Air Pollutant Concentration Prediction Based on GRU Method

Xinxing Zhou¹, Jianjun Xu¹*, Ping Zeng¹ and Xiankai Meng¹

¹ Laboratory of Software Engineering for Complex Systems, College of Computer, National University of Defense Technology, Changsha, China.
*Corresponding author’s e-mail: jjxu@nudt.edu.cn

Abstract. In recent years, the Chinese government has paid more and more attention to the construction of ecological civilization, and the governance of air pollution is one of the necessary links. Aiming at the problem of the time series prediction of air pollutants, this paper develops a time series prediction model based on deep learning method. In this paper, Beijing's hourly PM2.5 concentration information and weather information are used as input. Through GRU model, four models are trained according to the four seasons of spring, summer, autumn, and winter, and the effects of the four models on predicting the corresponding seasonal PM2.5 are evaluated by using corresponding test sets. After repeated experiments and constant adjustment of model parameters, the prediction error and prediction accuracy of the model are analyzed and compared, then the feasibility and advantages of this method are verified. The results delineate that the prediction accuracy of the model based on the GRU model is high, and the method is valid for the time series prediction of air pollutants.

1. Introduction

With the rapid development of China's economy and the continuous expansion of the urban scale, air pollution is becoming more and more serious. We are concerned that air quality is germane to human health. As the core function of the pollution weather monitoring and early warning system, air pollution forecasting and air quality forecasting have a crucial impact on the overall system's role [1]. Air pollution forecasting and air quality forecasting are complex systems engineering. How to improve the accuracy of prediction is a hot and challenging issue in the field of air pollution prevention and control.

The time series of the API (Air Pollution Index) and the AQI (Air Quality Index) are high-dimensional and complex where it is challenging to predict, analyze, and model. And deep learning methods usually provide better expression and classification. Asha B.Chelani and other scholars established a hybrid autoregressive nonlinear model and used this improved method to predict the NO₂ concentration in Delhi, India. The prediction results are incredibly satisfactory, which proves that the model has strong applicability in air quality prediction [2]. Based on the GA-ANN theory, Zhao Hong tried to combine the artificial neural network operator with the genetic operator and applied it to Tianjin's future air quality prediction. The prediction result is more accurate than the single method prediction, indicating two methods. The combination helps to improve the accuracy of air quality predictions [3]. Bing-Chun Liu and other scholars used the multidimensional collaborative SVR method to predict air quality in the Beijing-Tianjin-Shijiazhuang area [4]. Xie Shenjun used Hefei's air quality and weather data from July 1, 2016, to July 10, 2016, to establish a PM2.5-hour average concentration prediction model through LIBSVM to better describe the nonlinearity between PM2.5
and its impact factors. Relationships, and based on this model, can obtain higher prediction accuracy than BP networks [5].

The problem of air pollutant concentration prediction is a typical time series prediction problem with multiple input variables. With the GRU (Gated Recurrent Unit) recurrent neural network can learn long-term dependence information, and we can availably use it in the field of time series prediction. Meanwhile, Chung J's research has confirmed that GRU performs better in convergence on CPU time, parameter updating, and generalization compared to LSTM (Long Short Memory Network) [6]. Our work is mainly to use GRU to establish a PM2.5 hourly average concentration prediction model for Beijing, including the year-round model and four models corresponding to the four seasons of spring, summer, autumn, and winter, to improve the accuracy of PM2.5 concentration prediction.

2. Models
In this section, we will introduce the overall architecture and GRU details of the model of air pollutant concentration prediction used in our study, as well as the method of data preprocessing.

2.1 Overall Structure
Recurrent neural networks need to repeatedly apply the same operation at various moments of a long time series to construct deep computational graphs, which results in that the defect of long-term dependence is particularly prominent. To overcome the problem that RNN does not handle long-term dependencies well, Hochreiter and Schmidhuber proposed the LSTM [7] model. GRU [8] is a variant of LSTM that not only maintains the characteristics of LSTM but also makes the structure simpler. Figure 1 shows the overall architecture of the model of air pollutant concentration prediction used in this paper. The input layer is comprised of input vectors, and the input vector is entered into a GRU network.

2.2 Gated Recurrent Unit
Figure 1 also shows the expansion of the GRU model, with hidden layers interconnected at multiple times. Each recurrent neural network has a duplicate module, and the duplicate module structure of GRU is slightly simpler than that of LSTM. It only has two gates, namely the update gate and the reset gate, that is, $z_t$ and $r_t$ in Figure 2. The update gate is used to control the extent to which the information of previously hidden states is carry over into the current state. The larger the value of the update gate, the more the information of previous states is brought in. And the reset gate is used to control the extent of ignoring the information of previous states, then the smaller the value of the reset gate, the more it is ignored. Accordingly, the capture of short-term dependencies is usually accompanied by frequent activation of reset gates, while long-term dependencies are accompanied by the activation of update gates.
In our research, the Gated Recurrent Unit (Figure 2) is given by equation (1)-(5), where $\sigma$ is the logistic sigmoid function. $x$ and $h$ are the input and the previous hidden state. $W_r$, $W_z$ and $W_h$ are weight matrices which are learned. (The $[ ]$ denotes that the two vectors are connected, and * represents the multiplication of the matrix elements.)

\[
\begin{align*}
    r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]) \\
    z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]) \\
    \tilde{h}_t &= \tanh(W_h \cdot [r_t * h_{t-1}, x_t]) \\
    h_t &= (1-z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \\
    y_t &= \sigma(W_o \cdot h_t)
\end{align*}
\]

### 2.3 Data Preprocessing Details

In this paper, the datasets are standardized by z-score which is given by equation (6). Where $\mu$ is the average and $\sigma$ is the standard deviation.

\[
z = \frac{x - \mu}{\sigma}
\]

We divided the dataset into a training set and a test set, and the division method is as follows: the dataset is divided into several study periods, each of which consists of 1000 pieces of data, and the first 750 as a training set and the last 250 as a test set.

### 3. Experiments

#### 3.1 Datasets

Beijing is the capital of China and the GaWC (Globalization and World Cities). Hence it has essential research value for the prediction of Beijing’s air quality. The dataset used in this paper consists of two parts: Beijing PM2.5 average hourly concentration data from 2010 to 2014 and Beijing-to-Beijing meteorological data from 2010 to 2014. The data are from the official website of the China Meteorological Administration. The work of this paper considers seven characteristic variables for
Beijing PM2.5 concentration prediction: “dew point”, “temperature”, “pressure”, “wind speed”, “snowfall”, “rainfall”, and “wind direction”, as shown in Table 1.

| Variable      | Unit |
|---------------|------|
| dew point     | °C   |
| temperature   | °C   |
| pressure      | hPa  |
| wind speed    | m/s  |
| snowfall      | h    |
| rainfall      | h    |
| Wind direction| NA   |

The dataset is divided into four parts according to the spring (March, April, May), summer (June, July, August), autumn (September, October, November), winter (December, January, February). We divided four datasets into four training sets and four test sets according to the method of data preprocessing mentioned above. We used training sets to train the prediction model corresponding to the four seasons of spring, summer, autumn, and winter, and test sets to evaluate the effect of the model.

3.2 Parameter Setting
We adjust parameters through repeated experiments. Table 2 displays the parameter settings for our model.

| Parameter Name   | Value |
|------------------|-------|
| Number of GRU units | 200   |
| Initial state    | 0.0   |
| Initial bias     | 1.0   |
| Batch size       | 250   |
| Dropout rate     | 0.5   |
| Initial learning rate | 0.001 |
| Early stopping patient | 50    |
| Gradient clipping | 5.0   |

3.3 Result and Analysis
This experiment uses the regression prediction method for the prediction of air pollutant concentration. Therefore, we use MAE (equation 7), MSE (equation 8), and RMSE (equation 9) to evaluate the model effect. After the comparative analysis of experiments, we found that the effect of MAE is better than the other two. Thus we use the training model of the MAE loss function to exhibit our prediction results.

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |f_i - y_i|
\]

(7)

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (f_i - y_i)^2
\]

(8)
\[ RMSE = \sqrt{MSE} \] 

(9)

As is depicted in Figure 3, the predicted values of the GRU models for the four seasons of spring, summer, autumn, and winter correspond to four images (a), (b), (c), and (d), respectively. The blue line represents the actual value, and the red line represents the predicted value. From the figure, we can find that the prediction accuracy of different seasons is distinctly high, verifying the effectiveness of GRU model to predict pollutant concentration. The autumn forecast sequence fits the real value best, and the summer fitting result is slightly worse but still within the acceptable range. We can find that the GRU model in this paper is not competent for predicting PM2.5 value higher than 600, which is probably because the model does not consider the primary pollution sources and other factors that have an immediate impact on PM2.5 value.

![Figure 3. Prediction Result.](image)

4. Conclusion
In this study, the GRU model was used to predict the concentration of air pollutants according to the characteristics of high-dimensional and complex air pollution time series. The experiments using PM2.5 concentration information per hour in Beijing were conducted to verify the effectiveness of the GRU model in the field of air quality prediction. The new prediction method presented in this paper is of great significance for the statistical prediction of air quality.
In the future, we will further ameliorate our model and evaluate the model with a larger dataset. Furthermore, we will take advantage of the model to predict other pollutant concentration.

Acknowledgments
This study is supported by the Scientific Research Fund of Hunan Provincial Education Department (No. 15A007) and the National Natural Science Foundation of China (No. 61202116, No. 61602504).

References
[1] Yin, W.J., Zhang, D., Yan, J., et al. (2015) Deep Learning based Air Pollutant Forecasting with Big Data. J. Chinese Journal of Environmental Management, 7(6):46-52.  
[2] Chelani, A.B., Devotta, S. (2015) Air quality forecasting using a hybrid autoregressive and nonlinear model. J. Atmospheric Environment, 40(10):1774-1780.  
[3] Zhao, H., Liu A.X., Wang, K., BAI, Z. (2009) Improved GA ANN Model for Air Quality Forecasting. J. Research of Environmental Sciences, 22(11): 1276-128  
[4] Liu, B.C., Binaykia, A., Chang, P.C., et al. (2017) Urban air quality forecasting based on multi-dimensional collaborative Support Vector Regression (SVR): A case study of Beijing-Tianjin-Shijiazhuang. J. Plos One, 12(7).  
[5] Xie, S., Qian, B., Yang, B. (2017) Influence on Input Parameters of PM2.5 Concentration Prediction Model Based on LIBSVM. J. Journal of Luoyang Institute of Science andTechnology (Natural Science Edition), 27(2):9-12.  
[6] Chung, J., Gulcehre, C., Cho, K.H., et al. (2014) Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. J. Eprint Arxiv.  
[7] Hochreiter, S., Schmidhuber, J. (1997) Long Short-Term Memory. J. Neural Computation, 9(8):1735-1780.  
[8] Cho, K., Van, M.B., Gulcehre, C., et al. (2014) Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. J. Computer Science.  
[9] Bai, H.M., Shen, R.P., Shi, H.D., et al. (2013) Forecasting Model of Air Pollution Index Based on BP Neural Network. J. Environmental Science & Technology.