Science of the Total Environment 626 (2018) 1421–1438

Contents lists available at ScienceDirect

Science of the Total Environment

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

Research and application of a novel hybrid air quality early-warning system: A case study in China

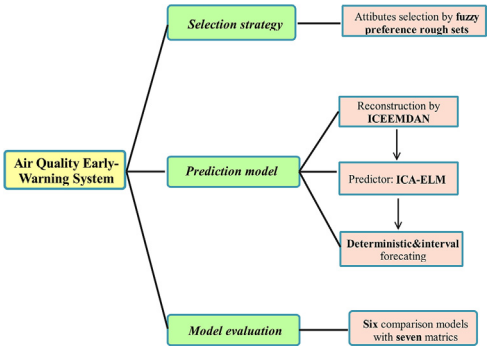
Chen Li, Zhijie Zhu *

School of Statistics, Dongbei University of Finance and Economics, No. 217, Jianshan Road, Shahekou District, Dalian, Liaoning Province 116025, China

HIGHLIGHTS

- A hybrid early-warning system is developed for air quality monitoring and analysis.
- The fuzzy preference rough sets theory is applied to select the attributes for cities.
- A hybrid model named ICEEMDAN-ICA-ELM is proposed to forecast the main pollution contaminants.
- Deterministic and interval prediction are conducted and validated well in six cities in China.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 17 October 2017
Received in revised form 10 January 2018
Accepted 19 January 2018
Available online xxxx

Editor: P. Kassomenos

Keywords:
Air quality early-warning system
Attributes selection
Pollution contaminants prediction
Hybrid model
Interval forecasting

ABSTRACT

As one of the most serious meteorological disasters in modern society, air pollution has received extensive attention from both citizens and decision-makers. With the complexity of pollution components and the uncertainty of prediction, it is both critical and challenging to construct an effective and practical early-warning system. In this paper, a novel hybrid air quality early-warning system for pollution contaminant monitoring and analysis was proposed. To improve the efficiency of the system, an advanced attribute selection method based on fuzzy evaluation and rough set theory was developed to select the main pollution contaminants for cities. Moreover, a hybrid model composed of the theory of "decomposition and ensemble", an extreme learning machine and an advanced heuristic algorithm was developed for pollution contaminant prediction; it provides deterministic and interval forecasting for tackling the uncertainty of future air quality. Daily pollution contaminants of six major cities in China were selected as a dataset to evaluate the practicality and effectiveness of the developed air quality early-warning system. The superior experimental performance determined by the values of several error indexes illustrated that the proposed early-warning system was of great effectiveness and efficiency.

1. Introduction

In this section, the research background, literature review and the aims and innovation of this paper are shown in detail.

* Corresponding author.
E-mail address: dufezhijie@163.com (Z. Zhu).

https://doi.org/10.1016/j.scitotenv.2018.01.195
0048-9697/

1.1. Background

Environmental pollution has been one of the most serious scourges because of the great damage it causes to economies and the lives of people. According to the World Health Organization (WHO) report from March 6, 2017, air pollution takes the lives of 1.7 million children under 5 years of age every year (http://www.who.int/mediacentre/news/releases/2017/pollution-child-death/en/). With industrialization and urbanization, air pollution and hazy weather have increased at a great rate, especially in developing countries (Wu and Zhang, 2017). In recent years, China has experienced severe and persistent air pollution in the winter, which has attracted worldwide attention (Liu et al., 2017). Relevant research about this issue has also flourished (Ma et al., 2017; Zhuo et al., 2017).

The dominant pollution contaminants, such as particulate matter (PM), sulfur dioxide (SO_2) and nitrogen oxides (NO_x), pose significant risks to environments. These pollution contaminants are barometers of reaction to environmental problems (L. Zhang et al., 2017). In most air quality monitoring systems, carbon monoxide (CO) and ozone (O_3) are the two main objective pollution contaminants (Maga et al., 2017). Additionally, other hazardous substances, such as radiation, soiling or specific toxic gases, cause great damage to air quality. In fact, finding the main factor that causes air pollution is the key to solving the problem. Moreover, precise prediction of pollution contaminants plays a vital role. Therefore, it is of great importance to propose an early-warning system and take effective corresponding protection measures.

1.2. Literature review

There are a series of environmental factors that cause air pollution, and the flow and diffusion of pollution contaminants is a rather complex process (Vidale et al., 2017). In early stages, many scholars focused on the relationship between air pollution and its environmental influence factors (Pahlavani et al., 2017). Based on these pioneering works, air quality monitoring systems emerged for monitoring the main pollution contaminants. Recently, studies on the emission sources of pollution contaminants have become a hot topic, and early-warning systems have shown practical importance (Yin et al., 2015; Liora et al., 2016; L. Li et al., 2017). However, little research has focused on the selection of pollutant indicators and the design of early-warning systems for particular cities.

On the other hand, a variety of models have been proposed to forecast pollution contaminants that can be roughly divided into two categories: statistical models and machine learning models. Traditional statistical approaches, including regression models (Lee et al., 2017), time series models (Nhung et al., 2017) and autoregressive moving average models (ARIMA) (Zafra et al., 2017), are applied in pollution contaminant prediction. Additionally, researchers have investigated the partial and multiple linear features of pollution contaminants and have shown different patterns of statistical models (Carlsen et al., 2018; Zhu et al., 2017). However, each of these approaches has difficulties in dealing with non-linear problems.

In addition to statistical models, machine learning models have been widely used in pollution contaminant forecasting. For example, support learning machines (SVMs) (Zhu et al., 2012), artificial neural networks (ANNs) and fuzzy logic have been applied to overcome non-linear limitations and uncertainties to achieve better forecasting accuracy (Elangasinghe et al., 2014). Generally, ANNs are considered as a rather better choice in tackling non-linear problems (Perez and Reyes, 2006). For example, Niu et al. established an efficient early-warning system based on a least squares support vector machine for $PM_{2.5}$ concentration prediction (Niu et al., 2017). Wang et al. indicated that a proposed model based on back propagation neural networks had better performance forecasting pollution contaminants than other models (i.e., ARIMA, SVM) in the case of three cities in China (J. Wang et al., 2017). However, ANN models may also have many disadvantages, such as the over-fitting problem and easily falling into the local optimum problem.

To improve the efficiency and accuracy of prediction performance, data pre-processing approaches and optimization methods have been widely investigated in pollution contaminants forecasting. Research on data pre-processing methods, such as empirical mode decomposition (EMD) family approaches (Zhou et al., 2014; D. Wang et al., 2017) and singular spectrum analysis (SSA) (Kumar and Jain, 2010), have flourished for years. For example, Niu et al. developed a hybrid model based on CEEMD for $PM_{2.5}$ concentration forecasting (Niu et al., 2016). On the other hand, intelligent optimization algorithms such as immune algorithms (IA) (Lin et al., 2011), genetic algorithms (GA) (Ruiz et al., 2018) and the grey wolf optimizer (GWO) algorithm (Niu et al., 2016) have been employed to optimize the parameters of ANNs to improve prediction performance.

Moreover, a variety of early-warning systems have been presented and applied for monitoring and analyzing air quality. Xu et al. proposed an air quality early-warning system that consists of prediction and evaluation (Xu et al., 2017a). An interoperable online air quality information system based on spatial-temporal distribution was provided by Wiemann et al. (Wiemann et al., 2016). Recent early-warning systems always focus on deterministic prediction combined with uncertainty prediction (Xu et al., 2017b). Yang et al. proposed an early-warning system that combined air quality prediction with assessment modules (Yang and Wang, 2017). Unfortunately, relevant studies are far from sufficient in China, despite their great significance. Accordingly, establishing a scientific and useful air quality early-warning system has become highly urgent for decision-makers.

1.3. Aims and innovations

Based on the above information on air quality monitoring and analysis, the shortcomings and deficiencies of the models mentioned in the literature review are as follows:

- Attributes selection is ignored, and the proper indicators are difficult to determine;
- The accuracy and stability of individual forecasting models can hardly meet the requirements;
- Uncertainty prediction is rarely applied in early-warning systems, and the current early-warning systems fail to provide specific strategies for different environments.

In this paper, a novel hybrid early-warning system for air quality monitoring is proposed that is composed of three parts: attributes selection, deterministic prediction and uncertainty prediction. With respect to the first part, a novel attribute selection method based on fuzzy preference relations and rough set theory was used to evaluate the relevance between pollution contaminants and air quality. The corresponding main pollution contaminants of each study city were decided by attribute dependency. For the deterministic prediction part, a novel hybrid model consisting of signal processing, an extreme learning machine (ELM) and a recent heuristic optimization algorithm called the imperialist competitive algorithm (ICA) was used to forecast the pollution contaminants, which is the core component of the proposed early-warning system. The corresponding details of the novel deterministic prediction model are as follows: An advanced de-noising method (ICEEMDAN) was used to reconstruct the original series of each pollution concentration. Then, the ELM optimized by a heuristic algorithm (ICA) was applied to obtain the final forecasting result. For the uncertainty prediction, an interval forecasting model based on the proposed deterministic forecasting model was developed to estimate the possible boundary of deterministic estimates. The combination of the point forecasting and the interval forecasting makes the early-warning system more credible and reliable. To assess the overall quality of the proposed