

Prediction of PM2.5 Concentration Based on CNN-BiGRU Model

Xinfang Li, Hua Huo

Henan University of Science and Technology, School of Information Engineering, China

Abstract: The issue of air pollution has always been a concern. Bad smog weather not only brings inconvenience to people's travel, but also poses a threat to people's health. PM2.5 concentration is an important indicator of air conditions. Therefore, it is of long-term significance to analyze and predict the concentration of PM2.5. Aiming at the problem that a single machine learning model cannot consider the impact of multiple factors on PM2.5 concentration changes, and the data characteristics are complex, which cannot better capture all the characteristics of the data, and cannot highlight the regularity of PM2.5 changes over time, the construction of a combined model further improves the prediction accuracy. Firstly, based on the PM2.5 concentration values, air quality data, and meteorological data at various stations in New Taipei City, Taiwan Province, through analyzing the spatiotemporal distribution characteristics of the PM2.5 concentration at the target station, as well as the correlation with various pollutant factors and meteorological factors, Spearman correlation analysis is used for feature selection. The combined model CNN-BiGRU constructed in this paper utilizes its unique convolution operation to extract features from one-dimensional data, and combines the circular neural network BiGRU with bidirectional transmission function to model and predict PM2.5 concentration based on the functional advantages of both parties.

Keywords: PM2.5, CNN, GRU, Combination Model.

1. Introduction

With the acceleration of China's industrialization process, the accompanying air pollution problem is becoming increasingly serious. Poor air quality can affect our daily lives and even our physical and mental health, with PM2.5 causing the most serious harm to the human body[1,2]. Accurate and timely prediction of future PM2.5 concentrations can provide assistance for people's defense.

In recent decades, many scholars have analyzed and studied the formation, diffusion, and concentration prediction of PM2.5. In previous years, most of the research on PM2.5 used traditional prediction methods, mainly based on mechanism models, which dynamically modeled the transmission and formation of PM2.5 based on the geographical environment, meteorological conditions, and even industrial level of a region. And most of them use CAMQ models[3-6]. For example, Lee et al[7] used the MM5-SMOKE-CMAQ modeling system to respectively generate meteorological fields, prepare emissions, and simulate air quality; Jiang et al[8] used a community multiscale air quality (CMAQ) simulation model to predict the daily average concentration of fine particulate matter (PM2.5). Another prediction method is to establish regression models. For example, Wang Weirud et al[9] used the Box-Jenkins theory to establish an ARIMA model to predict the PM2.5 concentration in Hangzhou, and found the optimal parameter value with good prediction accuracy; Wang Juan[10] used multiple regression analysis and gray correlation 2 degree method to conduct air pollution quality research in most areas of the city; Wang Jing et al[11] proposed a hierarchical autoregressive model based on Bayesian to handle the synchronous prediction of PM2.5; Most of these traditional PM2.5 concentration modeling methods require complex calculations and have a problem of low accuracy.

In recent years, machine learning algorithms have become a research craze. AmA et al[12] established two different

models, support vector machine SVM and artificial neural network ANN. The results show that the prediction accuracy of neural network for PM2.5 is better than that of support vector machine; Chen et al[13] established a random forest model and two traditional regression models to estimate ground PM2.5 concentrations. The results show that the prediction accuracy of the random forest model is significantly higher than that of the two traditional regression models; Joharestani et al[14] studied the importance of the characteristics of PM2.5 prediction in urban Tehran, and constructed random forests, limit gradient lifting, and deep learning machine learning methods to predict it. Liang Xiguan[15], Kong Yu, et al[16] established a tree model to predict PM2.5, and obtained its strong stability compared to other model tree models. These machine learning models have achieved good results in the prediction of PM2.5 concentration. However, for the single model is not accurate in the prediction of PM2.5 concentration, and cannot better capture the characteristics of multidimensional data changes, the combined model can further improve the prediction accuracy. For example, Ding et al[17] combined convolutional neural networks (CNN) with short-term and short-term memory (LSTM) to achieve better prediction results than other machine learning methods; Qian et al[18] used a generalized additive model to combine neural networks, random forests, and gradient enhanced PM2.5 estimates; Niu et al[19] used empirical mode decomposition method, gray wolf optimization algorithm, and SVR model to achieve high-precision prediction of PM2.5. These experiments all use a single model for comparative verification. Therefore, it is proved that the combined model can improve the prediction accuracy to a certain extent. In this paper, a CNN-BiGRU model was established to analyze and predict PM2.5. Its unique convolution operation was used to extract features from one-dimensional data. At the same time, the circular neural network BiGRU with bidirectional transmission function was combined to model and predict

PM2.5 concentration based on the functional advantages of both sides.

2. Proposed Algorithm

2.1. CNN

Convolutional Neural Networks is one of the most important networks in the field of deep learning. It is a deep learning model or multilayer perceptron similar to artificial neural networks, and is often used in analyzing visual images (such as face recognition, image segmentation, and image classification) and natural language processing. CNN is a feedforward neural network that can extract features from data, with local connectivity, weight sharing, and other characteristics. A convolutional neural network is mainly composed of five layers: data input layer, convolutional layer,

pooling layer, fully connected layer, and output layer[20].

2.2. BiGRU

BiGRU (Bidirectional Gate Recurrent Unit) is an extended model of GRU, which consists of two GRU models: a forward transitive GRU model that receives forward inputs; The other is the reverse transmission GRU model, which learns the reverse input of the model[21,22]. As the PM2.5 concentration value studied in this article belongs to time series data, which is greatly affected by the chronological order, the BiGRU model is selected to model it, which can not only learn the impact of historical information on subsequent information, but also ensure the correlation between historical information and subsequent information, thereby improving the prediction accuracy of PM2.5 concentration. The structure diagram of BiGRU module 3 is shown in the figure 1.

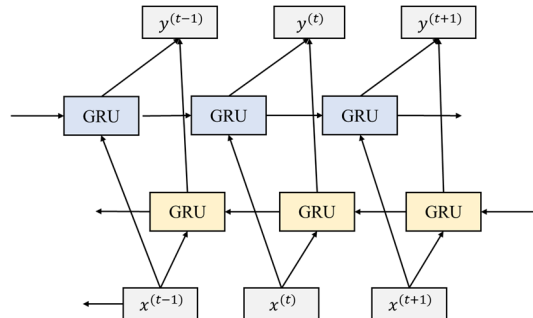


Figure 1. BiGRU model structure

3. Experiment

3.1. Feature selection

PM2.5 concentration is influenced by various factors and exhibits a state of coexistence of multiple change modes[23,24]. Therefore, in order to build an effective model and improve prediction accuracy, it is necessary to establish a reasonable input characteristic matrix. A good representation of the relationship between input data and PM2.5 is a key part of building a prediction model.

Feature selection can effectively improve the accuracy of the model, reduce runtime, and reduce the risk of overfitting by reducing redundant and unrelated features. From the correlation analysis in Chapter 3, it can be seen that most pollution factors and meteorological factors are significantly related to the increase and decrease of PM2.5 concentration.

In this paper, Spearman correlation coefficient is selected for feature selection[25]. Take the monitoring point in Xindian District as an example. Figure 2 and Figure 3 show the correlation between PM2.5 and pollutants and the correlation between meteorological factors. You can clearly see the Spearman correlation between each feature and PM2.5. The darker the color, the stronger the correlation. The correlation between PM10 and PM2.5 in pollutants reaches 0.83, with a strong correlation, followed by CO, while the correlation between meteorological factors and PM2.5 is weak, and most characteristic variables are inversely proportional to them, such as RH (relative humidity) and AMB_TEMP (temperature), etc. Therefore, PM10, CO, NO2, and SO2 with strong correlation are selected as pollutant factors, while RH and WS_HR are selected as meteorological factors, which are used as a correlation factor auxiliary variable.

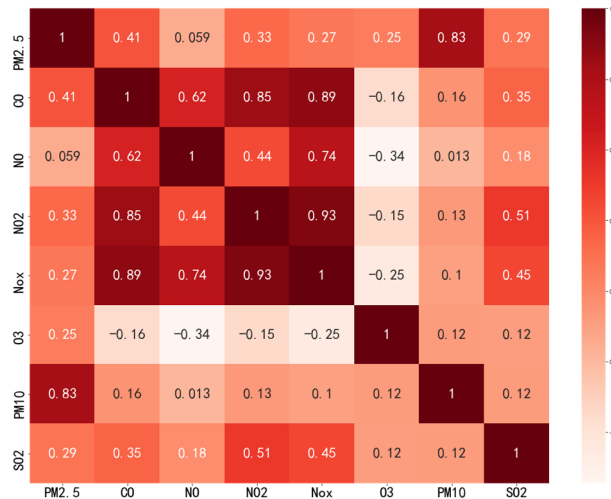


Figure 2. The correlation between PM2.5 and other pollutants

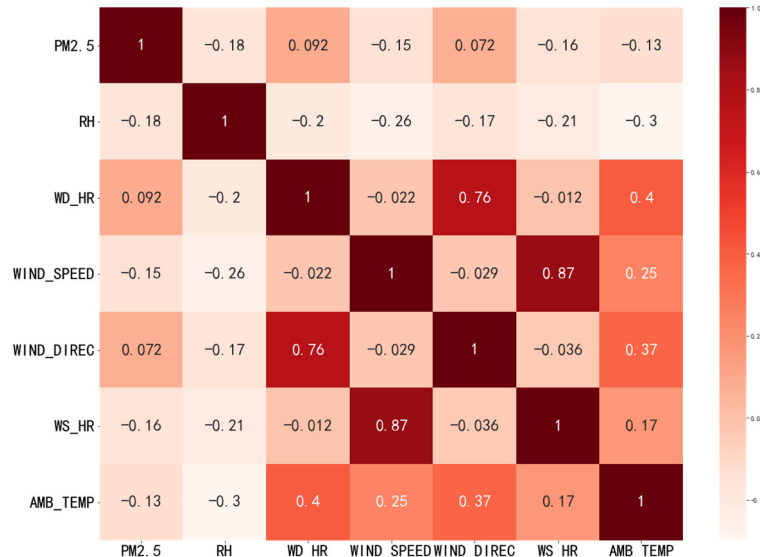


Figure 3. The correlation between PM2.5 and meteorological factors

3.2. Data preprocessing

Missing data will lead to the reduction of sample information, deviation of data analysis results, and “pseudo regression” of data outliers. To solve these problems, this paper adopts the following measures:

Data missing is filled with the nearest Neighbor (KNN) algorithm. The filling idea of KNN considers the “distance” between two samples, and selects the average value or distance weighting of the nearest several observations as the filling value of the missing samples.

Outlier refers to the index data containing incorrect input and illegal data, that is, the parameter value exceeds the normal range. The box diagram method is used to treat values less than $QL-1.5IQR$ or greater than $QU+1.5IQR$ as exceptions. QU and QL are the upper quartile and the lower quartile respectively, and IQR is the spacing of the quartiles. The detected outliers are regarded as missing values and filled with KNN filling method.

The data range of different features is different, and the difference between values may be large, so the data are standardized. Standard deviation is used to standardize data. The mean value and standard deviation of data processed by this method are 0 and 1 respectively. The transformation formula is shown in Equation (20):

$$X^* = \frac{X - \bar{X}}{\delta} \quad (1)$$

\bar{X} is the mean value of the original data, and δ is the standard deviation of the original data.

3.3. Model construction

The hybrid CNN-BiGRU model constructed in this paper first uses one-dimensional CNN to mine the nonlinear characteristics of data, improve the running speed, and also eliminate some unstable factors. Then, its output is used as the input of the next time series model, BiGRU, to extract

hidden temporal rules, which can capture long-term and short-term dependencies in time series, And feature vectors can be extracted through forward and backward transmission. This model combines the advantages of CNN in learning local and related features with the advantages of learning temporal rule representations from BiGRU, further improving the prediction accuracy of PM2.5 concentrations. In this study, there are eight data features used. The input is the pollutant concentration data for the past hour, and the output is the concentration value of PM2.5 for the next hour. That is, the input is a matrix of (1×8) size, and the output matrix is (1×1) .

The construction of CNN-BiGRU model is mainly divided into two parts, the structural design of one-dimensional CNN and the structural design of BiGRU model. The input data input sample constructed in this paper is divided into data segments of a certain length. Using one-dimensional CNN can effectively extract local features. Since this article uses historical 1-hour data of various pollutants and meteorological factors to predict the PM2.5 concentration in the next hour, the design of the network layer should not be complex. Based on previous experience, a layer of convolution layer and pooling layer has been designed to avoid increasing the calculation amount.

The structural design of the CNN-BiGRU model is also based on past experience. It mainly consists of the following parts: input layer, convolution layer, pooling layer, BiGRU layer, full connection layer, and output layer. The input layer is mainly used for data preparation. The convolution layer extracts features, and the pooling layer is used for data sampling. Generally, the convolution layer and pooling layer of a convolutional neural network are nested and used to extract and compress temporal dimension features. The convolution layer transmits the extracted data information to the bi-directional circular neural network BiGRU layer, which performs forward and reverse transmission of data, followed by a fully connected layer. The model structure diagram is shown in Figure 4.

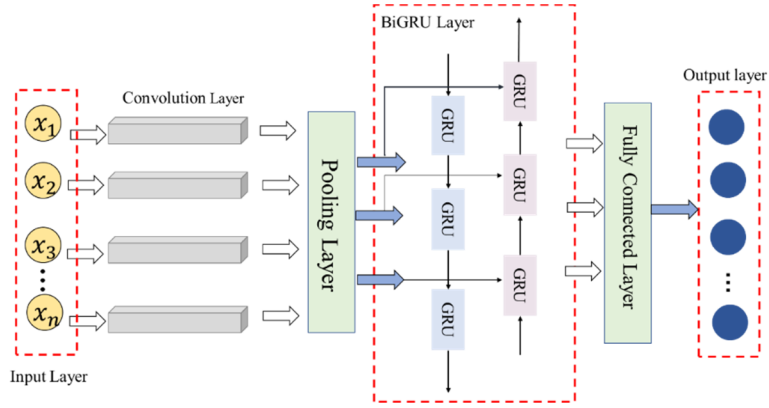


Figure 4. Structure of the CNN-BiGRU model

Due to the small dimension of the input matrix in this article, the number of layers for constructing the network should not be too large. This article sets up a one-dimensional convolutional layer, with a convolutional core size of 2, a convolutional core number of 64, and a step size of 1. The maximum pooling is selected for the pooling layer. The BiGRU layer is also set to one layer with a number of neurons of 200, which can save time while maintaining good accuracy.

3.4. Evaluation criteria

In this paper, root mean square error (RMSE), mean absolute error (MAE) and determination coefficient R^2 are selected to evaluate the prediction performance of the model. Its formula is defined as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^N |y_i - x_i|}{N} \quad (3)$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y - y_i)^2} \quad (4)$$

The smaller the RMSE and MAE values are, the better the effect of the model will be. The larger the R^2 value is, the better the effect of the model will be.

3.5. Result analysis

From the scatter diagram Figure 5 (a) (b) (c), it can be observed that the R^2 values of the three monitoring points are 0.9111, 0.8890, and 0.8768, respectively, which are greater than 0.85, which can prove that the fitting effect of the combined model at the three selected stations is very good. The univariate linear fitting function for the monitoring points in Xindian District is: $y=0.9318x+1.3457$, with a slope of 0.9318, which is closer to 1. Overall, the actual value and predicted value are relatively close. The slopes of the fitting equations for Taoyuan District and Songshan District are 0.87 and 0.88, and their data fitting is also relatively concentrated. Figure 6 (a) (b) (c) show a broken line diagram of the actual and predicted values of the three stations, and selects the PM2.5 concentration values for the first 200 hours. It can be seen that when the PM2.5 concentration value is less than $5 \mu\text{g}/\text{m}^3$, the fitting degree of the predicted value of the model is low, and when the PM2.5 concentration value is between $10\text{-}25 \mu\text{g}/\text{m}^3$, the fitting effect is good. Therefore, it can be concluded that the CNN-BiGRU combined model has a relatively good effect in predicting PM2.5 concentration.

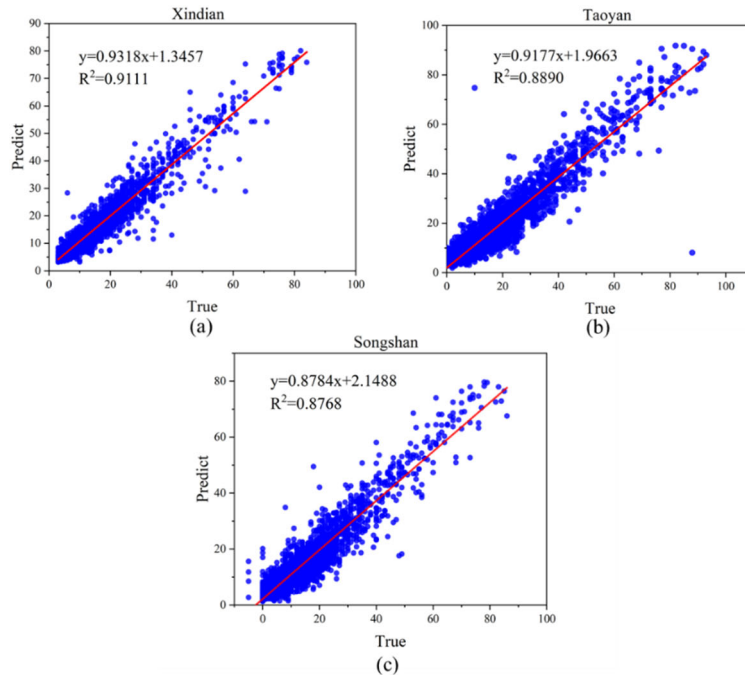


Figure 5. Forecast results for each site

In Figure 6(a)(b) and (c) are the loss iteration diagrams for the LSTM model, CNN-GRU model, and CNN-BiGRU model, respectively. The evaluation index selects the MAE graph, with the ordinate Loss representing the loss value, and the abscissa Epoch representing the number of iterations. The Epoch for all models is set to 60. As can be observed in the figure, the loss rates of the three models all decreased rapidly during the first few Epochs, and then decreased slightly. When the number of iterations is 5, it can be observed that the loss value of the LSTM model is about 2.9, the loss value of the CNN-GRU model is about 3, and the training set loss

value and test set loss value of the CNN-BiGRU model are both less than 3. It can be concluded that the model used in this article, CNN-BiGRU, can achieve the minimum loss value faster. Continue to observe that the LSTM's Loss value in Figure 6 (a) reaches the minimum value when Epoch is approximately 45 times, while the CNN-GRU model in Figure 6 (b) reaches the minimum value when Epoch is 50 times, and the number of iterations for the CNN-BiGRU model reaches the minimum value when the loss reaches the minimum value is approximately 30.

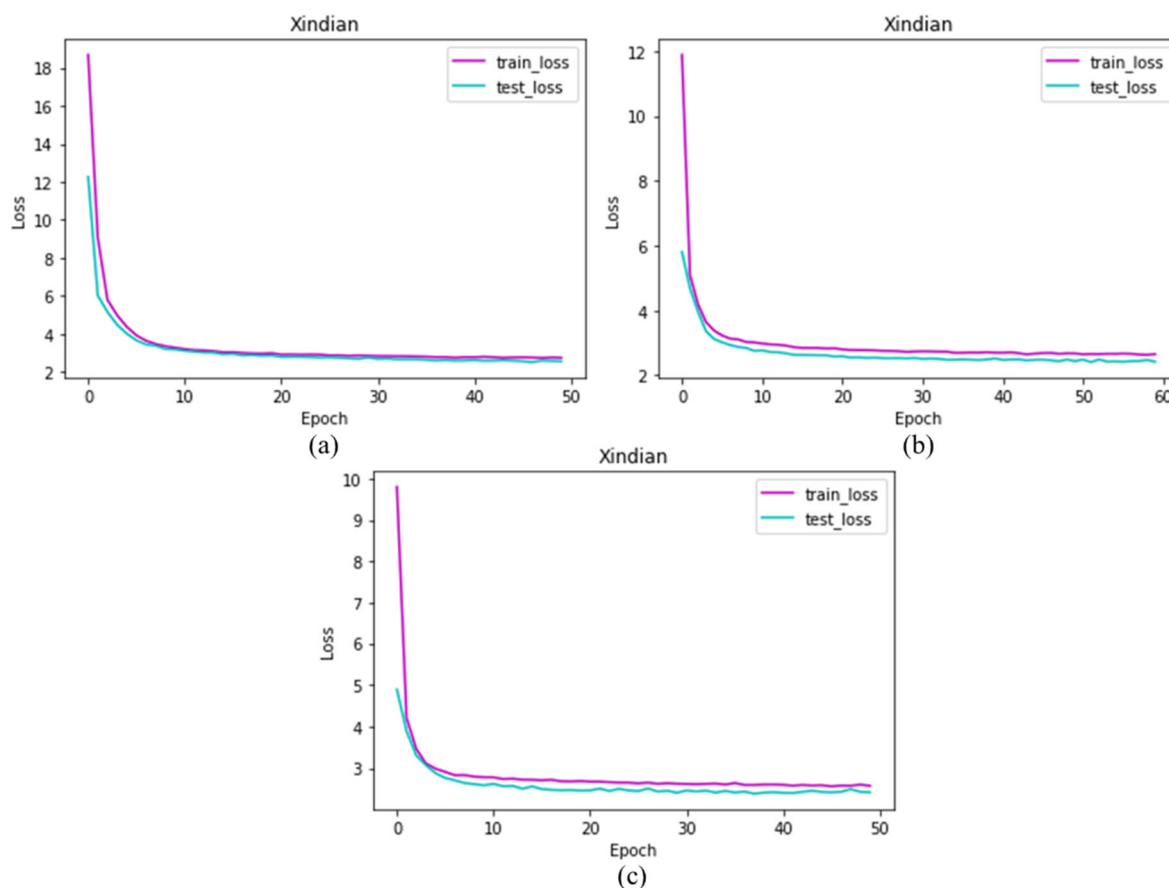


Figure 6. Loss iteration diagram of each model

From the histograms 7 and 8, it can be seen that the CNN-BiGRU model has the lowest MAE and RMSE values at the three target monitoring points, and the MAE values of the four models are LSTM>GRU>CNN-GRU>CNNBiGRU from high to low. Among them, the MAE predicted by each model at the monitoring points in Xindian District were 2.5932, 2.6157, 2.4596, and 8.24341. The MAE value of the CNN-BiGRU model in Xindian District increased 6.5% compared to the MAE value of LSTM, increased by 7.5% compared to GRU, and increased by 1% compared to CNN-GRU. The monitoring points in Taoyuan District increased by 1.6%, 1.3%, and 1.6% respectively. The monitoring points in Songshan District increased by 2.5%, 2.4%, 0.2% Model

CNN-BiGRU has significant advantages over single model LSTM and GRU. Among the different evaluation index values for the three target monitoring points, the R^2 predicted by the combined model for PM2.5 concentration at the three stations reached 0.9111, 0.8890, and 0.8768, respectively. Compared with LSTM's 0.8901, 0.8847, and 0.8593, the prediction accuracy of the combined model for PM2.5 concentration increased by 2.3%, 0.49%, and 2%, respectively. Compared with GRU's 0.8949, 0.8848, and 0.8613, the prediction accuracy of the combined model for PM2.5 concentration increased by 1.8%, respectively 0.5% and 1.8%.

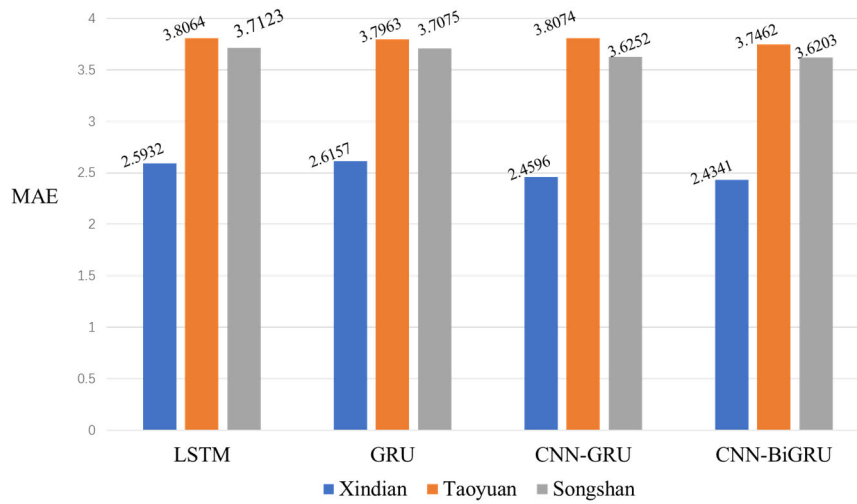


Figure 7. MAE diagram of each model

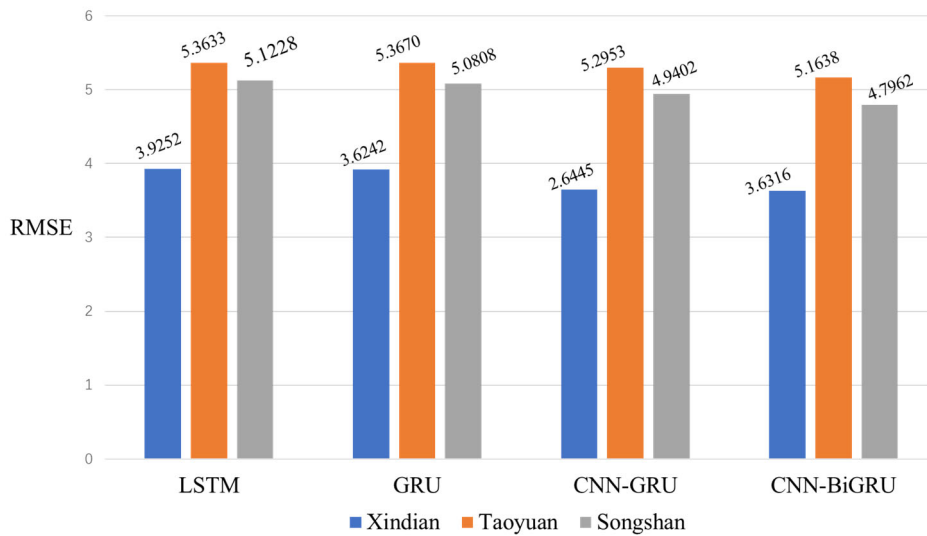


Figure 8. MAE diagram of each model

The R^2 of the CNN-GRU model is 0.9100, 0.8876, and 0.8692, respectively, and the prediction accuracy of the proposed combined model has been improved. Its RMSE can also be calculated as a percentage of its improvement. It can be concluded that this model has a significant improvement compared to LSTM, especially for monitoring points in the

new store area. Compared to the combined module CNN-GRU, the improvement is not significant, but there is also progress, because CNN-GRU also combines the ability of CNN to mine local features of data and the ability of GRU to process long-term and long-term memory. Therefore, it can be concluded that the combination model is better than the single model, and the proposed combination model has good prediction performance in most cases.

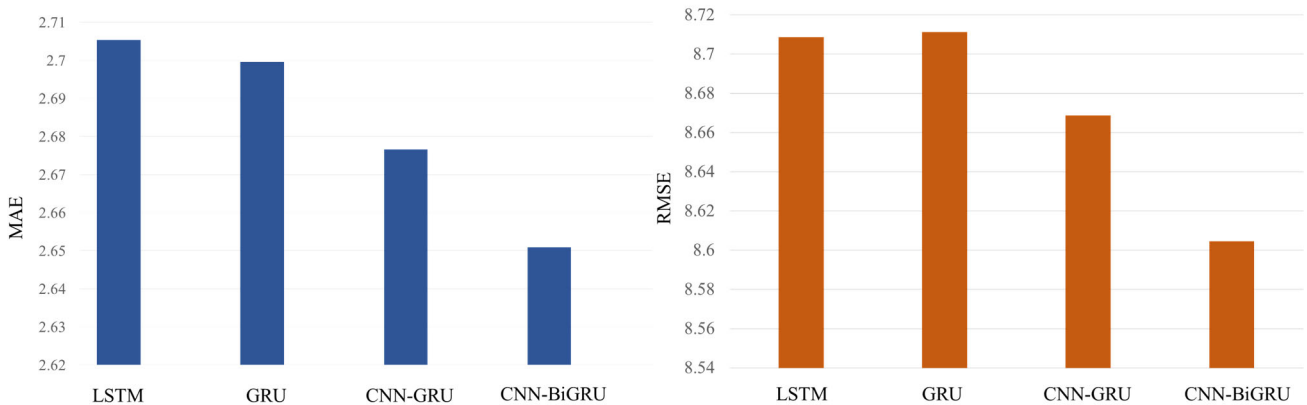


Figure 9. MAE diagram and RMSE diagram of each model

Data from 10 monitoring stations in Xinbei City were selected, and the average values of each characteristic were taken to predict the PM_{2.5} concentration value in Xinbei City. First, construct input features, select relevant feature variables, and then adjust parameters. Input the constructed CNN-BiGRU combination model to predict the PM_{2.5} concentration in Xinbei City, and also select LSTM, GRU, and CNN-GRU models as comparison models. Figure 9 shows the histogram of the evaluation indicators for predicting PM_{2.5} concentration in Xinbei City by all models. The evaluation indicators are MAE and RMSE. It can be observed that compared to using data from a single station to predict, the value of RMSE is higher, which is due to differences in pollutants and meteorological conditions across different regions, showing strong regional characteristics. However, in the MAE and RMSE diagrams of each model, the value of CNN-BiGRU is the lowest, followed by the CNN-GRU model, and the error value of the two single models is still the highest. Therefore, it can be concluded that the effect of this combined model in testing the PM_{2.5} concentration in the entire New Taipei City is still significant.

In summary, the CNN-BiGRU combination model used in this article can fully utilize the advantages of convolutional neural networks and two-way gated cyclic unit networks. CNN with strong feature extraction capabilities will input output features into the BiGRU network, thereby avoiding the singularity of single model prediction, and improving the prediction ability.

4. Conclusion

In this paper, we conducted a predictive analysis of PM_{2.5} concentration data from three monitoring stations in Xinbei City, Taiwan Province. First, we selected characteristic variables with strong correlation with them as predictive auxiliary factors, and then conducted modeling and prediction of PM_{2.5} concentration. We analyzed three evaluation indicators, and designed three models for comparison: LSTM model, GRU model, and CNN-GRU model. It was concluded that the prediction performance of the CNN-BiGRU model at the three target stations was good. Finally, a prediction analysis was conducted on the average PM_{2.5} concentration in Xinbei City, further verifying that the prediction effect of the combined model is better than that of a single model, and that the CNN-BiGRU model has more advantages in terms of measurement performance.

References

- [1] Chen Z, Chen D, Zhao C, et al. Influence of meteorological conditions on PM_{2.5} concentrations across China: A review of methodology and mechanism[J]. *Environment international*, 2020, 139: 105558.
- [2] Rajagopalan S, Al-Kindi S G, Brook R D. Air pollution and cardiovascular disease: JACC state-of-the-art review[J]. *Journal of the American College of Cardiology*, 2018, 72(17): 2054-2070.
- [3] Hogrefe C, Lynn B, Goldberg R, et al. A combined model-observation approach to estimate historic gridded fields of PM_{2.5} mass and species concentrations[J]. *Atmospheric Environment*, 2009, 43(16): 2561-2570.
- [4] Zhang Y, Pun B, Wu S Y, et al. Application and evaluation of two air quality models for particulate matter for a southeastern US episode[J]. *Journal of the Air & Waste Management Association*, 2004, 54(12): 1478-1493.
- [5] Chuang M T, Fu J S, Jang C J, et al. Simulation of long-range transport aerosols from the Asian Continent to Taiwan by a Southward Asian high-pressure system[J]. *Science of the total environment*, 2008, 406(1-2): 168-179.
- [6] San José R, Pérez J L, Morant J L, et al. Improved modelling experiment for elevated PM₁₀ and PM_{2.5} concentrations in Europe with MM5-CMAQ and WRF/CHEM[J]. *Air Pollution*, 2009: 377-86.
- [7] Lee D G. Comparison between Atmospheric Chemistry Model and Observations Utilizing the RAQMS-CMAQ Linkage, Part II: Impact on PM_{2.5} Mass Concentrations Simulated[J]. *Asian Journal of Atmospheric Environment*, 2014, 8(2): 108-114.
- [8] Jiang X, Yoo E. The importance of spatial resolutions of Community Multiscale Air Quality (CMAQ) models on health impact assessment[J]. *Science of the Total Environment*, 2018, 627: 1528-1543.
- [9] Wang W, Wu T, Zhang Z. PM_{2.5} prediction of Hangzhou based on ARIMA model [J]. *Journal of Natural Sciences*, Harbin Normal University, 2018, 34(3): 49-55.
- [10] Wang J. Prediction of PM_{2.5} based on multiple regression analysis [J]. *Microcomputer application*, 2020, 3.
- [11] Wang J, Cao C. PM_{2.5} prediction of Beijing based on Bayesian hierarchical autoregressive spatio-temporal model [J]. *Journal of Nanjing University of Information Science & Technology (Natural Science)*, 2023, 15(01): 34-41.
- [12] Masood A, Ahmad K. A model for particulate matter (PM_{2.5}) prediction for Delhi based on machine learning approaches[J]. *Procedia Computer Science*, 2020, 167: 2101-2110.
- [13] Chen G, Li S, Knibbs L D, et al. A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological and land use information[J]. *Science of the Total Environment*, 2018, 636: 52-60.
- [14] Zamani Joharestani M, Cao C, Ni X, et al. PM_{2.5} prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data[J]. *Atmosphere*, 2019, 10(7): 373.
- [15] Liang Y, Li Z, Jin Y, et al. Comparison of PM_{2.5} prediction effect in Beijing based on tree model [J]. *Environmental engineering*: 1-13 [2023-03-13].
- [16] Kong Y, Wang H, Zhang H, et al. PM_{2.5} concentration prediction based on integrated learning algorithm [J]. *Environmental protection science*, 2021, 47(4): 17-23.
- [17] Ding C, Wang G, Zhang X, et al. A hybrid CNN-LSTM model for predicting PM_{2.5} in Beijing based on spatiotemporal correlation[J]. *Environmental and Ecological Statistics*, 2021, 28(3): 503-522.
- [18] Di Q, Amini H, Shi L, et al. An ensemble-based model of PM_{2.5} concentration across the contiguous United States with high spatiotemporal resolution[J]. *Environment international*, 2019, 130: 104909.
- [19] Niu M, Wang Y, Sun S, et al. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM_{2.5} concentration forecasting[J]. *Atmospheric Environment*, 2016, 134: 168-180.
- [20] Zhou F, Jin L, Dong J. Review of Convolutional Neural Network [J]. *Journal of Computer Science*, 2017, 40(6): 1229-1251.
- [21] Liang R, Chang X, Jia P, et al. Mine gas concentration forecasting model based on an optimized BiGRU network[J]. *ACS omega*, 2020, 5(44): 28579-28586.

- [22] Zhang B, Jia M, Xu J, et al. Network Security Situation Prediction Model Based on EMD and ELPSO Optimized BiGRU Neural Network[J]. Computational Intelligence and Neuroscience, 2022, 2022.
- [23] Zhang H, Wang Y, Hu J, et al. Relationships between meteorological parameters and criteria air pollutants in three megacities in China[J]. Environmental research, 2015, 140: 242-254.
- [24] Rojas N, Galvis B. Relationship between PM2. 5 and PM10 in Bogotá[J]. Revista de Ingeniería, 2005 (22): 54-60.
- [25] Liu D, Cho S Y, Sun D, et al. A Spearman correlation coefficient ranking for matching-score fusion on speaker recognition[C]. in TENCON 2010-2010 IEEE Region 10 Conference. IEEE, 2010: 736-741.