CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.2020-0006

Vol. **1**, No. 1, pp. 186-206
March 2020

Detecting Suspected Epidemic Cases Using Trajectory Big Data

Chuansai Zhou[1], Wen Yuan[2], Jun Wang[2], Haiyong Xu[3], Yong Jiang[3], Xinmin Wang[1,4], Qiuzi Han Wen[1,4,*] and Pingwen Zhang[1,4,*]

[1] *School of Mathematical Sciences, Peking University, Beijing 100871, China.*
[2] *Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China.*
[3] *China Mobile Information Technology Co., Ltd., Beijing, China.*
[4] *National Engineering Laboratory for Big Data Analysis and Application, Peking University, Beijing, China.*

Received 3 March 2020; Accepted 20 March 2020

Abstract. Emerging infectious diseases are existential threats to human health and global stability. The recent outbreaks of the novel coronavirus COVID-19 have rapidly formed a global pandemic, causing hundreds of thousands of infections and huge economic loss. The WHO declares that more precise measures to track, detect and isolate infected people are among the most effective means to quickly contain the outbreak. Based on trajectory provided by the big data and the mean field theory, we establish an aggregated risk mean field that contains information of all risk-spreading particles by proposing a spatio-temporal model named HiRES risk map. It has dynamic fine spatial resolution and high computation efficiency enabling fast update. We then propose an objective individual epidemic risk scoring model named HiRES-p based on HiRES risk maps, and use it to develop statistical inference and machine learning methods for detecting suspected epidemic-infected individuals. We conduct numerical experiments by applying the proposed methods to study the early outbreak of COVID-19 in China. Results show that the HiRES risk map has strong ability in capturing global trend and local variability of the epidemic risk, thus can be applied to monitor epidemic risk at country, province, city and community levels, as well as at specific high-risk locations such as hospital and station. HiRES-p score seems to be an effective measurement of personal epidemic risk. The accuracy of both detecting methods are above 90% when the population infection rate is under 20%, which indicates great application potential in epidemic risk prevention and control practice.

AMS subject classifications: 62, 92

Key words: Trajectory big data, spatio-temporal modeling, machine learning, suspected case detection, epidemic risk prevention and control.

*Corresponding author. *Email addresses:* qiuzi.wh@pku.edu.cn (Q. H. Wen), pzhang@pku.edu.cn (P. Zhang), chuansai@pku.edu.cn (C. Zhou), yuanwen9510@pku.edu.cn (W. Yuan), jwwwang@pku.edu.cn (J. Wang), xuhaiyong@chinamobile.com (H. Xu), jiangyong@chinamobile.com (Y. Jiang)

1 Introduction

One of our greatest challenges is the continuing global impact of infectious diseases. On Mar 11th of 2020, The World Health Organization declared spread of the novel coronavirus COVID-19 is a global pandemic, when 118,000 confirmed cases of COVID-19 were found in 114 countries, with 4,291 deaths [1]. Although SARS-CoV-2, the virus that causes COVID-19, has been found to have lower case fatality rate than either SARS or Middle East respiratory syndrome-related coronavirus (MERS-CoV) [2], the sheer speed of its geographical expansion and surge in numbers of confirmed cases severely impact public health system. Human-to-human transmission seems to be the main method of transmission for SARS-CoV-2, according to CDC of the United States [3]. The human-to-human transmission routes of SARS-CoV-2 include direct transmission, such as cough, sneeze, droplet inhalation transmission, and contact transmission, such as the contact with oral, nasal, and eye mucous membranes [4]. But limitation of our knowledge on SARS-CoV-2 and diversified symptoms of the infected patients complicate the diagnosis of COVID-19 [2,5]. In addition, a significant percentage of infected patients, ranging from 18% to 50% as estimated in different research, are asymptomatic or with mild symptoms, but they could still be highly contagious [6]. Therefore, early identification of infected individuals and blocking their transmission paths are keys to effectively stop virus from spreading.

Classic epidemic models such as SIR or exponential growth model use mathematical and statistical approaches to quantify the dynamic mechanism of epidemic transmission and to predict the size of the infected population. In these epidemic models, basic reproduction number R_0 is the key parameter for quantifying the virus epidemic. Liu et al. reviewed published studies on estimation of R_0 of COVID-19 extracted from PubMed, bioRxiv and Google Scholar during Jan 1st to Feb 7th, 2020 [7]. Results show that the estimated R_0 varies from 2.2 to 6.49, and estimations obtained from the mechanism-based dynamic modeling is significantly higher than those obtained using statistical methods based on exponential growth model. Lack of robustness in estimation of the basic reproduction number may result in misleading estimates of the epidemic trend of COVID-19 as well as the size of infected population. In addition, these epidemic models cannot predict the infection status of a specific individual, thus can only play a limited role in practice of epidemic prevention and control. Epidemiology investigation used to be the main method of exploring transmission process of infected individuals, but "omissions and errors in previous activities can occur when the investigation is performed through only a proxy interview with the patient" [8].

With the development of emerging technologies such as cloud computing, big data and artificial intelligence, epidemiological research as well as epidemic prevention and control methods are undergoing innovation. For example, based on search engine query data, Google developed an approach, i.e., the Google Flu Trends (GFT) model, for detecting influenza epidemics through monitoring health-seeking behavior in the form of queries to online search engines. It provides estimates on the degree of influenza activ-