Information Systems ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Information Systems

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

Forecasting smog-related health hazard based on social media and physical sensor

Jiaoyan Chen [a], Huajun Chen [a,*], Zhaohui Wu [a], Daning Hu [b], Jeff Z. Pan [c]

[a] College of Computer Science, Zhejiang University, Hangzhou, China
[b] Department of Informatics, University of Zurich, Zurich, Switzerland
[c] Department of Computer Science, The University of Aberdeen, Aberdeen, UK

ARTICLE INFO

Keywords:
Smog disaster
Health hazard
Social media
Urban data
Forecasting
Data mining

ABSTRACT

Smog disasters are becoming more and more frequent and may cause severe consequences on the environment and public health, especially in urban areas. Social media as a real-time urban data source has become an increasingly effective channel to observe people's reactions on smog-related health hazard. It can be used to capture possible smog-related public health disasters in its early stage. We then propose a predictive analytic approach that utilizes both social media and physical sensor data to forecast the next day smog-related health hazard. First, we model smog-related health hazards and smog severity through mining raw microblogging text and network information diffusion data. Second, we developed an artificial neural network (ANN)-based model to forecast smog-related health hazard with the current health hazard and smog severity observations. We evaluate the performance of the approach with other alternative machine learning methods. To the best of our knowledge, we are the first to integrate social media and physical sensor data for smog-related health hazard forecasting. The empirical findings can help researchers to better understand the non-linear relationships between the current smog observations and the next day health hazard. In addition, this forecasting approach can provide decision support for smog-related health hazard management through functions like early warning.

© 2016 Published by Elsevier Ltd.

1. Introduction

Smog disasters are becoming more and more frequent and may cause severe consequences on the environment and public health in China. For example, in January 2013, smog had covered the capital of China, Beijing, for over 20 days. According to recent statistics [1], smog affects more than a quarter of the land and over 600 million people in China.

According to Virginia Hughes [2], smog is a health hazard that may adversely affect people's health. Sometimes it causes extreme and immediate public health emergency, like the one in 1952 London [3]. Therefore, it is necessary to develop a systematic approach to analyze, monitor and forecast smog-related health hazards in a timely manner.

On the other hand, social media as a real-time urban data source has become an increasingly important channel to observe events, trends and sentiment [4,5]. Negative comments on smog or complaints about smog-related health conditions from a small group of environment sensitive individuals can diffuse really fast on social media and cause much large scale of discussions and reactions. Therefore, social media with its network effects can be used to capture possible smog-related public health disasters in its early stage and provide warnings.

In the big data era, various technologies are developed to extract, process and analyze population-level social media

* Corresponding author.
E-mail address: huajunsir@zju.edu.cn (H. Chen).

http://dx.doi.org/10.1016/j.is.2016.03.011
0306-4379/© 2016 Published by Elsevier Ltd.

data, but few with the purpose of forecasting. Previous research [6] usually collected and analyzed such social media data for monitoring the impacts of nature environment on public health, but there is a lack of systematic approaches for forecasting smog-related health hazards with social media data.

Moreover, a variety of physical sensor platforms for monitoring smog status, including air quality stations, weather stations and earth observation satellites, are also widely deployed across China for both big cities and small towns [7], generating a huge amount of observational data about smog severity.

As Fig. 1 shows, we propose a predictive analytics approach that utilizes both social media and physical sensors for smog-related health hazard forecasting. It contains two major components: (1) modeling smog-related health hazards and smog severity with raw microblogging text and network information diffusion records and (2) forecasting the next day smog-related health hazards using an artificial neural network-based model.

To the best of our knowledge, our research is the first study to systematically model and analyze real-world social media and physical sensor data for smog-related health hazard forecasting. Firstly, this study can help researchers to better understand the non-linear relationships between current smog observations and the next day health hazard, in which physical sensors alone often fail to capture. Secondly, the proposed predictive analytics framework aims to provide decision support for smog-related health hazard management through functions like early warning for the coming smog-related public health emergency.

Moreover, we investigate the strengths of social media in smog-related health hazard forecasting. It can contribute more than physical sensors in forecasting the smog-related health hazards when the smog disasters are severe. Meanwhile, data about social observations' diffusion in social networks can further improve the forecasting accuracy.

2. Related work

2.1. Smog disaster and public health

On one hand, predictive analytics that are related to smog disasters or other kinds of air pollutions usually investigates the natural observations themselves without considering their related health hazard. Here are some examples. Merz et al. [8] conducted a time-series analysis of air monitoring data for the downtown Los Angeles station to detect the air pollution trends. Casado et al. [9] applied a series of geostatistics and visualization procedures to analyze hourly ozone measurements collected from 29 stations in the southeastern United States, which clearly confirmed the diurnal pattern of ozone fluctuations. Van et al. [10] investigated smog prediction problem in perspective of computational steering techniques which allow an optimal trade off between computation speed and prediction accuracy.

On the other hand, most studies that involve smog-related public health problems usually analyzed the impacts of smog on public health, but largely ignored real-time health hazard monitoring and forecasting. They mainly used objective indicators from physical sensors or statistics from hospitals. Pope and Dockery [11] conducted an extensive review on the research about health effects of particulate matter (PM) – the most harmful component in smog. They focused on the short-term and long-term PM exposure and its effects on mortality and some diseases. Recently, Hughes et al. [2] compared annual case numbers of chronic obstructive pulmonary disease (COPD) with smog trends in some cities to investigate the health effects of smog in past years.

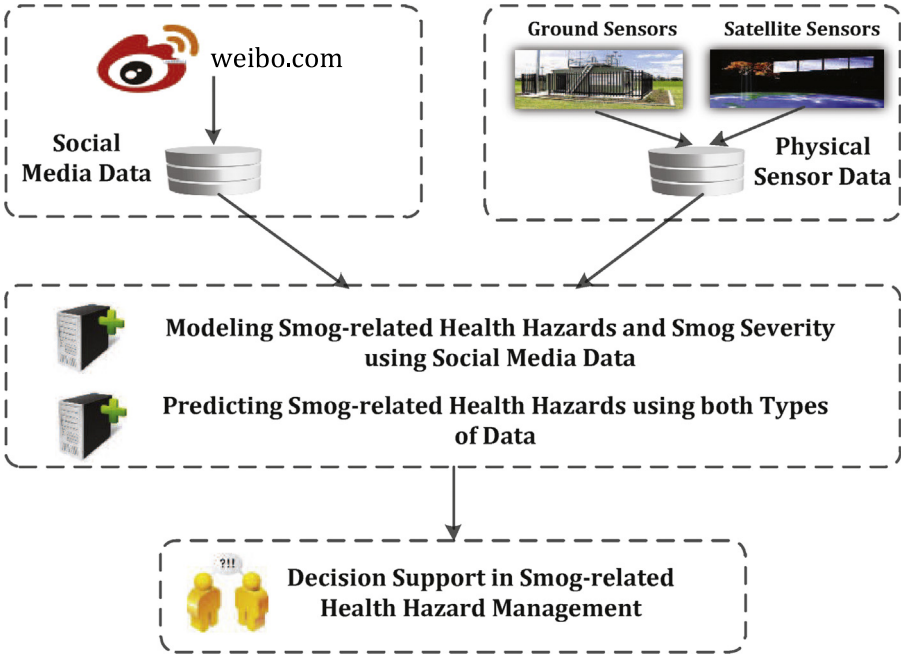

Fig. 1. Predicting smog-related health hazard with social media and physical sensor data.

Please cite this article as: J. Chen, et al., (2016), http://dx.doi.org/10.1016/j.is.2016.03.011

Download English Version:

https://daneshyari.com/en/article/4945161

Download Persian Version:

https://daneshyari.com/article/4945161

Daneshyari.com