local feature learning and dimensionality reduction in input variables. Compared with traditional convnets, the lack of max-pooling instead another convnet have a better performance on disposing the original data to emerge low-dimensional feature sequences. In the second part, the feature sequences are fed into the bidirectional GRU, which is an improvement on LSTM and the parameters in reset gate and update gate are constantly adjusted in the process of training, then the relationship between the features extracted from the convnets are learned from the time series data. At the end of the model, two fully connected layers are used and the last layer contains only one channel to predict the PM_{2.5} concentration value. The innovation of this method is to use one-dimensional convnet as the preprocessing process before GRU, and use an additional layer of convolution network to replace the max-pooling layer which combines the local feature extraction with the time series prediction ability of GRU. On the other hand, through bidirectional processing sequence, bidirectional GRU can capture patterns that may be ignored by unidirectional GRU. Figure 2 shows the architecture of our deep model.

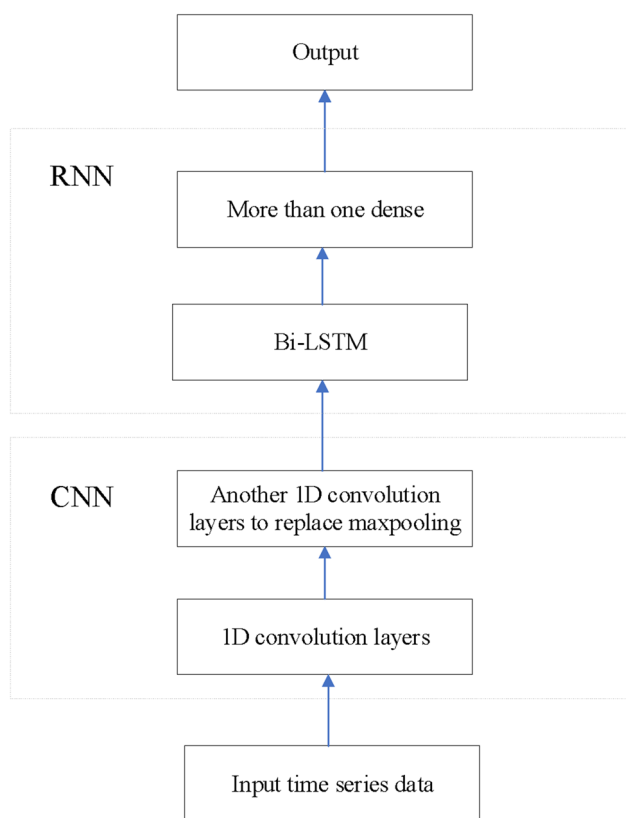


Fig. 2 The architecture of CDBGRU

3.2.2 1D convnet for local trend features learning

CNN not only has great work in image processing but also can be used for air quality data processing and time series analysis. The weight sharing features and the local perception of CNN can accelerate the learning of parameters for processing multi-variable time series, so as to improve learning efficiency. CNN usually has three layers: Convolutional layer, Activation layer and Pooling layer. The 1D convnets can be utilized for learning local trend features. Convnets can perform convolution operations to extract features from local input trend features, which allows modularity representation and data efficiency. These characteristics make convnets excellent in image processing and suitable for time series sequence processing. In this paper, we regard time series as a spatial dimension. The local perception and weight sharing of convnets can reduce the difficulty of handling in dealing with multiple time series and improve the learning efficiency. The patterns learned in one position at one sequence can be recognized at other positions in the future, because the patterns learned in the same position can be recognized in other positions and input conversion can be performed for each subsequence which is summarized as time shift invariants of convnets.

As shown in Fig. 3, with convolution kernel windows used in each convolution layer to process meteorological and PM2.5 time series, sequence fragments can be learned within one window size, and all these subsequences can be identified in the whole time series, so that the local trend feature change characteristics of multivariate time with time can be captured. In the general convolution network, after 1D convolution operation, another convolution layer instead of max-pooling layer was used for secondary sampling to output the subsequence extracted from the input time series because the weather data do not need to be desampled after convolution, and the sample does not require high translation invariance, so this paper chooses CNN + Relu with strip = 2 to replace the original pooling layer, which not only reduces the length of the input time series, but also retains most of the characteristics of the air pollution data.

3.2.3 Bidirectional gated recurrent unit for time series forecasting

In our model, bidirectional Gated Recurrent Unit (BiGRU) is shown in Fig. 4, which is often used for predicting time series process. It is well known that RNN is a distinct kind of neural network developed for processing sequential data. However, due to the disadvantages of RNN such as vanishing gradient and explosion gradient, it is difficult to learn long-term-dependent tasks such as long time series air quality data. To solve the problem, LSTM, GRU and their derivatives are proposed. The LSTM keeps track of long-term information by including gates (input gate, forget gate and output gate). And GRU is an improved version of the LSTM, which can also learn long-term dependencies. Different from LSTM, there are no memory unit, but 2 gates which is update gate and reset gate instead of 3 gates in GRU. Compared with LSTM, GRU has more simple architecture, requires less computation and is trained faster.

In the cell of GRU, the function of update gate is to decide which information can be retained and then delivered to the next state, and the reset gate represents how to combine the previous state information with the new input information. The criteria of state update for the next output and state value in the GRU unit are as follows.

$$z_t = \sigma(W_z * [x(t), h(t-1)]) \quad (1)$$

$$r_t = \sigma(W_r * [x(t), h(t-1)]) \quad (2)$$

$$\hat{h}(t) = \sigma(W_h * [x(t), (r_t * h(t-1))]) \quad (3)$$

$$h(t) = (1 - z_t) * h(t-1) + z_t * \hat{h}(t) \quad (4)$$

Where σ is the activation function, $x(t)$ is the input, $h(t-1)$ is the previous output, W_z , W_r and W_h are the weights of the update gate, reset gate, and candidate output.

Fig. 3 1D convnets processing time series

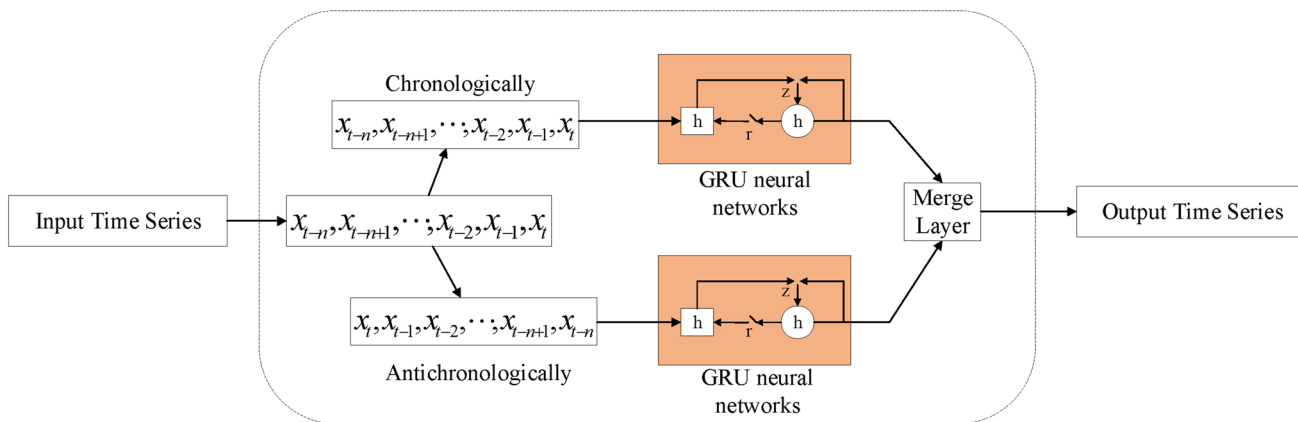
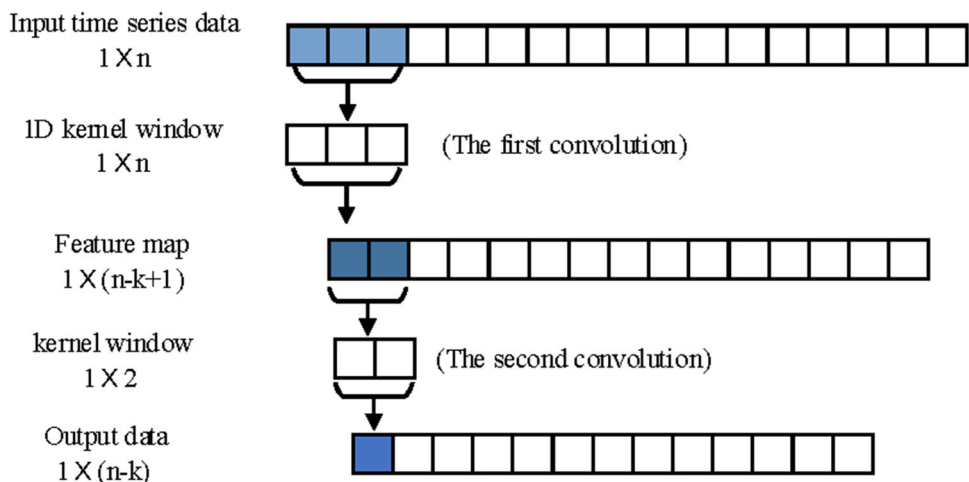


Fig. 4 Bidirectional GRU

The Bi-GRU consists of two GRUs, one of which processes a chronological input time series data and the others process an anti-chronological time series data, and then combine their representations in a total state. Features and time series data such as air quality data and meteorological data are subject to the distribution of some continuous function, in which we can fit a function from the previous data through the observation values to predict the future data. In the same way, future data can be used to a function to predict the value of the previous moment, which is often used on back propagation. For time series forecasting tasks, in previous research, only historical data can provide predictive power due to the continuous distribution; however, bidirectional training model can provide more useful information in modeling for the reverse continuity. By viewing air quality data from chronological and anti-chronological input enables the model to get more accurate representations and capture patterns that may be ignored when using chronological GRU, thereby improving the performance of ordinary GRU.

3.2.4 Full connected layer (dense)

Full connected layer often plays the role of Classifier or Regressor in the whole network. It is to map the learned distributed feature representation to the sample marker space. Using more than one full connected layer can improve the accuracy of our model due to so many parameters in these layers.

3.3 Discussion about limitation

This paper proposed CDBGU algorithm to predict the PM2.5 value by advanced neural network. However, there are some limitation on this model.

1. A gluttonous neural network will be enhance the loss of time. To some extent, the network of this paper is not complex. Compared with previous methods, our network has too much parameters, which increase the calculation time

to some degree. However, in this paper, it is worthwhile to sacrifice time complexity for more accurate results.

2. 1D convent is used to deal with nonlinear features. Before that, most methods used correlation coefficient to screen air quality data. Compared with the previous selection strategy, 1D convent may have advantages, but there is no evidence that correlation coefficient is worse than 1D convent. Parameters of high linearity from correlation coefficient selection will be also achieve good results.

4 Experiment

In this section, we will perform experiments on real air quality data to evaluate the proposed method. By comparing the classic shallow learning model, deep learning model and our model, the prediction performance and effectiveness of the model are verified.

4.1 Dataset

The experience in this paper uses Beijing air quality dataset which includes Pm2.5 value data and other feature such as date, time, temperature, humidity, wind speed, wind direction. The data interval in the dataset is one hour, and dataset used for experiment is ranged from 2018-01-01 to 2019-07-01.

4.2 Setup

4.2.1 Error measure

We choose Root Mean Square Error (RMSE) as the loss function, while RMSE can better reflect the true situation of the prediction mistake. In addition, R square is used as the error evaluation index of the model to evaluate the change degree and accuracy of the data and to measure the prediction quality of the model. The two of evaluating indexes are shown as follows:

$$\text{RMSE}_{(y',y)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y'_i - y_i)^2} \quad (5)$$

$$R\text{-Square}_{(y',y)} = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

where n is the number of samples, y_i is the real data, y'_i is the predicted data and \bar{y} is the average data.

4.2.2 Experiment setup

We choose eight other models to contrast our model. Besides GRU and bgru, other six model are introduced as follows:

1. *Support Vector Regression (SVR)* (Drucker et al. 1997) SVR is an important branch of support vector machine. It holds that as long as the deviation between the prediction value and true value is not too large, the prediction can be considered as correct without calculating the loss.
2. *Gradient Boosting Regressor (GBR)* (Huang and Oosterlee 2011) GBR is a boosting method through a series of iterations to optimize the regression results, each iteration introduces a weak regressor to overcome the shortcomings of the existing weak regressor combination.
3. *Decision Tree Regressor (DTR)* (Xu et al. 2005) DTR is an application of decision tree. DTR realizes reasonable regression prediction through continuous branching and pruning.
4. *Recurrent neural network (RNN)* (Schuster and Paliwal 1997): RNN is a kind of neural network with sequence data as input, recursion in the evolution direction of sequence, and all nodes (cycle units) are connected by chain. It usually deal with sequence data.
5. *LSTM* (Gers et al. 1999) LSTM is to deal with the problems in long-term dependence. It solves the problem of gradient disappearance to a certain extent.
6. *CBGRU* (Tao et al. 2019) CBGRU tries to predict the air quality data with fewer parameters. It uses a convolutional layer to achieve more effective information.

Some information and comparisons about each method are shown in Table 1

4.2.3 Results analysis

For deep learning model, all are trained for 200 epochs. Each result is averaged over 10 trials. We use dropout between layers with probability of 0.2 in order to avoid the over-fitting problem. In addition, all the deep learning model use the early stop condition in the training process. If the loss of validation data does not change in 20 epochs in training, the training will stop. After obtaining the trained models, each data points in testing dataset are tested and MAE, RMSE and R-Square are calculated.

Before we start the experiment, we need to preset the super parameters in the model for the best performance. Owing to GRU as our baseline, we will test the hyperparameters on GRU, including lookback and the number of neurons in hidden layers which mean how many timesteps should the input data go back and how many neuron nodes are needed to achieve an optimal prediction effect. When testing lookback, lookback was set to a value chosen from a candidate set of {5, 10, 15, 20, 25} and the experiment result is shown in Fig. 5.

From Fig. 5, we can find that with the increase in lookback, the forecasting performance first improves greatly and then begins to deteriorate. When lookback is 10, the model can

Table 1 The information and comparisons about each method

Methods	Type	Preprocessing techniques	Fixed parameter
SVR	Shallow model	Correlation coefficient selection	kernel = 'rbf', C = 1
DTR	Shallow model	Correlation coefficient selection	Depth = 10
GBR	Shallow model	Correlation coefficient selection	n_estimators = 100
RNN	Deep model	None	lookback = 15, epoch = 200
LSTM	Deep model	None	lookback = 15, epoch = 200
GRU	Deep model	None	lookback = 15, epoch = 200
BGRU	Deep model	None	lookback = 15, epoch = 200
CBGRU	Deep model	None	lookback = 15, epoch = 200

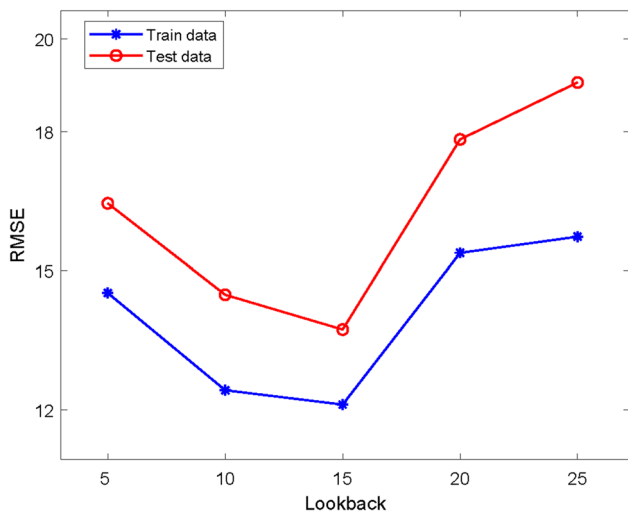


Fig. 5 The lookback value with RMSE

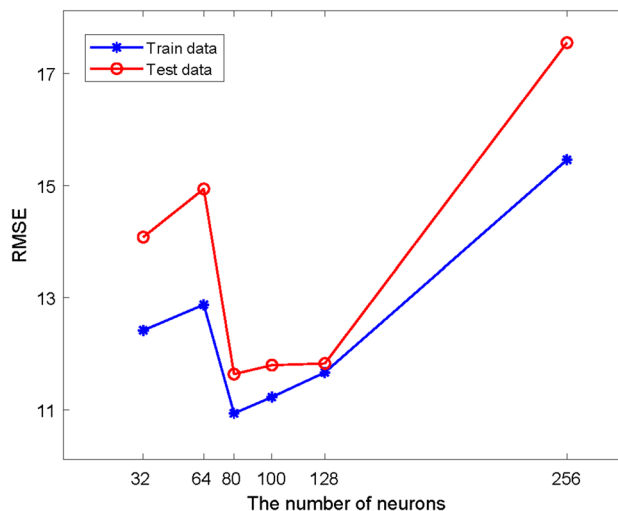


Fig. 6 The number of neurons with RMSE

achieve the best performance. In fact, small lookback cannot guarantee long-term memory input, and large lookback will have more redundant information input, which is not conducive to modeling. Next, we begin to search the suitable number of neurons. The number of neurons will be chosen from the candidate of {32, 64, 80, 100, 128, 256}. The results are shown in Fig. 6.

From Fig. 6, we can find with the increase in the number of neurons, the effect of the model becomes better first and worse later. When the number of neurons is 80, the effect of the model is the best. Combining Figs. 5 and 6, we choose the past 10 hours to predict the next hour’s data and the neurons of GRU hidden layer is 80.

As to convolutional neural networks, a common setting is 2 layers of convolutional layers with the activation functions of RELU where the length of kernel window is 1×3. Due to the lack of max-pooling layer, another convolution is used in which the convolution window is 1×2 for replacing max-pooling layer with the pool size of 2. At last, we add another fully connected layer with only 1 neuron.

4.3 Forecasting results and analysis

To verify the efficiency and accuracy of CDBGU, we develop several comparative models and our trained model. In SVR, we choose RBF kernel and the kernel coefficient is the reciprocal of feature number. In DTR, the maximum depth of the tree is 10 and the criterion is gini. In GBR, we set the loss as the least squares. For deep learning models, i.e., RNN, LSTM, GRU, BGRU, CBGRU, the number of hidden layers was all set as 2 in which the number of layer node is 80.

4.3.1 Results analysis

To verify the effectiveness of our methods, some experiments are performed. Table 2 shows the quantitative results by RMSE and R-Square. From Table 2, we can find that shallow models usually have a worse performance than deep learning models. Compared with deep learning models (such as RNN or LSTM), shallow models (such as SVR) have a larger RMSE, while R-Square are smaller. For deep learning methods, LSTM and GRU have similar performance and both

Table 2 The result in different methods

Methods	RMSE	R-Square
SVR	21.53	0.64
DTR	19.84	0.68
GBR	18.93	0.70
RNN	18.78	0.69
LSTM	11.85	0.73
GRU	11.64	0.73
BGRU	8.43	0.76
CBGRU	7.88	0.80
CDBGRU	7.32	0.84

better than RNN. When adding bi-direction training model, the performance gets better for bi-direction training model can enhance network stability. Compared with CBGRU, our model has a better performance due to the additional Convnets and full-connected layer instead of Max-pooling layer. The result shows that our model can learn more local trend information, time series information and long-term dependencies. Besides, in order to explain the validity of this model more intuitively, Fig. 7 showed the correctness of this method in some data. From Fig. 7, we can observe our method has a good performance on PM2.5 forecasting.

4.3.2 Parameters sensitivity analyze

Although the two parameters (lookback and the number of neurons) were tested before, there are still other parameters such as epoch and the size of conv kernel. To better evaluate the performance of the parameters, we chose to evaluate the results on both the training set and the test set. Figure shows the sensitivity of these two parameters

From Fig. 8, we can find with the increase in epochs, and the test data are getting better and better. However, when the

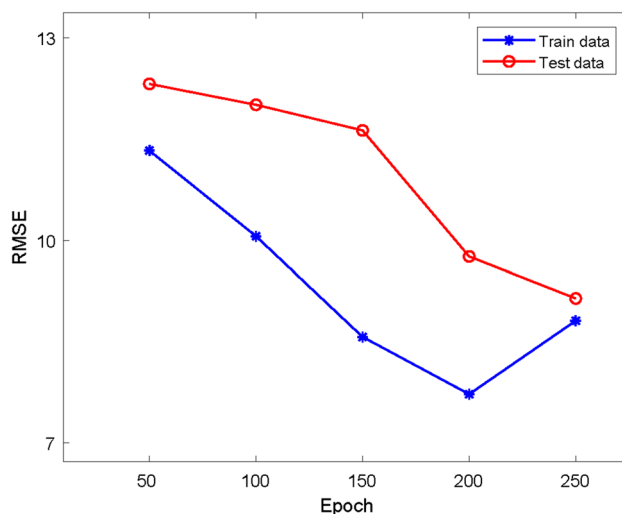


Fig. 8 The sensitivity of epochs

epoch is greater than 200, the effect on the training set will decrease, which is caused by the overfitting problem caused by the increase in training number. So, we chose 200 as the epoch. The same as above, the size of kernel performs great in each size on training data which is performed in Fig. 9, as well as 3 on the testing data. Too large convolution kernel will affect the loss of feature, too small convolution kernel will lead to feature redundancy. So we choose 1 × 3 kernel size as the best size.

4.3.3 Discussion about experiments

In this section, we will analyze the difference of each method and the advantages and disadvantages. From Table 2, we can easily find that our method achieved the best result. Compared the other deep learning method, especially CBGRU, our method has extra full connection layer architecture to ensure the prediction accuracy. However, from the experi-

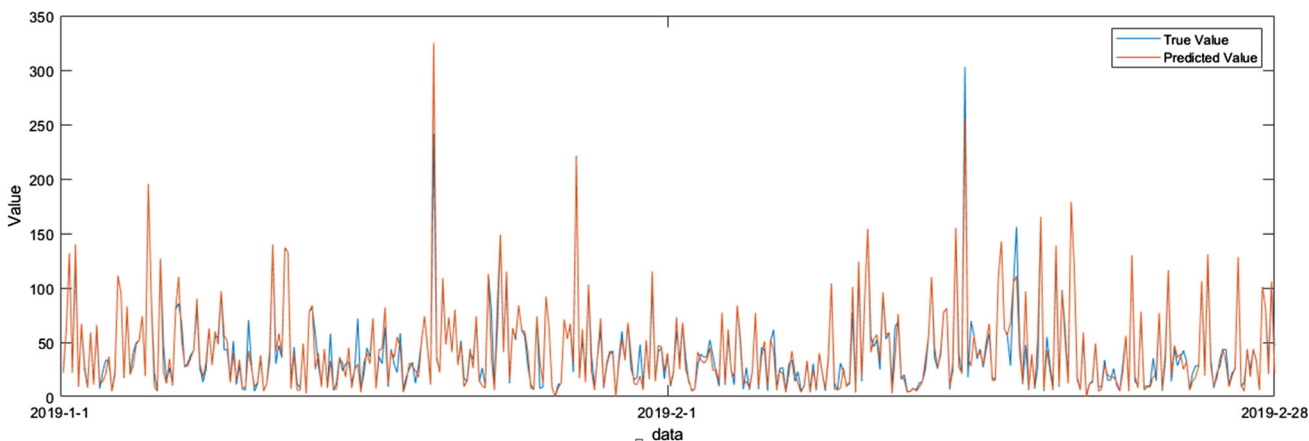


Fig. 7 The true value and the predicted value from 2019-1-1 to 2019-2-28

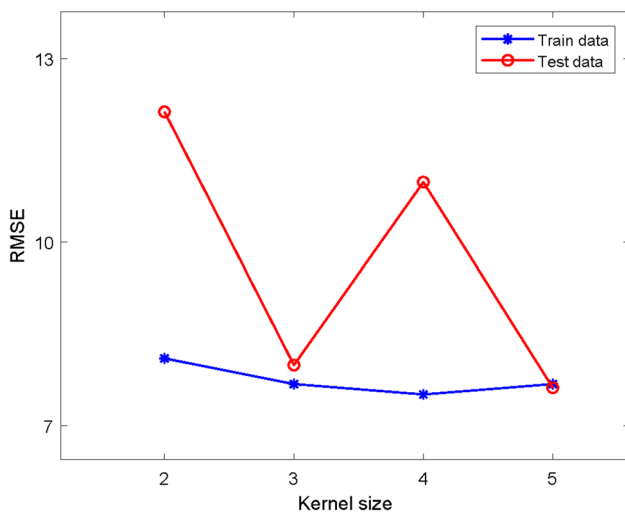


Fig. 9 The sensitivity of kernel size

ments of parameter sensitivity we can also find CDBGRU has no obvious parameter sensitivity, which will disturb us to find the best parameters. Moreover, additional Dense layer will increase the number of parameters, which will increase time complexity. At last, our model has a robust prediction, from Fig. 7 we can observe that there are two outliers, which has been forecasted correctly. Overall, CDBGRU has satisfied the current demand of PM_{2.5} prediction.

5 Conclusion

This paper proposed Convnet and Dense-based Bidirectional Gated Recurrent Unit (CDBGRU), to which a special type of RNN was instructed. Different from the classic RNN model, the feature in air quality data is extracted by Convnet and Dense which has a good performance on a large scale of data. Moreover, we choose the Bi-GRU network with better performance to deal with time series data, e.g., air quality data. Then, we compared it with traditional machine learning method and deep learning method. Through evaluation experiments, we verified the performance superiority and parameter sensitivity. Considering the complex network, our method cannot be called an excellent algorithm for the too many parameters. For the future work, we may consider a better network to predict more accuracy PM_{2.5} value.

Acknowledgements This work is supported by the National Natural Science Foundation of China under Grant 61972207, U1836208, U1836110, 61672290; the Major Program of the National Social Science Fund of China under Grant No. 17ZDA092, by the National Key R&D Program of China under grant 2018YFB1003205; by the Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET) fund, China; by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund.

Funding This work is supported by the National Natural Science Foundation of China under Grant 61972207, U1836208, U1836110, 61672290; the Major Program of the National Social Science Fund of China under Grant No. 17ZDA092, by the National Key R&D Program of China under grant 2018YFB1003205; by the Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET) fund, China; by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund.

Data availability All data can be downloaded from <https://quotssoft.net/air/>

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Code availability All Codes are available from <https://github.com/mc-boo/CDBGRU>

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent Written informed consent for publication was obtained from all participants.

References

- Baowei Wang, Weiwen Kong, Hui Guan, Xiong Neal N (2019a) Air quality forecasting based on gated recurrent long short term memory model in internet of things. *IEEE Access* 7:69524–69534. <https://doi.org/10.1109/access.2019.2917277>
- Baowei Wang, Weiwen Kong, Wei Li, Xiong Neal N (2019b) A dual-chaining watermark scheme for data integrity protection in internet of things. *Comput Mater Continua* 58(3):679–695. <https://doi.org/10.32604/cmc.2019.06106>
- Baowei Wang, Wei Li, Xiong Neal N (2019c) Time-based access control for multi-attribute data in internet of things. *Mobile Netw Appl.* <https://doi.org/10.1007/s11036-019-01327-2>
- Feng Lu, Dongqun Xu, Yibin Cheng, Shaoxia Dong, Chao Guo, Xue Jiang, Xiaoying Zheng (2015) Systematic review and meta-analysis of the adverse health effects of ambient pm_{2.5} and pm₁₀ pollution in the chinese population. *Environ Res* 136:196–204. <https://doi.org/10.1016/j.envres.2014.06.029>
- Gers Felix A, Jürgen Schmidhuber, Fred Cummins (1999) Learning to forget: continual prediction with lstm. *Neural Comput* 12(10):2451–2471. <https://doi.org/10.1049/cp:19991218>
- Giovanni Lonati, Michele Giugliano, Paola Butelli, Laura Romele, Ruggero Tardivo (2005) Major chemical components of pm_{2.5} in milan (italy). *Atmos Environ* 39(10):1925–1934
- Harris Drucker, Burges Christopher JC, Linda Kaufman, Smola Alex J, Vladimir Vapnik (1997) Support vector regression machines. *Adv Neural Inf Process Syst.* https://doi.org/10.1007/978-0-387-77242-4_9
- Jianzhou Wang, Bai Lu, Siqi Wang, Chen Wang (2019d) Research and application of a hybrid forecasting model based on secondary denoising and multi-objective optimization for an air pollution early warning system. *J Clean Prod* 234:54–70. <https://doi.org/10.1016/j.jclepro.2019.06.201>
- Li Anna, Xu Xiao (2018) A new pm_{2.5} air pollution forecasting model based on data mining and bp neural network model. In: 2018 3rd

- International Conference on Communications, Information Management and Network Security (CIMNS 2018)
- Lili Pan, Jiaohua Qin, Hao Chen, Xuyu Xiang, Cong Li, Ran Chen (2019) Image augmentation-based food recognition with convolutional neural networks. *Comput Mater Continua* 59(1):297–313. <https://doi.org/10.32604/cmc.2019.04097>
- Mike Schuster, Paliwal Kuldip K (1997) Bidirectional recurrent neural networks. *IEEE Trans Process* 45(11):2673–2681. <https://doi.org/10.1109/78.650093>
- Min Long, Yan Zeng (2019) Detecting iris liveness with batch normalized convolutional neural network. *Comput Mater Continua* 58(2):493–504. <https://doi.org/10.32604/cmc.2019.04378>
- Min Xu, Pakorn Watanachaturaporn, Varshney Pramod K, Arora Manoj K (2005) Decision tree regression for soft classification of remote sensing data. *Remote Sens Environ* 97(3):322–336. <https://doi.org/10.1016/j.rse.2005.05.008>
- Najafabadi Maryam M, Flavio Villanustre, Khoshgoftaar Taghi M, Naeem Seliya, Randall Wald, Edin Muharemagic (2015) Deep learning applications and challenges in big data analytics. *J Big Data* 2(1):1. <https://doi.org/10.26599/bdma.2019.9020007>
- Qian-Rao FU (2016) School Of Science, and Northwestern Polytechnical University. Research on haze prediction based on multivariate linear regression. *Comput Sci* 43:526–528
- Qing Tao, Fang Liu, Yong Li, Denis Sidorov (2019) Air pollution forecasting using a deep learning model based on 1d convnets and bidirectional gru. *IEEE Access* 7:76690–76698. <https://doi.org/10.1109/access.2019.2921578>
- Ritika Singh, Shashi Srivastava (2016) Stock prediction using deep learning. *Multimed Tools Appl*. <https://doi.org/10.31979/etd.bzmm-36m7>
- Samaher Al-Janabi (2020) Smart system to create an optimal higher education environment using ida and iots. *Int J Comput Appl* 42(3):244–259. <https://doi.org/10.1080/1206212x.2018.1512460>
- Samaher Al-Janabi, Alkaim Ayad F (2020) A nifty collaborative analysis to predicting a novel tool (drfls) for missing values estimation. *Soft Comput* 24(1):555–569. <https://doi.org/10.1007/s00500-019-03972-x>
- Samaher Al-Janabi, Muhammed Abaid Mahdi (2019) Evaluation prediction techniques to achievement an optimal biomedical analysis. *Int J Grid Util Comput* 10(5):512–527. <https://doi.org/10.1504/ijguc.2019.102021>
- Samaher Al-Janabi, Alkaim Ayad F, Zuhail Adel (2020a) An innovative synthesis of deep learning techniques (dcapsnet & dcom) for generation electrical renewable energy from wind energy. *Soft Comput* 24(14):10943–10962. <https://doi.org/10.1007/s00500-020-04905-9>
- Samaher Al-Janabi, Mustafa Mohammad, Ali Al-Sultan (2020b) A new method for prediction of air pollution based on intelligent computation. *Soft Comput* 24(1):661–680. <https://doi.org/10.1007/s00500-019-04495-1>
- Seun Deleawe, Jim Kusznir, Brian Lamb, Cook Diane J (2010) Predicting air quality in smart environments. *J Ambient Intell Smart Environ* 2(2):145–154. <https://doi.org/10.3233/ais-2010-00>
- Wei Hao, Lingyun Xiang, Yan Li, Peng Yang, Xiaobo Shen (2018) Reversible natural language watermarking using synonym substitution and arithmetic coding. *Comput Mater Continua*. <https://doi.org/10.3970/cmc.2018.03510>
- Xinzheng Huang, Cornelis Oosterlee (2011) Generalized beta regression models for random loss-given-default. *J Credit Risk* 7(4):45–70
- Yan Leiming, Wu Yaowen, Yan Luqi, Zhou Min (2018) Encoder-decoder model for forecast of pm2.5 concentration per hour. In: 2018 1st International Cognitive Cities Conference (IC3). pages 45–50. IEEE. <https://doi.org/10.1109/ic3.2018.00020>
- Yuhong Zhang, Qinqin Wang, Yuling Li, Xindong Wu (2018) Sentiment classification based on piecewise pooling convolutional neural network. *Comput Mater Continua* 56:285–297. <https://doi.org/10.30534/ijeter/2020/63892020>
- Zaremba Wojciech, Sutskever Ilya, Vinyals Oriol (2014) Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*
- Zheng Yu, Liu Furui, Hsieh Hsun-Ping (2013) U-air: When urban air quality inference meets big data. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. pages 1436–1444. ACM. <https://doi.org/10.1145/2487575.2488188>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.