Applications of SVR-PSO Model and Multivariate Linear Regression Model in PM2.5 Concentration Forecasting

Guo-Feng Fan, School of Mathematics & Statistics Science, Ping Ding Shan University, Ping Ding, China
Meng-Qi Liang, School of Mathematics & Statistics Science, Ping Ding Shan University, Ping Ding, China
Jing-Ru Li, School of Economics and Management, Ping Ding Shan University, Ping Ding, China
Wen-Lu Ma, School of Economics and Management, Ping Ding Shan University, Ping Ding, China

ABSTRACT

At present, the fog and haze problem is intensified, which has a great impact on the production of enterprises and living of the residents. PM2.5 is an important indicator of air pollution and it also receives much concern. This article collects the reliable data of PM2.5 in the five industrial cities in Henan Province from Weather Report Network, and PM2.5 Data Network since 2015. The effective approaches to forecast PM2.5 concentration is proposed, i.e., the improved multivariate linear regression (namely IMLR) model and support vector regression with particle swarm optimization algorithm (namely SVR-PSO) model. The empirical results demonstrate that the proposed IMLR and SVR-PSO forecasting models are effective, and also, could be an instructive reference for weather quality forecasting, safe travel, and safe production.

KEYWORDS

Multivariate Linear Regression (MLR), Particle Swarm Optimization (PSO), PM2.5 Concentration Forecasting, Support Vector Regression (SVR)

1. INTRODUCTION

In recent years, along with the economic development and the increasing request of high quality life, the degree of industrialization has also been promoted to a high level, the basic infrastructures of cities are also gradually improving. Henan is an industrial province with a large amount of population in China. The trade-off issue of balance developments from economic and environmental sides is more and more important (Zhang et al., 2015a). Environmental pollution problem is prominent, the range of fog and weather is expanding. The automobile exhaust and industrial pollution are the primary pollution sources of PM2.5, which result the declining of the air quality. Therefore, accurate air quality forecasting is becoming an urgent air quality governance approach. There are three kinds of methodology for PM2.5 concentration forecasting. The first one is based on the pollution caused mechanism to establish the regression relationships among PM2.5 concentration and other relevant factors. In which, PM2.5 concentration is viewed as the dependent variable, and the other relevant factors are viewed as the independent variables. The disadvantage of this method is, on the one hand, that it is difficult to determine suitable independent variables both with high correlation and reasonable social-economic relationships. On the other hand, the data of the relevant factors are difficult to acquire. The second method is the so-called kinetic method, which is based on PM2.5’s

DOI: 10.4018/IJAEC.2017100105

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
chemistry and diffusion, i.e., establishing the PM2.5 diffusion kinetics model to forecast PM2.5 concentration. This method could theoretically improve the accuracy of the PM 2.5 concentration forecasting; however, it is difficult to collect the diffusion kinetic values in actual collection process. The third method is based on the data structure to establish the univariate forecasting model, such as ARIMA model. The advantage of this method is that it does not need to consider the complicate relationships among the relevant factors. The disadvantage is that this kind of forecasting models theoretically require higher data regularity and stability, and the forecasting performance would not be satisfied while suffering from the mutation effect of PM 2.5 concentration caused by holidays and meteorological factors (Mao et al., 2017). Comprehensively considering the limitations of the abovementioned three categories of PM2.5 concentration forecasting methodologies, this paper applies the nonlinear selection of multiple factors linear regression (Chiou et al., 2017) and SVR-PSO (Jia et al., 2017; Li et al., 2013; Li & Li, 2016) method to propose the daily average PM2.5 concentration forecasting models, respectively, to forecast the PM2.5 concentration from 2015 to 2017 years. The forecasting results also compared with each other.

The remain of this paper is organized as follows. Section 2 introduces the collected data sets and relevant analysis methodologies. Section 3 introduces the forecasting models. Section 4 demonstrates the experimental results. Section 5 concludes this paper.

2. THE DATA SOURCES AND PROCESSING

In this paper, the data are collected from the Weather Report and the PM2.5 Data Network, in which five cities of Henan Province, China, are selected. These five selected cities include Zheng Zhou City (the provincial capital of Henan Province), An Yang City, Luo Yang City, Ping Ding Shan City, and Jiao Zuo City; in which, the previous two cities are the industrial cities and the other three cities are belonged to the resource-based cities (Zhang et al., 2015b). The data contents mainly contain PM2.5 concentration and the dependent variable AQI index. The period of the collected data is from 23 April 2015 to 31 July 2015; 24 May 2016 to 31 August 2016; and 23 June 2017 to 30 September 2017; totally 300 days in the collected data set. In which, 200 days are used as the training set to establish the training model, the other 100 days are used to establish the testing model (Yang et al., 2013). In the training stage, the parameters of an SVR model would be determined with minimal forecasting errors, and the un-used raw data will be employed to calculate the forecasting errors. The testing results has demonstrated that the proposed approach receives more accurate forecasting performances, which could provide valuable and meaningful governance guidance than the simple linear regression model (Lv et al., 2016).

3. ANALYSIS METHOD

3.1. Multivariate Linear Regression Model

The multivariate linear regression (MLR) model is composed of the dependent variable and two or more explanatory (independent) variables, shown as Equation (1)

\[ Y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \varepsilon \]  

(1)

In general, a phenomena is often associated with a number of factors, the MLR model is applied to find out the optimal combination of these factors (independent variables) to forecast the dependent variable. The forecasting results of a MLR model are often superior to the univariate regression model. This paper firstly employs the MLR model to forecast PM2.5 concentration in each city, then, authors would be based on the shortcomings of the results to provide some improving approach.
Recurrence Indicators for the Estimation of Characteristic Size and Frequency of Spatial Patterns
www.igi-global.com/chapter/recurrence-indicators-estimation-characteristic-size/69463?camid=4v1a

Complex Adaptive Systems Theory and Military Transformation
www.igi-global.com/chapter/complex-adaptive-systems-theory-military/5142?camid=4v1a