2016 15th IEEE International Conference on Machine Learning and Applications

# Forecasting PM2.5 Concentration using Spatio-Temporal Extreme Learning Machine

Bo Liu, Shuo Yan, Jianqiang Li, Yong Li
School of Software Engineering, Beijing University of Technology, Beijing 100124, China
boliu@bjut.edu.cn, yanshuok@163.com, lijianqiang@bjut.edu.cn, li.yong@bjut.edu.cn

*Abstract*—**In recent years, air quality has become a severe environmental problem in China. Since bad air quality brought significant influences on traffic and people's daily life, how to predict the future air quality precisely and subtly, has been an urgent and important problem. In this paper, a Spatio-Temporal Extreme Learning Machine (*STELM*) method is proposed for air quality prediction. STELM considers temporal and spatial characteristics of air quality data and related meteorological data, constructs a prediction model based on ELM, and realizes air quality prediction with more than 80% precision. A prototype system is implemented and the experiments on practical air quality data in Beijing validate the effectiveness of our method and system.**

***Keywords-PM2.5 concentration; prediction; extreme learning machine***

## I. INTRODUCTION

In the big data era, large amounts of temporal and spacial data have been accumulated in environment, meteorology, traffic, medicine, etc. Generally speaking, a spatio-temporal series are usually used to describe a variable which changes along with time and spatial location. For instance, PM2.5 concentration is monitored and recorded every hour in 35 air quality monitoring stations in Beijing.

Spatio-temporal modeling aims at describing a relationship between spatio-temporal variables and predicting the value of spatio-temporal series. This paper proposes a novel method based on ELM (Extreme Learning Machine) to predict air quality. ELM is proposed by Guang Bin Huang [1], in which the connection weight between input layer and hidden layer is generated randomly, and the threshold value of hidden layer is not adjusted during the training process. It is easy to obtain the unique optima, and we only need to set the number of hidden layer neurons. Compared to traditional training method, it has faster learning rate and better generalization performance.

So far, there are mainly two methods on air quality prediction, one is numerical simulation, and the other is statistical regression. Numerical simulation uses physical and chemical equation to simulate the pollution process, and then predicts air quality. It needs plenty of input parameters and complex calculation. Statistical regression has two branches, logistic regression and artificial intelligence. They need fewer input parameters and simple calculation, but the prediction accuracy is fewer than that of numerical simulation.

In terms of artificial intelligence, ANN (Artificial Neural Net) is widely used. Most existing ANN methods need many configuration parameters, and it is time-consuming to train an ANN, and more importantly, it is easy to fall into local optima, so the prediction accuracy may be influenced by input data and cannot meet the actual needs. Besides, most models didn't take meteorology especially wind into account. However, wind direction and wind force have strong influence on PM2.5. In practice, due to large dataset and low processing speed, the execution time is relatively long.

In order to overcome the drawbacks of existing methods, the authors propose Spatio-Temporal Extreme Learning Machine (*STELM*) method for air quality prediction, which can improve both the accuracy and processing speed. STELM is based on ELM model, so it has the advantages of fast training, less configuration parameters, and ease of obtaining global optima. In order to further improve the prediction accuracy, wind force and wind direction are brought into modeling. For large dataset, we can separate spatio-temporal data into different subareas by clustering to improve the training speed and accuracy efficiently.

## II. RELATED WORK

As big data emerges, data mining has brought new solutions to air quality prediction. Recently logistic regression is seldom used by researchers; instead, ANN is much more popular. Lal *et al.* [2] and Nejadkoorki *et al.* [3] has proved that ANN is a promising method on air quality prediction. But training an ANN is time-consuming and always has local optima. Russo *et al.* [4] used a set of stochastic variables that represent the relevant information on a multivariate stochastic system as input for an ANN model for air quality forecasting. It reduces the training time without decreasing the accuracy. Jiang *et al.* [5] proposed SOCNN (self-organizing competitive neural network) based on the self-organizing clustering of samples and improved the accuracy efficiently. Samia *et al.* [6] combined ARIMA (Autoregressive Integrated Moving Average) and ANN. Kuma *et al.* [7] used PCA (Principal Component Analysis) and ANN to predict air quality of four seasons. It can accelerate the training speed. Zheng *et al.* [8] used a temporal predictor to model the local factors of air quality, a spatial predictor to model global factors, and a dynamic aggregator to combine the predictions of spatial and temporal predictors. They attained 44.4% to 74.9% predict precision from a full day to one hour.

978-1-5090-6167-9/16 $31.00 © 2016 IEEE

DOI 10.1109/ICMLA.2016.100

Most existing methods try to solve two main problems of traditional ANN, i.e., slow convergence speed and ease of falling into local optima. They just reduce the possibility, while ELM can achieve unique global optima, and has a good convergence speed.

## III. Proposed Method

### A. Main Steps of STELM

As introduced above, the ELM algorithm provides a better learning strategy in the application of traditional data. However, the special properties of PM2.5 data make it necessary to extend the ELM into the field of geographical domain. Then, STELM is developed in this work with the incorporation of the special properties, e.g., spatial dependency and spatial heterogeneity.

The main steps of STELM are described in Figure 1. The input of the system are spatio-temporal sequences regarding air quality and meteorology data. Then we group spatio-temporal sequences into several clusters through clustering algorithm, e.g., GeoSOM (Geographic Self-Organizing Maps). For each cluster, a STELM model is built based on historical air quality, meteorological spatio-temporal sequences, and adjacent air quality data. Meteorological spatio-temporal sequences include temperature, humidity, wind direction, wind force, and precipitation. Since PM2.5 and PM10 have close relation with wind direction and strength, the air quality data in adjacent stations are considered as input features to highlight the influence of wind. Then we construct STELM models for each cluster, and predict future air quality based on the trained model.

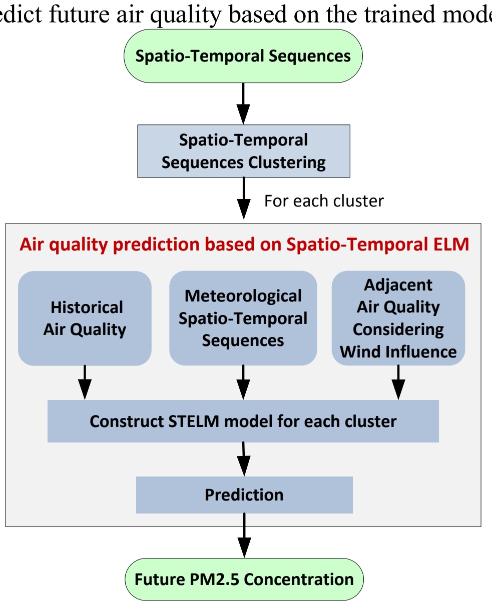


Figure 1. Main steps of STELM.

### B. Clustering

Considering the spatio-temporal properties of air quality data, we adopt clustering algorithm to group the sequences into different subareas according to their space locations. The similarity of sequences in one subarea should be different from that with other subareas as much as possible, at the same time the positions of sequences in one subarea should be adjacent in space.

In this paper, we select GeoSOM algorithm for spatio-temporal sequences clustering. GeoSOM adapts Self-Organizing Maps (SOM) to consider the spatial nature of geographic data [9]. SOM is usually used for mapping high-dimensional data into one, two or three-dimensional feature maps, which are grids of units or neurons. In GeoSOM, the search for the Best Matching Unit (BMU) is implemented by two phases, the first phase settles the geographical neighborhood where it is admissible to search for BMU, and the second phase performs the final search using the other multidimensional components. The search neighborhood is controlled by a parameter $k$, defined in the output space [10].

### C. STELM Model

As seen in Figure 2, there are two differences in the structure of STELM compared with the traditional network structure. 1) Spatio-temporal autocorrelation variable is incorporated into input layer, whose aim is to deal with spatial dependence. 2) The connecting weight vector is regarded as the function of spatial location, rather than independent of it, whose aim is to deal with spatial heterogeneity.

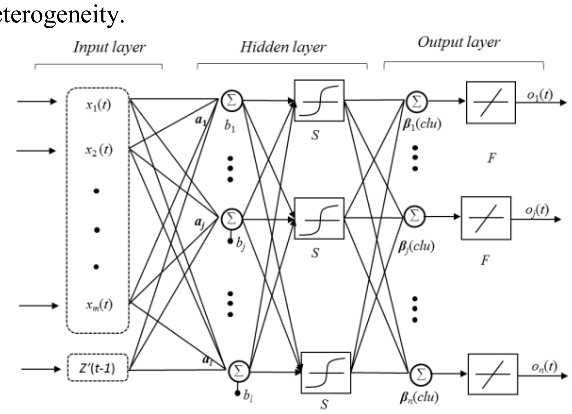


Figure 2. The structure of STELM

The input variables are composed of two parts: the covariates and autocorrelation variables. In Figure 2, $x_1(1)$, …, $x_m(t)$ represent the covariates, namely the meteorological elements, while $z'(t\text{-}1)$ refers to the autocorrelation variables, namely previous concentrations of PM2.5. In the field of spatial analysis, spatial autocorrelation variable is defined via spatial weights matrix $W$ ($N \times N$). Assume that $\boldsymbol{Z}(t)$ is the vector of PM2.5 concentration, then the spatio-temporal autocorrelation variable $\boldsymbol{Z'}(t\text{-}1)$ ($\boldsymbol{Z'}(t\text{-}1) = [\, z_1'(t\text{-}1), \ldots, z_n'(t\text{-}1)]$) can be defined as $W\boldsymbol{Z}(t)$. General methods of determining $W$ are based on spatial distance or spatial contiguity. Both of them make the isotropy assumption. However, since air pollutant always transport along with the direction of wind, the spread process of air pollution exists obviously anisotropic.

In the case of the central point $p_0$ in Figure 3, if the wind direction belongs to NE, the PM2.5 concentrations at $p_0$ may be affected by the $p_1$, $p_2$, $p_3$, and $p_4$. Obviously, the affect degree has negative correlation with the angle and the distance. The angle is defined through the wind direction and the edge between two points, e.g., the angel NE$p_0p_1$, and the distance is defined through the spatial location of two points, e.g., $d_{01}=\sqrt{(x_0-x_1)^2+(y_0-y_1)^2}$ (Euclidean distance).

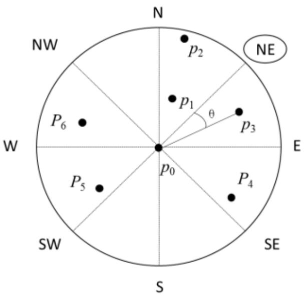


Figure 3. The influence of wind

In order to deal with anisotropy, Gauss vector weight (GVW) is defined based on traditional Gauss kernel function. GVW combines the direction effect and distance effect to accord with the transport process of air pollutants.

$$w_{ij}\left(d_{ij},\theta_{ij}(t)\middle|c,m\right)=\begin{cases} e^{-\frac{d_{ij}^2\bullet\sin^m\theta_{ij}(t)}{2c^2}} & if\ \ 0^\circ\le\theta_{ij}(t)\le 90^\circ \\ 0 & if\ \ 90^\circ<\theta_{ij}(t)\le 180^\circ \end{cases}$$

in which $d_{ij}$ and $\theta_{ij}$ represent the distance variable and the angle variable. The distance is calculated by spatial location directly, and the angle variable is estimated by a combination of dynamic wind direction. Two conditional parameters $m$ and $c$ are used to control the direction and distance effect. $m$ is regarded as anisotropic strength parameter. When $m$ equals 0, GVW is equal to traditional Gauss kernel function. $c$ is also called bandwidth, which can adjust the smoothing of GVW.

STELM builds region-model through clustering partition. Spatio-temporal clustering is a process of grouping spatio-temporal data into meaningful clusters according to its similarity in spatial and temporal domains [11]. The general geo-referenced time series clustering is used for clustering partition, then $\beta(loc)$ is estimated in each cluster. Because parameters $a$ and $b$ are given randomly, they are independent of spatial location. Hence, these parameters are identical in all local models in order to simplify the calculation.

Algorithm 1 presents the main steps of STELM.

**Algorithm 1**: *STELM* (*SQ, P*)

**Input**: Spatio-temporal sequences *SQ*; the location of each spatial position (monitoring station) *P*

**Output**: Prediction model *Mdl*.

**BEGIN**

Construct the space distance matrix $M$;

For each spatial position $i$, select a bandwidth $c$, and compute the Gauss vector weight matrix $W_{i,c}$

Assign randomly the input weight vectors $\boldsymbol{a}_i$ ($i = 1, \ldots, l$) and the threshold $b_i$ ($i = 1, \ldots, l$) of the feedforward neural network;

Calculate the hidden layer input vectors $H$;

Compute the output weight vectors

$\hat{\beta}=(H^T\bar{W}_{i,c}H)^{-1}H^T\bar{W}_{i,c}Z$ ;

Select the best matching model with minimum $CV$ as the final prediction model for position $i$, in which $CV=\sum_{t=1}^{k}\|z_i(t)-o_i(t)\|$

**Return** *model Mdl*

**END**

The space distance matrix M can be represented as:

$$M=\begin{bmatrix} d_{11} & \cdots & d_{1j} \\ \vdots & \ddots & \vdots \\ d_{i1} & \cdots & d_{ij} \end{bmatrix}$$

in which $d_{ij}$ is the Euclidean distance between position $i$ and $j$.

For each spatial position $i$, select a bandwidth $c$, the Gauss vector weight matrix $W_{i,c}$ can be represented as:

$$W_{i,c}=\begin{bmatrix} w_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{ij} \end{bmatrix}$$

in which $w_{ij}=\exp(-(d_{ij}/c))$, $c \in \{c_1, c_2, \cdots, c_m\}$, and the value of $c$ can be selected between minimum and maximum $d_{ij}$.

Suppose the length of each time series is $k$, we further define Hybrid spatial weight matrix as:

$\bar{W}_{i,c}=\mathrm{diag}(w_{i1},\ldots,w_{i1},w_{i2},\ldots,w_{i2},\ldots,w_{in},\ldots,w_{in})$

The numbers of $w_{i1}$, $w_{i2}$, …, $w_{in}$ are $k$, respectively.

To calculate the hidden layer input vectors $H$,

$$\boldsymbol{H}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_l,b_1,\ldots,b_l,\boldsymbol{x}_1(1),\ldots,\boldsymbol{x}_1(\mathrm{k}),\ldots,\boldsymbol{x}_n(1),\ldots,\boldsymbol{x}_n(\mathrm{k}))$$
$$=\begin{bmatrix} S(\boldsymbol{a}_1\cdot\boldsymbol{x}(1)+b_1) & \cdots & S(\boldsymbol{a}_l\cdot\boldsymbol{x}(1)+b_l) \\ \vdots & \cdots & \vdots \\ S(\boldsymbol{a}_1\cdot\boldsymbol{x}_n(\mathrm{k})+b_1) & \cdots & S(\boldsymbol{a}_l\cdot\boldsymbol{x}_n(\mathrm{k})+b_l) \end{bmatrix}_{nk\times l}$$

$$\boldsymbol{H}_{nk\times l}\boldsymbol{\beta}_l=\boldsymbol{o}_{nk}$$

To compute the output weight vectors, the minimum cost function is calculated:

$$E_{i,c}=\sum_{t=1}^{k}\sum_{j=1}^{n}w_{ij}\|z_j(t)-o_j(t)\|$$

$$\hat{\beta}=(H^T\bar{W}_{i,c}H)^{-1}H^T\bar{W}_{i,c}Z$$

Finally, we compare the models with different bandwidths $c$, and select $\min(CV=\sum_{t=1}^{k}\|z_i(t)-o_i(t)\|)$ as the final prediction model for position $i$.

## IV. Implementation and Experiment

To evaluate the performance of our method, we use the air quality and meteorology data in Beijing during Apr. and May in 2014, to predict the PM2.5 in later 3 days, i.e., May 29-31, 2014. There are 35 air quality monitoring station in Beijing. Each station records the concentrations of 6 main pollutants ($NO_2$, CO, $SO_2$, $O_3$, PM10, and PM2.5). As far as meteorology data, temperature, humidity, wind direction, wind force, and precipitation are collected.

After pre-processing like interpolation of missing value and elimination of abnormal value, all the spatio-temporal sequences are input to our system. First, GeoSOM is applied to cluster the PM2.5 sequences of 35 stations. After analysis, we found that when the number of clusters is 14, the maximum Sil index and minimum DB index could be achieved. Accordingly, the 35 stations are grouped into 14 clusters. For each cluster, we build a STELM model, based on which the values of PM2.5 in future 72 hours are predicted.

Taking the Olympic Sports Center Station as an example, the prediction results in future 24 hours are shown in Figure 4. Compared with proposed STELM method, the results of ELM and MLR (Multiple Linear Regression) are also listed as baselines. We can see that on May 29, 2014, the real value of PM2.5 varied greatly. The MLR method can only predict the coarse tendency, while ELM and STELM have much better prediction performance. Moreover, STELM outperforms ELM slightly in first 15 hours and greatly after 15 hours, especially when PM2.5 has sudden changes in the $17^{th}$ and $22^{nd}$ hour.

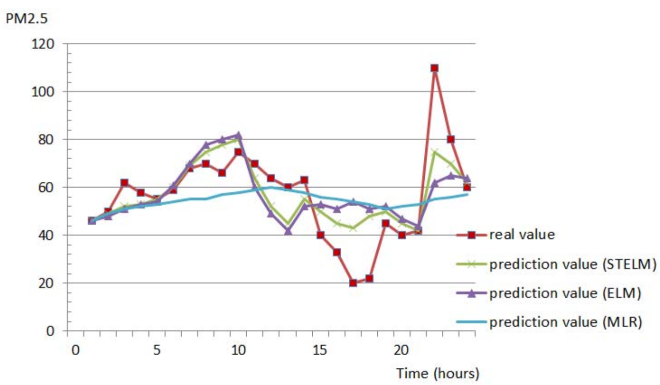


Figure 4. Prediction results (Olympic Sports Center Station).

The Mean Relative Error (MRE) for 35 monitoring stations in future 12 hours is shown in Figure 5. It can be seen that the MRE in first 12 hours is relatively low. The Mean Absolute Error (MAE) for 35 monitoring stations in future 24 hours increases as time passes. The calculation equations of MRE and MAE are listed below:

$$MRE_t = \frac{1}{35}\sum_{i=1}^{35}\left|\frac{Prediction(i)_t - Observation(i)_t}{Observation(i)_t}\right|$$

$$MAE_t = \frac{1}{35}\sum_{i=1}^{35}\left|Prediction(i)_t - Observation(i)_t\right|$$

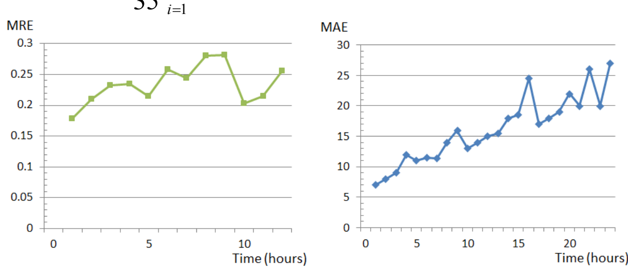


Figure 5. Average relative error for 12h prediction (left). Mean absolute error for 24h prediction (right).

To validate the prediction precision, we believe the predicted value is correct if the predicted value and real value are in the same level. As reported by China's State Environment Protection Agency, the air pollution level is as follows: 0-50, Excellent; 51-100, Good; 101-150, Slightly Polluted; 151-200, Lightly Polluted; 201-250, Moderately Polluted; 251-300, Heavily Polluted; 300+, Severely Polluted. The overall precision in first 12, 24, 48, 72 hours are 82%, 78%, 71%, and 63%, respectively.

## V. Conclusions and Future Work

This paper proposes an algorithm for air quality prediction called STELM for short. Combining the advantages of spatio-temporal ANN and extreme learning machine, it is suitable to spatio-temporal sequences modeling and prediction, with the properties of fast training, less configuration parameters, and high precision. STELM considers the spatial heterogeneity and the influence of wind, and achieves effective generalization performance and higher precision. The real-world application on PM2.5 prediction in Beijing demonstrates the usefulness and effectiveness of STELM. In the future, we will improve the precision and reduce the absolute errors, and merge more data sources to further increase the precision.

## Acknowledgment

This work is supported by Beijing Natural Science Foundation (4152007), China National Key Technology Research and Development Program project with No. 2013BAH05F02, Beijing Postdoctoral Research Foundation, and China Postdoctoral Science Foundation.

## References

[1] G. B. Huang, Q.Y. Zhu, C. K. Siew. Extreme learning machine: theory and applications. Neurocomputing, 2006, 70(1), pp.489-501.

[2] B. Lal, S.S. Tripathy. Prediction of dust concentration in open cast coal mine using artificial neural network. Atmospheric Pollution Research, 2012, 3(2), pp.211-218.

[3] F. Nejadkoorki, S. Baroutian. Forecasting Extreme PM10 Concentrations Using Artificial Neural Networks. International Journal of Environmental Research, 2012, 6(1), pp.277-284.

[4] A. Russo, F. Raischel, P.G. Lind. Air quality prediction using optimal neural networks with stochastic variables. Atmospheric Environment, 2013, 79(7), pp.822-830.

[5] Z. Jiang, et al. An air quality forecast model based on the BP neural network of the samples self-organization clustering. International Conference on Natural Computation. IEEE, 2010, pp. 1523-1527.

[6] A. Samia, et al. A Hybrid ARIMA and Artificial Neural Networks Model to Forecast Air Quality in Urban Areas: Case of Tunisia. Advanced Materials Research, 2012, Vols. 518-523, pp. 2969-2979.

[7] A. Kumar, P. Goyal. Forecasting of Air Quality Index in Delhi Using Neural Network Based on Principal Component Analysis. Pure & Applied Geophysics, 2013, 170(4), pp. 711-722.

[8] Y. Zheng, et al. Forecasting Fine-Grained Air Quality Based on Big Data. Proceeding of the 21th SIGKDD conference on Knowledge Discovery and Data Mining, KDD2015, pp. 2267-2276.

[9] GeoSOM Home - Universidade Nova de Lisboa. http://www.novaims.unl.pt/labnt/geosom/. .

[10] R. Henriques, et al. Spatial Clustering with SOM and GeoSOM: Case Study of Lisbon's Metropolitan Area. IEEE International Conference on Advanced Geographic Information Systems, 2010, pp.148-152.

[11] M. Deng, Q.L. Liu, J.Q. Wang, Y. Shi. A general method of spatio-temporal clustering analysis. Science China: Information Science, 2011, 54(10), pp.1-14.