Multi-Station Collaborative Spatio-Temporal Learning for PM2.5 Prediction

s2179908 (YUMING DING)

Abstract

Air pollution is a critical environmental problem worldwide, and accurate PM2.5 predictions are essential for environmental management and human health. A common approach to reasoning about future PM2.5 is through their historical air quality data. However, the impact of geography is ignored, which means that beneficial environmental information is discarded. Although some methods have been proposed to utilize information from surrounding stations, they only model spatial correlations based on geographic formulas and mathematical statistical methods, which do not effectively address all kinds of geospatial environments. In this paper, a multi-station collaborative learning with a spatio-temporal tensor fusion model for PM2.5 prediction is proposed, which introduces a novel environmental convolution mechanism and a causal convolution mechanism to model the spatial and temporal information among multiple stations, respectively. Finally, we evaluate the proposed multi-site collaborative learning model on the Beijing air pollution dataset. The experimental results demonstrate that the performance of the proposed method is not weaker than the state-of-the-art PM2.5 prediction methods.

1. Introduction

With accelerated urban development and modernization, a variety of modern scenes including deforestation, industrialization, vehicle emissions and super population explosion have made great contributions to various types of air pollution. Air pollutants are mainly composed of gaseous pollutants and particle matters(PM)(Xing et al., 2016). High levels of PM2.5(particulate matter with aerodynamic diameter $< 2.5\mu m$) have posed a severe environmental threat to both of public health and economics of China and have gained great attention from the society(Maji et al., 2018). PM2.5 are small in diameter but large in surface, good properties for carrying various toxic substances. Previous studies has shown that PM2.5 has chronic health effects on people of all age groups. Exposure to PM2.5 would lead to larger possibilities of having lung cancer(Vinikoor-Imler et al., 2011) and cardiovascular disease(Xing et al., 2016), as well as higher risks of ischemic heart disease(Alexeeff et al., 2021), acute bronchitis(Xie et al., 2014), nervous system breakdown(Zhang et al., 2018) and asthma(Tecer et al., 2008). Due to the adverse health effects, PM2.5 has affected economic activities greatly in terms of gross domestic product(GDP)(Maji et al., 2018). Therefore, accurate prediction of PM2.5's mass concentration has played a significant role in making atmospheric management decisions(Zhang & Li, 2015). Forecasting the concentration of PM2.5 in advance is the basis for strengthening the prevention and control of air pollution and achieving comprehensive environmental management, which is of great significance to public health and governmental decision-making(Zheng et al., 2015).

A wide variety of methods have been used to forecast regional air quality. These research methods can be roughly divided into 2 categories as deterministic methods and statistical methods. Deterministic methods (C., 1996; D., 1999; I. et al., 2001; Grell et al., 2005; Kim et al., 2010; Cheng et al., 2013) typically rely on current knowledge of atmospheric physics and chemical processes and use mathematical methods to build a numerical model of the dilution and diffusion of pollutant concentrations in the atmosphere. High-speed calculations and simulations can be used to predict the dynamic changes in mass concentrations of air pollutants. to predict. However, it can be difficult to explain the non-linearity and heterogeneity of many influences due to the lack of highly informative source data (e.g. emission sources and amounts) and the absence of important information on physical processes. In addition, the different types and large number of parameters to be determined would also lead to limited prediction accuracy, as many of them are determined by experience in the field.

With the increasing availability of historical data collected from multiple air pollution monitoring stations, data-based statistical methods for air pollutant prediction have attracted growing interest. Statistical methods use extensive historical data to build statistical models that serve as links between multiple explanatory variables and PM2.5 prediction results. Aerosol optical depth (AOD)(W. et al., 2017) is one of the most typical satellite data products for predicting the surface distribution of particulate pollutants. Although the problem of limited prediction time due to missing values and low temporal resolution has been addressed by the introduction of geostationary satellites, the small coverage of the global monitoring area have significantly affected the utility of these models. Considering this for statistical models, ground-based data including both of spatial and temporal information would be of greater use in the prediction task.

Using the ground-based data, several studies proposed linear statistical models((Ma et al., 2014),(He & Huang, 2018),(A. et al., 2017),(Donnelly et al., 2015),(Xiao et al., 2018)) to represent the linear relationship between feature variables and PM2.5 prediction levels. However, given that linear constraint may not fit the real situation of multiple feature variables playing both of independent and collaborative role in affecting PM2.5 level, it is better to take into account nonlinearity of the feature space when building statistical models. K-nearest neighbour regression (KNN), Support Vector Regression (SVR) (Choudhury et al., 2022), Multi-Layer Perceptron (MLP) model, Random Forest models (RF) (Murugan & Palanichamy, 2021) and Artificial Neural Network (ANN) have been widely used in air pollutant prediction. However, there is a serious problem - these models cannot integrate and analyze heterogeneous data from multiple sources that have a large impact on air pollution levels.

With an aim of representing the complex spatial-temporal dependency in air pollutant historical data, deep learning methods which are capable of effectively discovering non-linear relationship in high-dimensional feature space are adopted for air pollution prediction. Back Propagation Neural Network(BPNN), Gradient Boosting Regression Tree (GBRT), Long-short term memory (LSTM)-based network(Tian et al., 2021). Among those deep learning-based prediction models, the LSTM model works well in capturing nonlinear temporal auto-correlations from data and is widely used in prediction tasks. Improvements of corresponding weight parameters, adding optimization algorithms or adding auxiliary data has been used for improving the forecasting performance of the existing methods. Hybrid models that combine two or more single models have been studied to further improve the prediction accuracy of the model.

Convolutional Neural Network(CNN) is a network composed of convolutional layer, pooling layers and fully-connected layers. CNNs are widely-used in feature extraction tasks such as image classification, face detection and video recognition. For sequence data modeling tasks, a general method called Temporal Convolutional Network(TCN) is developed(Lea et al., 2017). The model designed for spatiotemporal input features is the dilated TCN, whose main ingredient are causal convolutions. The same as audio generation task finished in WaveNet(oord et al., 2016), causal convolutions can prevent the violation of the ordering in which the data is modelled. The prediction result of the next time step will only depend on historical and current data, and no data from future time step will involve in the model training process at current time step. In order to incorporate a wider range of historical data rather than the several nearest data, a dilated convolution which skips input values with a certain step is applied. In addition, stacked extended convolution allows networks to have large receptive fields with multiple layers, while maintaining input resolution across the network as well as computational efficiency.

With an aim of capturing both of spatial and temporal features, we created a hybrid model CSTNet. CSTNet used CNN for the spatial dependency representation between various stations and dilated TCN for temporal feature extraction.

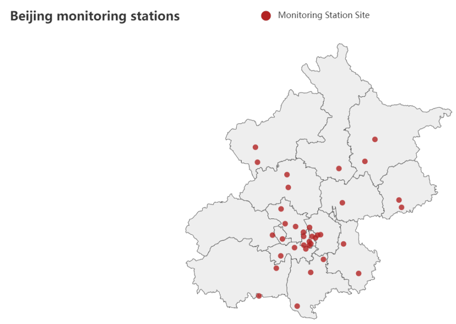

Figure 1. Monitoring stations in Beijing

1029	1030	1006	1007	1027	1032
1031	1024	1008	1013	1023	1028
1025	1001	1015	1021	1014	1026
1004	1002	1009	1020	1012	1033
1005	1003	1016	1010	1022	1019
1036	1035	1011	1017	1018	1034

Figure 2. Grid representation of the 36 air quality monitoring stations in Beijing

2. Data set and task

2.1. Dataset

2.1.1. Data Source

In this project, we mainly focus on the historical PM2.5 concentration changes in Beijing from May 1st 2014 to April 30th 2015. We collected hourly PM2.5 concentration data and daily PM2.5 averages from 36 air quality monitoring stations(Figure 1) in Beijing during one year. Our dataset was sourced from the National City Air Quality Real-Time Publishing Platform.

2.1.2. Abstract to a square distribution

In order to fit the LSTM encoder and convolutional layer, we abstracted the 36 air quality monitoring stations data to a regular square distribution as Figure 2. According to Tobler's First Law of Geography(Tobler, 1970), we reorganized the unevenly distributed 36 stations into a square uniform distributions based on the latitude and longitude of each station.

2.2. Data Pre-processing

2.2.1. Time step generation

The original dataset contains information of different stations' pollution data including station ID, time, PM2.5 level, PM10 level, NO2 level, CO level, O3 level, SO2 level, weather, temperature, pressure, humidity, wind speed, wind direction. (All the information apart from station ID and time would be called as feature in the next section.) Then we need to generate a timestamp for each row of the dataset according to the time of date format. The formula for timestamp generation is

$$timestamp = (original{-}date{-}70*365{-}19)*86400{-}8*3600 \quad (1)$$

then turn the timestamp column in excel into the regular format so timestamp can display normally. And for utility of the data, we scale down the timestamp to time step using this formula:

$$timestep = (timestamp - 1398898800) + 12 \quad (2)$$

After the generation of timestamp and time step, we performed an ascending sort on station ID.

2.2.2. Input tensor generation

For all the data, there are 2 arrays staTracks and staTraj. staTracks is an array of dataset number rows and station number columns. In each cell, there is an array of 12*N size, composed of records of 12 features at this station in N time steps in this dataset. staTraj is a large array. Each row represents 12 cells of data that includes 12 feature information of target station and neighboring stations at current time step. As for each column, the meanings are listed in the Appendix.

Each cell contains certain feature information of target station and neighboring stations in the grid. For representation in a lower dimension, we transferred the 2D grid feature representation to a 1D feature vector. The index of each station is listed in Appendix.

2.3. Task

Given the input feature information, the task of our model is to generate PM2.5 prediction concentration values in one hour.

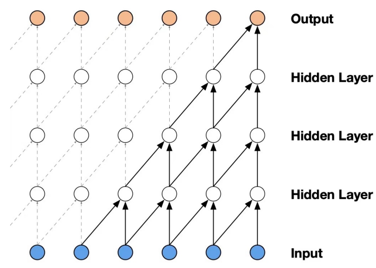

Figure 3. Basic structure of plain causal convolutional neural network

3. Methodology

3.1. Inputs and outputs

Given N historical time-step air quality data set for $H \times W$ monitoring stations, the input features can be represented as a 4D vector, $\boldsymbol{X} \in \mathbb{R}^{N \times F \times H \times W}$, where F represents the environmental features associated with PM2.5 concentrations, and H and W represent the vertical and horizontal coordinates of the monitoring stations located in the regional spatial station representation, respectively. X can also be represented as a vector of N feature matrices, $[\boldsymbol{X}_{t_1}, \ldots, \boldsymbol{X}_{t_i}, \ldots, \boldsymbol{X}_{t_N}]$, where $X_{t_i} \in \mathbb{R}^{F \times H \times W}$ denotes the feature matrix of all monitoring stations at the ith time step $t = t_1, t_2, \ldots, t_N$.

The output of the model is the predicted PM2.5 value for the next 1 hour for the predicted station.

3.2. Architecture of the proposed network

The architecture of the proposed CSTNet is illustrated in Figure 4. The model has two parts. The first part is the multi-station convolutional environment pooling module, as illustrated in Fig. 4(A), and the second part is the causal convolutional network, as illustrated in Fig. 4(B). The details of these two parts are described in the following two subsections.

3.2.1. Regional Spatial Station Representation Method and Frame of Reference

Due to the uneven distribution of the 36 stations in Beijing, the air quality data of the stations are not available as direct inputs to our network model. We employ the regional spatial station representation proposed in the paper (Wang et al., 2021) to reorganize the 36 stations in a grid of 6 rows * 6 columns according to Tobler's first law of geography (Tobler, 1970; Wang et al., 2019). The station locations (coordinates in the grid) are taken as inputs to our proposed network model.

Considering the excellent capability of CNN in image processing to extract image features, we adopted CNN for mining complex spatial relationships between target station and neighbouring stations. In our study, CNN is employed to capture and quantify the complex spatial correlations of

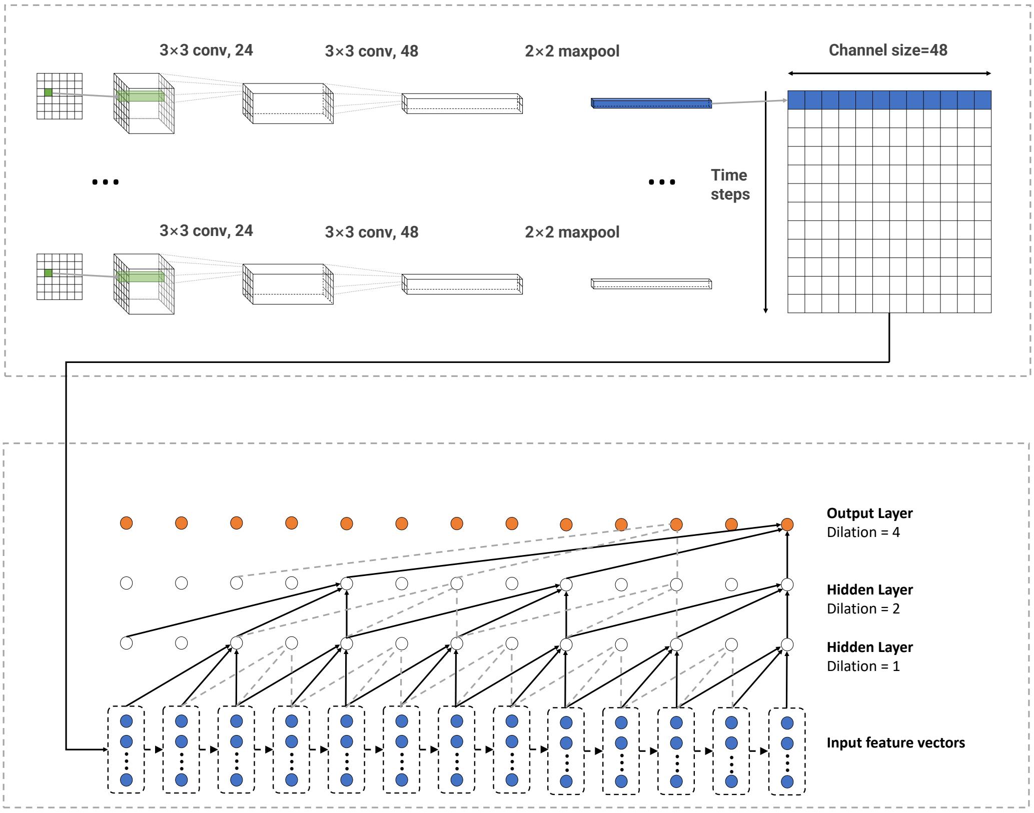

Figure 4. CSTNet PM2.5 prediction model architecture, which includes two main parts: (A) Multi-station convolutional environment pooling module; (B) Causal convolutional network with kernel size = 3 and d (dilations) = 1, 2, and 4

air quality. In a two-dimensional convolutional operation, the computation is performed within the spatial neighborhood of each cell of a grid represented by a regional spatial station, which represents the extraction of a certain feature pattern within the spatial neighborhood(He et al., 2018). Compared with traditional geographic methods, such as selecting stations with a high degree of mutual influence based on Pearson correlation coefficients, the convolution operation is effective in learning the spatial information of a specific geographic environment.

3.2.2. Causal convolutional network module

Generally, in traditional neural networks, each layer of neurons is in the form of fully connected. It is not difficult to find that fully connected networks violate the basic constraint of temporal succession. The output leaning forward (previous moment) neuron is connected to the input leaning backward (later moment) neuron, which is not supposed to be allowed in processing time series. As illustrated in the Figure 3, an acceptable approach is to remove some of the connections in the network layer by layer through a mask, preserving those connections from the front-to-back neurons, so that the network satisfies the principle of temporal dependence.

However, such a causal convolutional network architecture does not cover much historical information. For example, in the Figure 3, the output of the model can only be calculated based on the last 5 neurons in the first input layer, and the input information further ahead is deserted. A simple approach to extend the coverage of historical data by causal convolution is to increase the network depth. The capture field of historical information expands as the number of layers of the network increases. However, this approach is limited to expanding the capture field of historical information on a linear time scale. Moreover, as the number of layers and parameters in the network increase, this also raises the difficulty of finding the optimal solution during model training. Therefore, we employ dilated convolution to expand the reception field of the causal convolutional network for historical information according to the method proposed by (Van Den Oord et al., 2016). As illustrated above, the introduction of the dilated mechanism enables the current causal convolutional network to expand the reception field of the network by increasing the convolutional kernel size or increasing the dilated parameter with constant depth, which also improves the capability of the causal convolutional network to model temporal sequence inference tasks for long-term memory.

4. Experiments

4.1. Experimental settings

The model is trained in an end-to-end fashion using the Adam(Kingma & Ba, 2014) optimizer performing stochastic gradient descent optimization with a learning rate of 0.001. Ideally, we minimize the loss function Mean Square Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2. \tag{3}$$

In the convolutional environment pooling module, the convolutional kernel size is 3*3 and the numbers of filters for two convolutions are 24 and 48. In the causal convolution module, we use a 4-layer causal convolutional network with a kernel size of 3 and the hidden layer size of 32. We activate all layers using the ReLU function and dropout with p=0.75. The model is implemented using PyTorch 1.11.0.

We chose the RMSE, MAE and R^2 as the evaluation metrics of PM2.5 prediction results. The formula for calculating these three measures is as follows

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2}, \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i|, \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2}. \tag{6}$$

.

4.2. Model performance results

We chose the monitoring station 1007 in Beijing as the centre station and applied the proposed model CSTNet on predicting hourly PM2.5 concentration values from May 2014 to April 2015 on the test dataset as an example. Figure 5 shows the result of comparison between the predicted hourly PM2.5 concentrations and the true observed values at the station 1007. We can obtain that the predicted values represented by the red line have roughly the same trend as the observed values represented by the blue line except for a long-period observations missing in December 2014. Thus the predicted results of the CSTNet were sound and the model is feasible for PM2.5 concentration prediction.

The prediction results for the six example monitoring sites shown in the Table 1 represent that the proposed model achieved good performance at these sites as well. Figure 6 shows the correlation (R^2) between the predicted PM2.5 concentrations and true observed values on the evaluation dataset of the monitoring station 1007. We can see that CSTNet performs a good fit on the validation dataset.

4.3. Comparison with other baselines

We choose ANN, GRU, LSTM and ST-CausalConvNet(Tian et al., 2021) as four baseline experiments. Table 2 shows the prediction accuracy of the chosen centre station 1007 using different methods. Our proposed model CSTNet outperforms all the traditional neural networks (ANN, GRU and LSTM) and almost achieves the results of the best performing state of ST-CausalConvNet. However, unlike ST-CausalConvNet, which needs to calculate Pearson correlation coefficient between each other stations and the target site and test the accuracy performance of the network under different Pearson correlation coefficient thresholds to select the best threshold for the ST-CausalConvNet, CSTNet learns the spatial relationships between monitoring stations completely automatically. Another major drawback of the ST-CausalConvNet model is that each time the target monitoring site to be predicted is changed, the process of calculating Pearson coefficient and testing candidate thresholds needs to be repeated. In contrast, our proposed model uses Convolutional Neural Network to learn spatial relationships between monitoring stations without the need for overly complex repetitive experiments. And from the experimental results, this automatic learning is very effective (RMSE = 19.101 µg/m^3, MAE = 12.635 µg/m^3 and R^2 = 0.916) and almost achieves the best accuracy performance of the ST-CausalConvNet using the optimal threshold (RMSE = 18.847 µg/m^3, MAE = 11.896 µg/m^3 and R^2 = 0.923). From this we can conclude that CSTNet performs well and solves the disadvantage that ST-CausalConvNet requires repeated testing.

Station	RMSE (µg/m^3)	MAE (µg/m^3)	R^2
1001	18.374	12.145	0.928
1007	19.101	12.635	0.916
1013	18.538	12.974	0.921
1020	17.983	11.867	0.932
1025	18.427	12.176	0.924
1030	17.832	11.786	0.938

Table 1. Performance of the CSTNet model for 6 example stations

Model	RMSE (µg/m^3)	MAE (µg/m^3)	R^2
ANN	36.932	25.254	0.702
GRU	29.253	20.348	0.814
LSTM	20.632	13.153	0.896
ST-CausalConvNet	18.004	12.051	0.927
CSTNet	**19.101**	**12.635**	**0.916**

Table 2. Performance comparasion with other models

5. Related work

5.1. Spatiotemporal prediction

Recently, there are multiple studies introducing PM2.5 forecasting models based on spatiotemporal feature information. For temporal data, most(Wang et al., 2021; Zhao et al., 2019; Wen et al., 2019; Zhou et al., 2019) used LSTM for representation of sequential feature information. As for representation of spatial feature information,

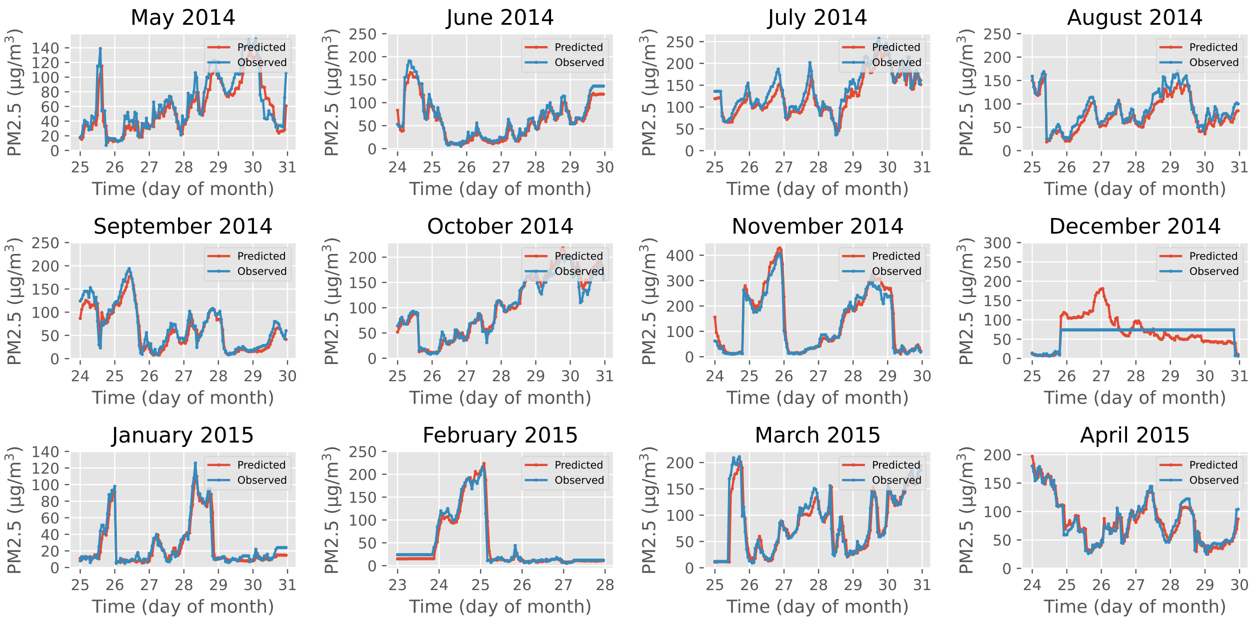

Figure 5. Results of PM2.5 concentration prediction values compared with observation values

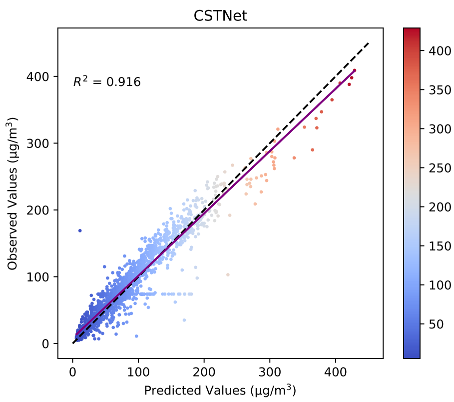

Figure 6. Correlations between the predicted and observed PM2.5 concentrations with the CSTNet model for the 1007 station

various approaches has been adopted. ConvLSTM and CR-LSTM(Wang et al., 2021) added a convolution operation to the traditional LSTM model, giving it a property of capturing the spatial correlations among neighboring monitoring stations similarly to the CNN. LSTM-FC(Zhao et al., 2019) adopted a fully connected layer to integrate spatial information. C-LSTME(Wen et al., 2019) and ST-CausalConvNet(Zhang et al., 2021).

5.2. Pearson correlation coefficient

(Zhang et al., 2021) used the method of calculating Pearson correlation coefficients as measurement of the degree of correlation between stations. Thresholds are set manually to filter out the surrounding stations that have a high impact on the air quality of the target station. Only the historical data of the stations that passed the filtering are used as inputs for the training model. The combination of this method and causal convolutional network they used demonstrated advanced prediction performance. However, this approach has two limitations.

1. When a high threshold is set, the surrounding stations with high correlation are removed, which means that the beneficial environmental information is deserted.

2. The ideal threshold of Pearson correlation coefficient needs to be set by training the model several times for comparison experiments to explore the best parameters.

3. Different geographic environments have different optimal thresholds.

When large-scale data sets are applied to model training, this feature engineering method of manually setting parameters and thus filtering correlated stations appears to be difficult to implement. Therefore, we propose a convolutional environment pooling module to replace the method of filtering based on Pearson correlation coefficients. The convolutional neural network can learn the geospatial dependency between stations, which actually uses the convolutional neural network to filter stations instead of manually setting Pearson thresholds.

5.3. 3D CNN

Models such as (Lee & Shin, 2019) uses 3D CNN for representing temporal feature information. With an aim of preventing information leakage from future data to current data, we decided not to use 3D CNN, which incorporates information of all of previous, current and future data in the convolution operation, for constructing spatiotemporal PM2.5 forecasting model.

6. Conclusions

This study constructed a hybrid model CSTNet for PM2.5 pollutants prediction. This model takes into account both of spatial and temporal air pollution feature data to make predictions of PM2.5 concentration. For representation of spatial feature information, CNN which has shown great ability in extracting features is applied on input feature data, instead of traditional mathematical formula such as the Pearson correlation. Compared to model with the Pearson correlation, which has to be trained several times and checked for correlation threshold manually to maximize the prediction accuracy, CSTNet has shown greater transferability on different training datasets. Furthermore, a dilated Causal Convolution Network functions as extractor of temporal features while effectively preventing information leakage from future to past compared with other deep learning models. The experimental results show that our proposed CSTNet model has good accuracy in predicting the hourly PM2.5 concentration in Beijing. Compared with the baseline models, CSTNet has a significant improvement over ANN, GRU, and LSTM by decreasing the RMSE by 48.3%, 34.7%, and 7.4%, respectively. Meanwhile the performance of CSTNet is comparable to that of ST-CausalConvNet in the best-case scenario, and the repetitive parameter selection effort is omitted.Finally, we conclude that the proposed CSTNet network is a effective method for forecasting PM2.5 concentration.

References

A., Shamsoddini, M.R., Aboodi, and J., Karami. Tehran air pollutants prediction based on random forest feature selection method. volume 42, pp. 483 – 488, 2017. doi: 10.5194/isprs-archives-XLII-4-W4-483-2017. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85032384989&doi=10.5194%2fisprs-archives-XLII-4-W4-483-2017&partnerID=40&md5=0f614a7960f769643d954f1f978ee2b1. Cited by: 12; All Open Access, Gold Open Access.

Alexeeff, Stacey E, Liao, Noelle S, Liu, Xi, Van Den Eeden, Stephen K, and Sidney, Stephen. Long-term pm2. 5 exposure and risks of ischemic heart disease and stroke events: review and meta-analysis. *Journal of the American Heart Association*, 10(1):e016890, 2021.

C., Coats. Fast emissions modeling with the sparse matrix operator kernel emissions (smoke) modeling system. *Air Waste Management Association Pittsburgh Pa*, 13:584 – 588, 1996. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85027972764&partnerID=40&md5=3e61366b923653d995c07318b3017a95. Cited by: 1.

Cheng, Zhen, Wang, Shuxiao, Jiang, Jingkun, Fu, Qingyan, Chen, Changhong, Xu, Bingye, Yu, Jianqiao, Fu, Xiao, and Hao, Jiming. Long-term trend of haze pollution and impact of particulate matter in the yangtze river delta, china. *Environmental Pollution*, 182:101–110, 2013. ISSN 0269-7491. doi: https://doi.org/10.1016/j.envpol.2013.06.043. URL https://www.sciencedirect.com/science/article/pii/S0269749113003679.

Choudhury, Amartya, Middya, Asif, and Roy, Sarbani. *A Comparative Study of Machine Learning and Deep Learning Techniques in Forecasting Air Pollution Levels*, pp. 607–619. 01 2022. ISBN 978-981-16-5119-9. doi: 10.1007/978-981-16-5120-5_46.

D., Byun. Science algorithms of the epa models-3 community multiscale air quality (cmaq) modeling system. *J. Jpn. Soc. Atmos. Environ.*, 43:79 – 91, 1999. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061178852&partnerID=40&md5=10188340b170581187c90d4d3c78fafd. Cited by: 1.

Donnelly, Aoife, Misstear, Bruce, and Broderick, Brian. Real time air quality forecasting using integrated parametric and non-parametric regression techniques. *Atmospheric Environment*, 103:53–65, 2015. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2014.12.011. URL https://www.sciencedirect.com/science/article/pii/S1352231014009595.

Grell, Georg A., Peckham, Steven E., Schmitz, Rainer, McKeen, Stuart A., Frost, Gregory, Skamarock, William C., and Eder, Brian. Fully coupled "online" chemistry within the wrf model. *Atmospheric Environment*, 39(37):6957–6975, 2005. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2005.04.027. URL https://www.sciencedirect.com/science/article/pii/S1352231005003560.

He, Jialv, Li, Xia, Yao, Yao, Hong, Ye, and Jinbao, Zhang. Mining transition rules of cellular automata for simulating urban expansion by using the deep learning techniques. *International Journal of Geographical Information Science*, 32(10):2076–2097, 2018.

He, Qingqing and Huang, Bo. Satellite-based mapping of daily high-resolution ground pm2.5 in china via space-time regression modeling. *Remote Sensing of Environment*, 206:72–83, 2018. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2017.12.018. URL https://www.sciencedirect.com/science/article/pii/S003442571730593X.

I., Bey, D.J., Jacob, R.M., Yantosca, J.A., Logan, B.D., Field, A.M., Fiore, Q., Li, H.Y., Liu, L.J., Mickley, and M.G., Schultz. Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research Atmospheres*, 106(D19):23073 – 23095, 2001. doi: 10.1029/2001JD000807. URL https://www.scopus.

com/inward/record.uri?eid=2-s2.0-0034761947&doi=10.1029%2f2001JD000807&partnerID=40&md5=0dad095fe9bf69df59a51d11ddf89586. Cited by: 1458; All Open Access, Bronze Open Access, Green Open Access.

Kim, Yunhee, Fu, Joshua S., and Miller, Terry L. Improving ozone modeling in complex terrain at a fine grid resolution: Part i – examination of analysis nudging and all pbl schemes associated with lsms in meteorological model. *Atmospheric Environment*, 44(4):523–532, 2010. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2009.10.045. URL https://www.sciencedirect.com/science/article/pii/S1352231009009200.

Kingma, Diederik P and Ba, Jimmy. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Lea, Colin, Flynn, Michael D, Vidal, Rene, Reiter, Austin, and Hager, Gregory D. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 156–165, 2017.

Lee, Seonggu and Shin, Jitae. Hybrid model of convolutional lstm and cnn to predict particulate matter. *International Journal of Information and Electronics Engineering*, 9(1), 2019.

Ma, Zongwei, Hu, Xuefei, Huang, Lei, Bi, Jun, and Liu, Yang. Estimating ground-level pm2.5 in china using satellite remote sensing. *Environmental Science & Technology*, 48(13):7436–7444, 2014. doi: 10.1021/es5009399. URL https://doi.org/10.1021/es5009399. PMID: 24901806.

Maji, Kamal Jyoti, Ye, Wei-Feng, Arora, Mohit, and Shiva Nagendra, S.M. Pm2.5-related health and economic loss assessment for 338 chinese cities. *Environment International*, 121:392–403, 2018. ISSN 0160-4120. doi: https://doi.org/10.1016/j.envint.2018.09.024. URL https://www.sciencedirect.com/science/article/pii/S0160412018307785.

Murugan, Rishanti and Palanichamy, Naveen. Smart city air quality prediction using machine learning. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1048–1054, 2021. doi: 10.1109/ICICCS51141.2021.9432074.

oord, Aaron, Dieleman, Sander, Zen, Heiga, Simonyan, Karen, Vinyals, Oriol, Graves, Alex, Kalchbrenner, Nal, Senior, Andrew, and Kavukcuoglu, Koray. Wavenet: A generative model for raw audio. 09 2016.

Tecer, Lokman Hakan, Alagha, Omar, Karaca, Ferhat, Tuncel, Gürdal, and Eldes, Nilufer. Particulate matter (pm2.5, pm10-2.5, and pm10) and children's hospital admissions for asthma and respiratory diseases: A bidirectional case-crossover study. *Journal of Toxicology and Environmental Health, Part A*, 71(8):512–520, 2008.

Tian, Hao, Zhao, Yongquan, Luo, Ming, He, Qingqing, Han, Yu, and Zeng, Zhaoliang. Estimating pm2.5 from multisource data: A comparison of different machine learning models in the pearl river delta of china. *Urban Climate*, 35:100740, 2021. ISSN 2212-0955. doi: https://doi.org/10.1016/j.uclim.2020.100740. URL https://www.sciencedirect.com/science/article/pii/S2212095520302728.

Tobler, Waldo R. A computer movie simulating urban growth in the detroit region. *Economic geography*, 46 (sup1):234–240, 1970.

Van Den Oord, Aäron, Dieleman, Sander, Zen, Heiga, Simonyan, Karen, Vinyals, Oriol, Graves, Alex, Kalchbrenner, Nal, Senior, Andrew W, and Kavukcuoglu, Koray. Wavenet: A generative model for raw audio. *SSW*, 125:2, 2016.

Vinikoor-Imler, Lisa C., Davis, J. Allen, and Luben, Thomas J. An ecologic analysis of county-level pm2.5 concentrations and lung cancer incidence and mortality. *International Journal of Environmental Research and Public Health*, 8(6):1865–1871, 2011. ISSN 1660-4601. doi: 10.3390/ijerph8061865. URL https://www.mdpi.com/1660-4601/8/6/1865.

W., Wang, Z., Wang, Q., Guo, C., Zhang, and P., Sun. Doped expectation propagation for low-complexity message passing based detection. *Electronics Letters*, 53(6):403 – 405, 2017. doi: 10.1049/el.2016.3467. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85015756891&doi=10.1049%2fel.2016.3467&partnerID=40&md5=f21606b1e2ca33401cf8b3bfadb1f6ae. Cited by: 1.

Wang, Weilin, Zhao, Suli, Jiao, Limin, Taylor, Michael, Zhang, Boen, Xu, Gang, and Hou, Haobo. Estimation of pm2. 5 concentrations in china using a spatial back propagation neural network. *Scientific reports*, 9(1):1–10, 2019.

Wang, Weilin, Mao, Wenjing, Tong, Xueli, and Xu, Gang. A novel recursive model based on a convolutional long short-term memory neural network for air pollution prediction. *Remote Sensing*, 13(7):1284, 2021.

Wen, Congcong, Liu, Shufu, Yao, Xiaojing, Peng, Ling, Li, Xiang, Hu, Yuan, and Chi, Tianhe. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Science of the total environment*, 654:1091–1099, 2019.

Xiao, Lu, Lang, Yichao, and Christakos, George. High-resolution spatiotemporal mapping of pm2.5 concentrations at mainland china using a combined bme-gwr technique. *Atmospheric Environment*, 173:295–305, 2018. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2017.10.062. URL https://www.sciencedirect.com/science/article/pii/S1352231017307355.

Xie, Yuan-Bo, Chen, Juan, and Li, Wei. An assessment of pm2. 5 related health risks and impaired values of beijing residents in a consecutive high-level exposure during heavy haze days. *Huan Jing ke Xue= Huanjing Kexue*, 35(1):1–8, 2014.

Xing, Yu-Fei, Xu, Yue-Hua, Shi, Min-Hua, and Lian, Yi-Xin. The impact of pm2.5 on the human respiratory system. *Journal of thoracic disease*, 8:E69–E74, 02 2016. doi: 10.3978/j.issn.2072-1439.2016.01.19.

Zhang, Lei, Na, Jiaming, Zhu, Jie, Shi, Zhikuan, Zou, Changxin, and Yang, Lin. Spatiotemporal causal convolutional network for forecasting hourly pm2. 5 concentrations in beijing, china. *Computers & Geosciences*, 155: 104869, 2021.

Zhang, Tianliang, Zheng, Xinrui, Wang, Xia, Zhao, Hui, Wang, Tingting, Zhang, Hongxia, Li, Wanwei, Shen, Hua, and Yu, Li. Maternal exposure to pm2. 5 during pregnancy induces impaired development of cerebral cortex in mice offspring. *International journal of molecular sciences*, 19(1):257, 2018.

Zhang, Ying and Li, Zhengqiang. Remote sensing of atmospheric fine particulate matter (pm2.5) mass concentration near the ground from satellite observation. *Remote Sensing of Environment*, 160:252–262, 2015. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2015.02.005. URL https://www.sciencedirect.com/science/article/pii/S0034425715000516.

Zhao, Jiachen, Deng, Fang, Cai, Yeyun, and Chen, Jie. Long short-term memory-fully connected (lstm-fc) neural network for pm2. 5 concentration prediction. *Chemosphere*, 220:486–492, 2019.

Zheng, Yu, Yi, Xiuwen, Li, Ming, Li, Ruiyuan, Shan, Zhang, Chang, Eric, and Li, Tianrui. Forecasting fine-grained air quality based on big data. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.

Zhou, Yanlai, Chang, Fi-John, Chang, Li-Chiu, Kao, I-Feng, and Wang, Yi-Shin. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *Journal of cleaner production*, 209: 134–145, 2019.

7. Appendix

Station	Index
1	PM2.5 level
2	PM10 level
3	NO2 level
4	CO level
5	O3 level
6	SO2 level
7	weather
8	temperature
9	pressure
10	humidity
11	wind speed
12	wind direction

Station	Index
1029	1
1030	2
1006	3
1007	4
1027	5
1032	6
1031	7
1024	8
1008	9
1013	10
1023	11
1028	12
1025	13
1001	14
1015	15
1021	16
1014	17
1026	18
1004	19
1002	20
1009	21
1020	22
1012	23
1033	24
1005	25
1003	26
1016	27
1010	28
1022	29
1019	30
1036	31
1035	32
1011	33
1017	34
1018	35
1034	36