IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

Cross-city PM2.5 predictions with recurrent neural network

Cross-city PM2.5 predictions with recurrent neural network

RH Zong[1], TY Zhang[2], Z Chen[3], Y Zhu[4]

[1] School of Computer Science, Sichuan University, Chengdu, Sichuan, China
[2] School of Statistics, Jiangxi University of Finance and Economics, Qingshanhu, Nanchang, China
[3] School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan, China
[4] School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China

Email: 2016141093007@stu.scu.edu.cn

Abstract. PM2.5 is inhalable particulate with a diameter less than 2.5 μM that easily enters the lungs and causes diseases and non-accidental death. The generation and dissipation of PM2.5 are strongly affected by a variety of environmental factors, thus the concentration of PM2.5 is presumably predictable with the observations of environmental conditions. This paper used multi-year meteorological and PM2.5 concentration data across multiple megacities in China (Beijing, Chengdu, and Shenyang) and sought for a universal predictive model. Our results showed that data-driven machine-learning model was able to not only capture PM2.5 dynamics at the city where the model was trained but also could be generalized to predict PM2.5 concentrations over other cities. Therefore, the modeling results indicated a universally existing predictive relationship between PM2.5 source-sink dynamics and the environmental drivers.

1. Introduction

PM2.5 are particulate matters having a diameter of fewer than 2.5 micrometers. These fine particulate matters are the critical pollutant, which could cause adverse impacts on the human health system. The size feature of PM2.5 makes it difficult to be protected from. They are small in diameter and the large surface areas that are capable of suspending in the atmosphere for a long time and carrying a large number of toxic substances. In addition, the small size enables itself to pass throw the nose hair and impair the function of lung and other parts of the body. A growing number of researches revealed that particulate matter enhanced the relevant diseases and mortality rate. For example, "Harvard Six Cities Study" demonstrates that there is a linear dependence between PM2.5 and non-accidental mortality [1].

Observational evidence came from Beijing Normal University site, from 2001 to 2006, revealed that atmospheric PM2.5 were black carbon crustal elements, nitrates, ammonium salts, and sulfates [2], which indicated that the PM2.5 concentrations were source emission controlled mainly from industrial emission, soil dust, coal combustion, secondary aerosol, and biomass burning [2]. Besides, the evolution of atmospheric PM2.5 concentration is also dependent on the variability of boundary layer depth, wind speed, air humidity, and many other environmental drivers that control the sink strength of PM2.5.

Predictive modeling of PM2.5 concentrations across multiple cities is challenging. Traditional process-based modeling framework of PM2.5 source-sink dynamics is comprised of atmospheric transport and chemical models, such as WRM-Chem and CMAQ [3]. These models can be used to predict multiple atmospheric pollutants over the whole region, but their accuracy is highly dependent on the prescribed source emissions that are very difficult to generate [3]. Data-driven times series based data mining models are the alternative numerical framework to predict atmospheric PM2.5 concentrations, including e.g., Auto Regression Moving Average, Stochastic Volatility, Stock-Watson Model, Support Vector Machine, and artificial Neural Network [4]. These frequently used machine-learning methods vary significantly by its complexity and prior assumption about the modeled system. However, none considers the influence of the system memory (previous moment) on the future dynamics. Furthermore, each model is subject to the specific condition in the city where the model is trained. Thus, it is hard to be generalized to other cities. In this study, we employ Recurrent Neural Network, to tackle the system memory of PM2.5 dynamics and aim to generate a universal machine learning models suitable to predict PM2.5 concentrations over multiple megacities in China.

2. Methodology

2.1 Data

We used the meteorology data and PM2.5 concentration data recorded in Beijing, Chengdu, and Shenyang in China from 2010 to 2015 with the hourly resolution [5]. The density distribution of the data is shown in Figure 1. The red, green and blue curves correspond to Beijing, Chengdu, and Shenyang, respectively.

The mean values of PM2.5 concentration in Beijing, Chengdu, and Shenyang are 95.83, 84.66, and 75.08, and the variances are 91.80, 57.96, and 65.96, respectively. Among them, Beijing has the highest average value, but its variance is also the highest, indicating that it is likely to be more serious when it is polluted. The mean value of Chengdu is medium, but the lowest variance indicates that its fluctuation is the smallest, revealing that PM2.5 pollution event could last longer (sink slowly) than the other two cities.

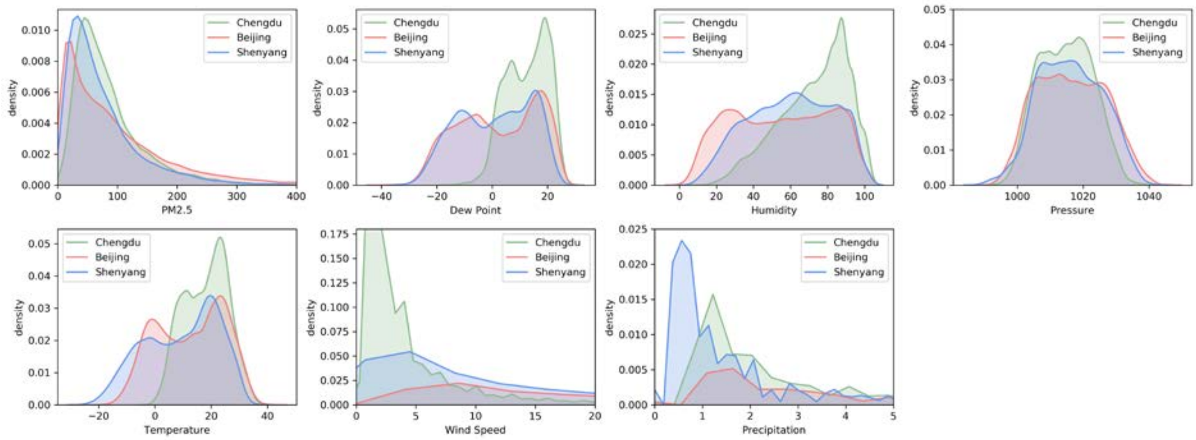

Figure 1. Density distribution of various different drivers and PM2.5 concentrations at Beijing, Chengdu, and Shenyang.

2.2 LSTM Model

We used memory-enabled Recurrent Neural Network (RNN) to predict PM2.5 concentrations across Beijing, Chengdu, and Shenyang at the same time. RNN maintains and updates the model hidden state h_t by a nonlinear activation of a linearly combining of input matrix X, previous hidden state h_{t-1} with weight parameters *W* and error *b*. Then, the output prediction matrix y is computed by multiplying the value of the hidden layer by its corresponding weighted matrix (Eq. 1, Eq. 2).

$$h_t = \sigma(W_{ih}x_t + W_{hh}h_{t-1} + b_h) \quad (1)$$

$$y_t = W_{ho}h_t + b_o \quad (2)$$

Specifically, we used Long Short-Term Memory Networks (LSTM) [6] implementation of RNN with memory cells that can add the cell value of the last time to the new model with a certain weight. LSTM can process data that is highly dependent on historical state or currently input by optimizing the value of the input gate and the forget gate.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (4)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (6)$$

$$h_t = o_t \tanh(c_t) \quad (7)$$

W in the equations is the corresponding weight matrix, and σ is activation function. In the following equations, i, f, and o are the input gate, the forget gate, and the output gate, all of which are scaled value between 0 and 1 that are typically activated by the sigmoid function. New input matrix and hidden layer matrix of the last time are activated by $\tanh$ function and multiply by the input gate i (Eq. 3, Eq. 5). New cell activation c (Eq. 5) is computed by linearly combining the product of the last c and the forget gate f (Eq. 4), and the activated input value multiplied by the input gate. The output gate o (Eq. 6) is computed by linearly combining the input matrix X, the hidden layer matrix h of the last time, and the cell activation value. Then the memory cell value activated by $\tanh$ function is multiplied by the output gate to calculate the new hidden layer matrix h (Eq. 7).

We aim to use the data from three cities (Beijing, Chengdu, and Shenyang) to build and train the LSTM model and to predict the respective PM2.5 concentration across all three cities. The input matrix X is meteorological features, and specifically includes dew point, humidity, pressure, temperature, wind speed, and precipitation. The output matrix y is the predicted PM2.5 concentration at the city where LSTM is trained. After the respective predictions, the models of the three cities are used to predict the PM2.5 concentration of the other two cities to test whether the model is scalable. We find the most representative of the models that best predict other cities. It may be possible to extend it to a wider range of forecasts in the future.

3. Results and Discussion

3.1 Optimize model performance at a single city

Since the various important hyper-parameters have significant impacts on the overall model performance, we test and compare the effects of those parameters on the model predictability. The batch size was tested with six values: 100, 200, 1000, 2000, 5000, 10000. We found that the value for the best test results was 100, while 10000 was significantly inferior to other results. As the value increases, the Pearson coefficients become smaller and smaller, which is consistent with the empirical knowledge that training fewer data at a time can lead to more accurate results.

We tested six epoch values: 10, 20, 40, 60, 80, 100, and the results are shown in the figure below (Figure 2) as an example of the experiments in different parameters. The results corresponding to 10 and 20 were significantly worse than the other results, while the results of the other values of epoch increased slightly as the value increases, which was consistent with training more times can get better results. Nevertheless, the result corresponding to 100 in the model trained in Chengdu had a large error, which may be due to over-fitting caused by excessive training. Therefore, epoch size 80 was considered the optimal parameter for selection in this experiment.

We also tested the number of nodes per layer: 10, 20, 30, 40, 50. The difference in prediction results is less obvious than the first two parameters, and there are single values do not match the overall trend. However, in general, the prediction result is more accurate as the number of nodes increases.

We also conducted leave-one-out experiment to generate insights on which input feature was more important in modeling PM2.5 dynamics. The correlation coefficients of the prediction results were not much different, but the mean square error was the smallest when the feature 'pressure' was removed, which confirmed the potential importance of pressure fields in controlling the local atmospheric transport of pollutants.

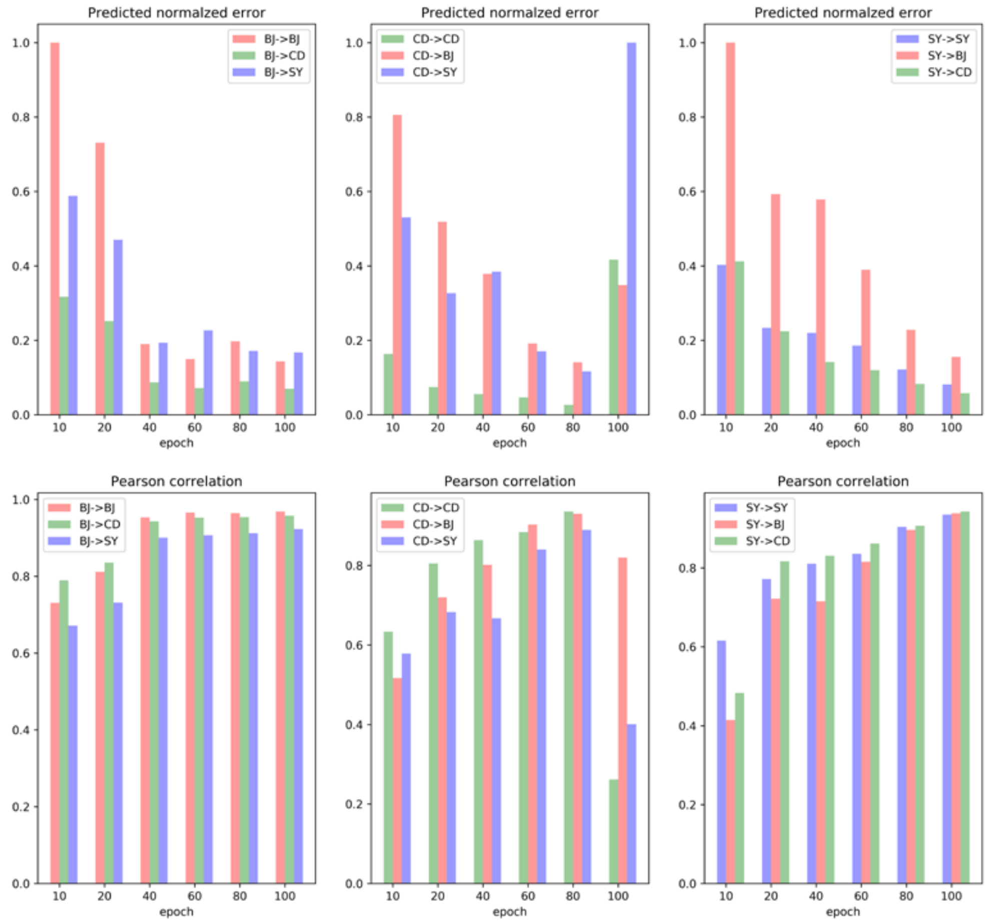

Figure 2. Example of hyper-parameter tuning of RNN PM2.5 modeling in three cities.

In Beijing, Chengdu, and Shenyang, the Pearson correlation coefficients of the predicted and tested values of the PM2.5 predictions using the best parameters were 0.975, 0.962, and 0.959, respectively. Among them, Beijing has the best results, thus, we showed the distribution of the observed versus modeled PM2.5 concentrations in Beijing as an example (Figure 3).

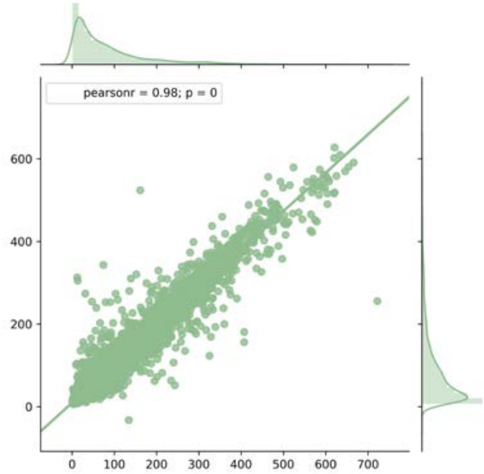

Figure 3. Best model performance in Beijing.

3.2 PM2.5 predictions across multiple cities

The second goal of this study is to evaluate the generality of RNN models from one city to another. The results of the cross-city PM2.5 predictions with those best hyper-parameters (section 3.1) across cities indicated that individual RNN trained in one city could be generalized to the other two cities (Figure 4). Specifically, using the model trained by Beijing's data, the Pearson correlation coefficients of the predicted and tested values of Beijing, Chengdu, and Shenyang were 0.975, 0.962, 0.936, respectively. With the model trained by Chengdu's data, the Pearson coefficients of the predicted and tested values of Chengdu, Beijing, and Shenyang were 0.962, 0.963, 0.940, respectively. With the model trained by Shenyang's data, the Pearson coefficients of the predicted and tested values of Shenyang, Beijing, and Chengdu were 0.959, 0.965, 0.964, respectively. Among them, RNN trained in Shenyang was the most representative model because it had better results both in Beijing and Chengdu.

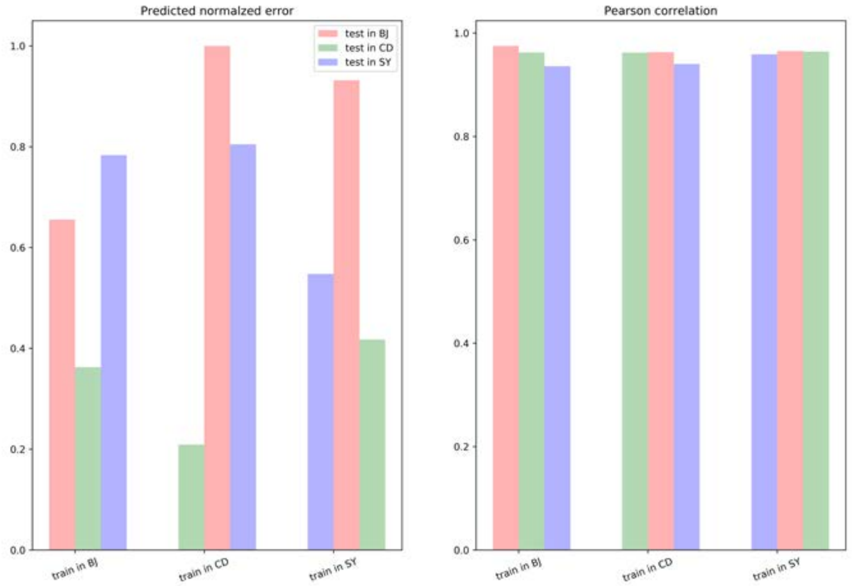

Figure 4. Across-city model performances shown by normalized error and Pearson correlation.

4. Conclusion

Particular matter pollution over megacities is increasing severely due to the rapid industrial and economic development. This paper aimed to use the advanced memory-enabled neural network to model PM2.5 dynamics across Beijing, Chengdu, and Shenyang. Our results demonstrated that multiple hyper-parameters significantly contributed to the overall performance of the model at each individual city. More importantly, the machine-learning model trained in one city could be effectively generalized to other cities, without significantly losing model performance. It implied that the internal relationship between PM2.5 source-sink dynamics and the environmental drivers were closely linked and the linkage might be universal across cities.

5. References

[1] Dockery D, Pope C, Xu X, Spengler J and Ware J 1993 An association between air pollution and mortality in six U.S. cities *NEJM*, 329(24): 1753

[2] Wang H, Zhuang Y, Wang Y, et al 2008 Long-term monitoring and source apportionment of PM2.5/PM10 in Beijing, China *Journal of Environmental Sciences*, 20 (11) :1323-1327

[3] Xia X, Zhao W, Rui X, et al 2016 A comprehensive evaluation of air pollution prediction improvement by a machine learning method *IEEE International Conference on Service Operations and Logistics*, 176-181

[4] Shen J 2012 PM2.5 concentration prediction using times series based data mining *City*, 2013, 2014

[5] Liang X, Li S, Zhang S, Huang H, Chen SX, et al 2016 PM2.5 data reliability, consistency, and air quality assessment in five Chinese cities[J] *Journal of Geophysical Research Atmospheres*, 2016, 121

[6] Gers FA, Schraudolph NN, et al 2003 Learning precise timing with LSTM recurrent networks *JMLR*, 3 (1) :115-143