Environmental Pollution 231 (2017) 997–1004

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation☆

Xiang Li [a, b], Ling Peng [a], Xiaojing Yao [a, *], Shaolong Cui [a], Yuan Hu [a, b], Chengzeng You [a, b], Tianhe Chi [a]

[a] Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 24 February 2017
Received in revised form
29 August 2017
Accepted 31 August 2017

Keywords:
Air pollutant concentration predictions
Long short-term memory neural network (LSTM NN)
Recurrent neural network
Spatiotemporal correlation
Multiscale prediction

ABSTRACT

Air pollutant concentration forecasting is an effective method of protecting public health by providing an early warning against harmful air pollutants. However, existing methods of air pollutant concentration prediction fail to effectively model long-term dependencies, and most neglect spatial correlations. In this paper, a novel long short-term memory neural network extended (LSTME) model that inherently considers spatiotemporal correlations is proposed for air pollutant concentration prediction. Long short-term memory (LSTM) layers were used to automatically extract inherent useful features from historical air pollutant data, and auxiliary data, including meteorological data and time stamp data, were merged into the proposed model to enhance the performance. Hourly $PM_{2.5}$ (particulate matter with an aerodynamic diameter less than or equal to 2.5 μm) concentration data collected at 12 air quality monitoring stations in Beijing City from Jan/01/2014 to May/28/2016 were used to validate the effectiveness of the proposed LSTME model. Experiments were performed using the spatiotemporal deep learning (STDL) model, the time delay neural network (TDNN) model, the autoregressive moving average (ARMA) model, the support vector regression (SVR) model, and the traditional LSTM NN model, and a comparison of the results demonstrated that the LSTME model is superior to the other statistics-based models. Additionally, the use of auxiliary data improved model performance. For the one-hour prediction tasks, the proposed model performed well and exhibited a mean absolute percentage error (MAPE) of 11.93%. In addition, we conducted multiscale predictions over different time spans and achieved satisfactory performance, even for 13–24 h prediction tasks (MAPE = 31.47%).

1. Introduction

Air pollution is a serious environmental problem that has attracted increasing attention worldwide (Kurt and Oktay, 2010). Certain air pollutants, such as $PM_{2.5}$ (particulate matter with an aerodynamic diameter less than or equal to 2.5 μm), can traverse the nasal passages during inhalation and reach the throat and even the lungs. Long-term exposure to ambient fine particulate matter can negatively affect human health (Dockery et al., 1993; Iii et al., 2002; Krewski et al., 2009) and cause respiratory and cardiovascular diseases and some other illnesses (Kappos et al., 2004; Neuberger et al., 2004; Wilson et al., 2005; Bravo and Bell, 2011). Therefore, obtaining air pollutant concentration information in real time is significant for air pollution control and the prevention of health issues due to air pollution (Zheng et al., 2013).

In recent years, many research efforts have focused on enriching approaches to predicting air pollutant concentrations. In general, methods of predicting air pollutant concentrations fall into two major categories: deterministic and statistical methods.

Deterministic methods adopt meteorological principles and statistical methods to model the emission, dispersion, transformation, diffusion and removal processes of pollutants based on atmospheric physics and chemical reactions; thus, the spatiotemporal distributions of air pollutants are simulated at different scales and orientations (Bruckman, 1993; Coats, 1996; Lurmann, 2000; Guocai, 2004; Baklanov et al., 2008; Kim et al., 2010; Jeong et al., 2011). These methods are viewed as model-based methods because their structures are predefined based on certain theoretical

☆ This paper has been recommended for acceptance by Dr. Hageman Kimberly Jill.
* Corresponding author.
E-mail address: yaoxj@radi.ac.cn (X. Yao).

http://dx.doi.org/10.1016/j.envpol.2017.08.114
0269-7491/

hypotheses, and the parameters can be calculated via specific priori knowledge. Many air quality models have been developed to simulate the complicated process of air pollutant diffusion. Representative methods, such as the Community Multiscale Air Quality (CMAQ) model (Chen et al., 2014), Nested Air Quality Prediction Modeling System (NAQPMS) (Wang et al., 2001) and WRF-Chem model (Saide et al., 2011), are commonly adopted for air pollutant concentration forecasting in urban areas. Although developed theories provide valuable insights for understanding pollutant diffusion mechanisms, most of these theoretical models are relevant to sophisticated priori knowledge, unreliable and limited data, and various usage constraints (Vautard et al., 2007; Stern et al., 2008).

Statistical methods, however, avoid sophisticated theoretical models and simply apply statistics-based models to predict air quality. Widely used methods include the multiple linear regression (MLR) method (Li et al., 2011), the autoregressive moving average (ARMA) method (Box and Jenkins, 1976), the support vector regression (SVR) method (Nieto et al., 2013), the artificial neural network (ANN) method (Hooyberghs et al., 2005), and hybrid methods (Díaz-Robles et al., 2008; Chen et al., 2013). Among these models, the ANN method, which can perform nonlinear mapping and is self-adaptive and robust, generally provides satisfying performance; therefore, it has been widely used in time series forecasting fields (Yoon et al., 2011). Recently, various ANN structures have been developed to improve predictions of air pollutant concentrations. Typical examples include the widely employed multilayer perceptron (MLP) (Paschalidou et al., 2011), back propagation neural network (BPNN) (Kolehmainen et al., 2001), radial basis function neural network (RBF NN) (Lu et al., 2002), neuro-fuzzy neural network (NFNN) (Mishra and Goyal, 2016), general regression neural network (GRNN) (Antanasijević et al., 2013), and recurrent neural network (RNN) (Feng et al., 2011). Due to the dynamic nature of relevant atmospheric environments, RNNs are particularly suited to capturing the spatiotemporal evolution of air pollutant distributions because RNNs can handle arbitrary sequences of inputs, thereby guaranteeing the capacity to learn temporal sequences (Ma et al., 2015). Certain RNNs, such as the time delay neural network (TDNN) (Ong et al., 2016) and Elman neural network (Prakash et al., 2011), have been used for air pollutant prediction in previous studies. However, these RNN models face two issues: 1) in the RNN structure, the time lag must be determined in advance, which requires a considerable number of experiments to identify the optimum time lag; and 2) traditional RNNs fail to capture long time dependencies in input sequences, and training RNNs with long time lags is difficult because vanishing gradient and exploding gradient problems may be encountered (Hochreiter and Schmidhuber 1997).

To resolve these issues, a special RNN architecture referred to as a long short-term memory neural network (LSTM NN) was developed by Hochreiter and Schmidhuber (1997). Unlike traditional RNNs, LSTM NNs are capable of learning long time series and are not affected by the vanishing gradient problem. These features are especially important for modeling spatiotemporal air pollutant processes in which the air pollutant concentration of one station is related to the previous status and those at nearby stations because of pollutant transport processes.

In recent years, the LSTM NN has been successfully applied to many studies involving time series prediction, such as traffic flow prediction (Lv et al., 2015), wind power prediction (Felder et al., 2010), human trajectory prediction (Alahi et al., 2016), etc. Recently, Sak et al. (2016) (Sak et al., 2016) adopted the LSTM NN for pollution risk prediction, but they only classified the pollution risk ranking without conducting real-value predictions of air pollutant concentrations. Moreover, they made predictions separately for individual cities without considering the spatial correlations between monitoring stations. To the best of our knowledge, the LSTM NN has not been applied in the domain of air pollutant concentration prediction. This paper aims to extend the LSTM NN to spatiotemporal correlation modeling and air pollutant concentration prediction.

The contributions of this paper are as follows: (1) an LSTM NN is extended to capture the long-term spatiotemporal dependency of air pollutant concentrations, and a multiscale prediction framework which can forecast the air pollutant concentration over the next 24 h is presented; (2) the proposed method can effectively and automatically extract the spatiotemporal correlations within air pollutant concentration data; and (3) auxiliary data are integrated into a traditional LSTM NN model, and the integrated model exhibits better performance than traditional methods.

2. Data and methods

2.1. Data description

Hourly $PM_{2.5}$ concentration data from 12 air quality monitoring stations in downtown Beijing collected from Jan/01/2014 to May/28/2016 were obtained from the Ministry of Environmental Protection of China (http://datacenter.mep.gov.cn/). Concentrations were measured using a Thermo Fisher 1405F detector and calculated based on the tapered element oscillating microbalance (TEOM) method. Meteorological data from the same period were downloaded from The National Oceanic and Atmospheric Administration's (NOAA's) national climate data center (https://www.climate.gov/). In successive experiments, we chose four main factors from the meteorological dataset that are highly related to $PM_{2.5}$ concentrations: temperature, humidity, wind speed and visibility (Díaz-Robles et al., 2008; Saide et al., 2011; Guocai, 2004). Fig. 1 shows the distribution of the air quality monitoring stations (blue triangles) and the location of the meteorological station (green triangle). Simple linear interpolation was performed to fill in the missing values in both datasets. Our dataset contained 20196 records for each station. In our experiment, we randomly selected 80

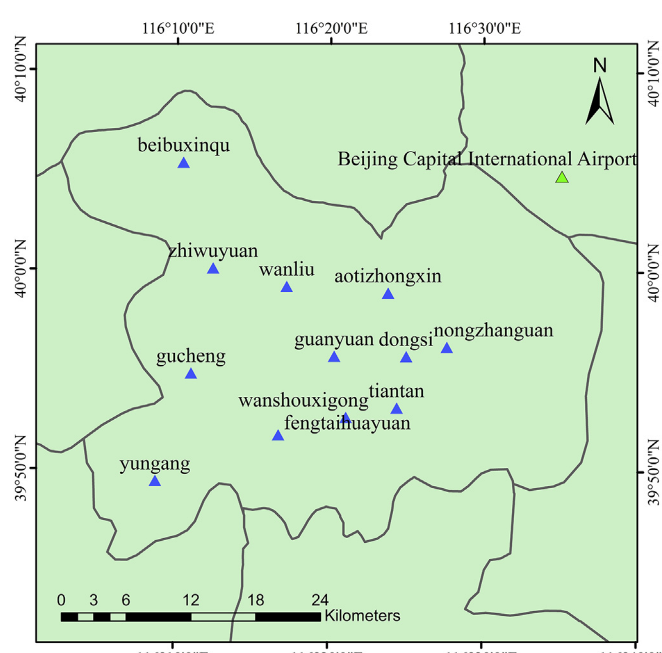

Fig. 1. Distribution of air quality monitoring stations in Beijing City.