Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

Detecting influenza states based on hybrid model with personal emotional factors from social networks

Xiao SUN [a,*], Jiaqi YE [a], Fuji REN [a,b]

[a] School of Computer and Information, Hefei University of Technology, TunXi Road No.193. 230009, Hefei, Anhui, China
[b] Faculty of Engineering, The University of Tokushima, 770–8506, Tokushima, Japan

ARTICLE INFO

Article history:
Received 15 April 2015
Received in revised form
5 January 2016
Accepted 10 January 2016

Keywords:
Influenza detection
Personal emotional factors
Transition time detection
Social web mining
Hybrid model

ABSTRACT

In this paper, we exhibit how social media data can be used to detect and analyze real-word phenomena with several data mining techniques. We investigate the real-time flu detection problem and propose a flu state detection model with personal emotional factors and semantic information (Em-Flu model). First, we extract flu-related microblog posts automatically in real-time using a hybrid model composed by Support Vector Machine with features extracted from Restricted Boltzmann Machine. In order to overcome the limitation of 140 words for posts, expect for sentiment related features, association semantic rules are also adopted as additional features, such as bag of words, negative words, degree adverbs and sentiment words dictionary. For flu state detection at specific location, we propose an unsupervised model based on personal emotional factors to figure out what state of flu in specific place. For comparison, a supervised model is also built by adopting Conditional Random Fields to decide whether a poster has "really" catch flu and what influenza stage the poster is in. Some statistic methods and prior rules are adopted in supervised model to get the flu state of specific locations by counting the number of microblog posts in different flu states. By considering personal emotional factors, spatial features and temporal patterns of influenza, the performance of unsupervised and supervised models are both improved. The system could tell when and where influenza epidemic is more likely to occur at certain time in specific locations. In different experiments results, the hybrid models show robustness and effectiveness than state-of-the-art unsupervised and supervised model only considering the number of posts.

1. Introduction

1.1. Background

Influenza is a highly contagious acute respiratory disease caused by influenza virus with a widespread activity which occurs every year. As the highly genetic variation, influenza can cause global epidemic, which not only brings huge disaster to people's life and health, but also has significant disruptions for country economy. There are about 10–15% of population who get influenza every year, which results in up to 50 million illnesses and 500,000 deaths in the world each year [1]. Influenza is a worldwide public health problem and there are no effective measures to control its epidemic at present [2,3]. Global attention has been drawn to this issue from both medical and technical perspectives. The prevalence of influenza in China is one of the most notable problems [2]. Detection and prevention of Influenza in its earliest stage might reduce the spread range of the illness. However, influenza is unable to be detected under traditional surveillance system both effectively and efficiently, making flu-monitoring a challenging topic.

In recent years, the epidemic of SARS [4], H1N1 and H5N9 influenza make us realize that people really need to expand surveillance efforts to establish a more sensitive and effective precaution indicator system for infectious disease forecasting. In order to detect influenza epidemic timely and improve the ability of early precaution, the research of early forecasting technique is urgently needed.

1.2. Related works

Nowadays, influenza surveillance systems have been established via the European Influenza Surveillance Scheme (EISS) in Europe and the Centre for Disease Control (CDC) in US to collect data from clinical diagnoses. The research of forecasting methods started relatively late in China via Chinese National Influenza Center (CNIC) and all these systems have approximately two

* Corresponding author.
E-mail address: sunx@hfut.edu.cn (X. SUN).

http://dx.doi.org/10.1016/j.neucom.2016.01.107
0925-2312/

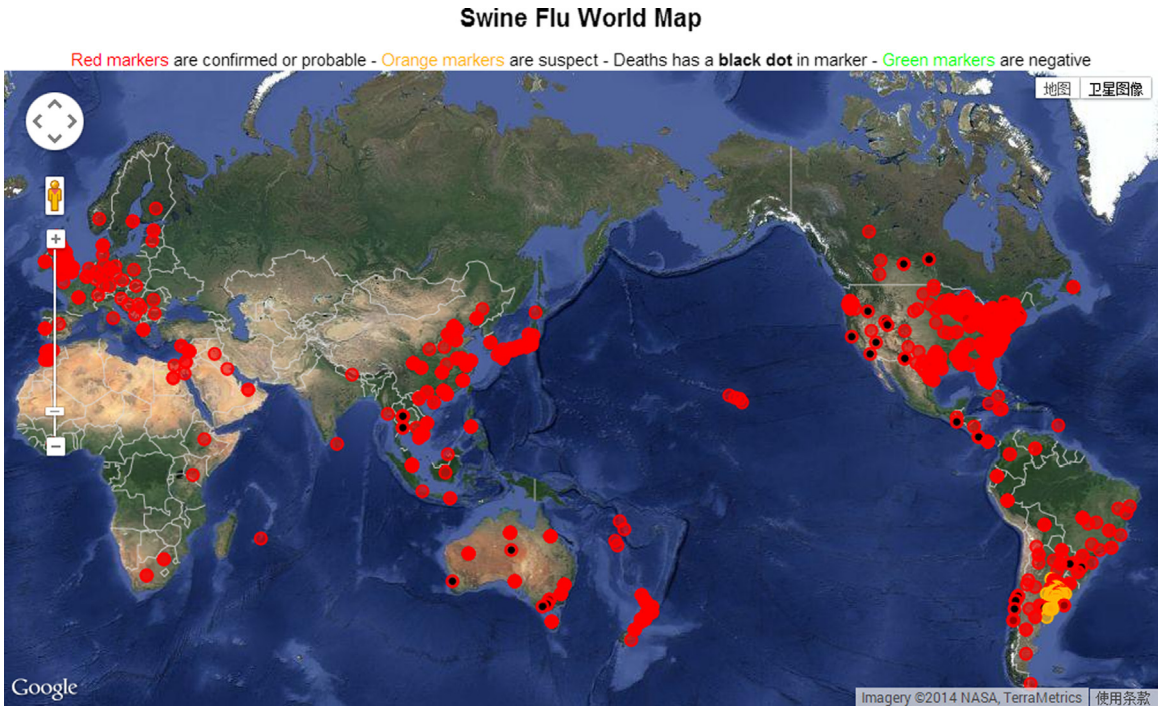

Fig. 1. World Flu Map of the year 2014.

week's delay [2]. The need for efficient sources of data for forecasting have increased due to the public health authorities' need to forecast at the earliest time to ensure effective treatment. Another popular surveillance system is Google's flu trends service [5–8] which is web search based flu reporting system. Google's flu trend adopts keywords statistic and uses linear model to link the influenza-like illness visits. Fig. 1 shows the world flu map of the year 2014 from Google's flu trends service which shows that the crisis of influenza is still serious.

In recent years, social media has become important and omnipresent for social contact and information sharing. Social media, for instance, Facebook, Twitter, Sina Weibo, has become a popular platform among people on which they could create, share, and propagate information easily. Social media gains ascendancy over traditional media because of its better performance in stability, fast propagation, and efficient resource utilization [9]. Sina Weibo is a Chinese popular microblog service that can potentially provide a perfect source for early-stage flu detection due to its large data scale and real-time characteristic as large number of active users posting about their daily life continually. When flu breaks out, infected users might post related microblog with corresponding emotions in a timely way which can be regarded as indicators or sensors of influenza states. Researchers have begun to mine social media data to predict varieties of social, economic, health, and entertainment related phenomena. Recent works have demonstrated that prediction of varieties of phenomena can be made by adopting social media data. Based on the real-time data of microblog, there has been some applications such as earthquake detection [10], public health tracking [11,12], and also flu detection [13]. Espino proposed a public health early warning system by utilizing data from telephone triage which is a public service to give advice to users via telephone in 2003 [14]. They obtained data from a healthcare call center services and software company. By investigating the relationship between the number of telephone calls and influenza epidemics, they reported a signification correlation. Magruder utilized the amount of over the computer drug sales to build a possible early warning indicator of human disease like influenza [15]. Influenza patients' requirement for anti-influenza drugs makes this approach reasonable. They reported the magnitude of correlations between clinical data and some OTC sales data and then measured the time lead after controlling for day of week effects and some holiday effects. Ginsberg built a system, utilizing Google web search queries to generate more comprehensive models for use in influenza surveillance [16]. Their approach demonstrated high precision, obtaining an average correlation of 0.97 with the CDC percentage. Lampos proposed a regression model, by applying Balasso which is a bootstrapped version of Lasso, for tracking the prevalence of ILI in part of UK using the contents of Twitter [17]. Aramaki proposed a system to detect influenza epidemics. First, the system extracts influenza related tweets via Twitter API [18]. Next, a support vector machine based classifier is used to extract tweets that mention actual influenza patients.

Previous approaches in flu state detection by social media have their advanced achievements but also have some shortcomings such as limited words of posts may impede the classification [19–23], semantic information was seldom considered which might be important for the flu detection because people often show obvious unstable emotions during the period of flu, spatial and temporal information on detect flu state were not fully considered as solutions. To meet with the limitations above, we take SVM-RBM with association rules and sentiment analysis to extract flu-related microblog posts automatically and then we build an unsupervised and a supervised model to compare the results on detecting influenza situation when deploying other state-of-the-art methods.

1.3. Proposed method

In this paper, we illustrate how social media can be used to detect and analyze influenza epidemics in China by utilizing the posts from Sina microblog. When flu breaks out, infected users might post related microblog with corresponding emotions in a timely way which can be regarded as indicators or sensors of influenza. The outline of our flu detection system is shown in Fig. 2.

We first extract influenza-related posts from Sina microblog. The most common influenza symptoms are fever, chills, sore

Please cite this article as: X. SUN, et al., Detecting influenza states based on hybrid model with personal emotional factors from social networks, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.01.107