Chapter 44
Modelling Propagation of Public Opinions on Microblogging Big Data Using Sentiment Analysis and Compartmental Models

Youjia Fang
Virginia Tech, USA

Zheng Song
Virginia Tech, USA

Xin Chen
Virginia Tech, USA

Tianzi Wang
Virginia Tech, USA

Yang Cao
Virginia Tech, USA

ABSTRACT

Compartmental models have been used to model information diffusion on social media. However, there have been few studies on modelling positive and negative public opinions using compartmental models. This study aimed for using sentiment analysis and compartmental model to model the propagation of positive and negative opinions on microblogging big media. The authors studied the news propagation of seven popular social topics on China's Sina Weibo microblogging platform. Natural language processing and sentiment analysis were used to identify public opinions from microblogging big data. Then two existing (SIZ and SEIZ) models and a newly developed (SE2IZ) model were implemented to model the news propagation and evaluate the trends of public opinions on selected social topics. Simulation study was used to check model fitting performance. The results show that the new SE2IZ model has a better model fitting performance than existing models. This study sheds some new light on using social media for public opinion estimation and prediction.

DOI: 10.4018/978-1-7998-0417-8.ch044

INTRODUCTION

With the fast development of Internet, last decade has witnessed information explosion over social media. Major microblogging platforms such as Twitter in the United States and Sina Weibo in China provide the instant access to rapid-changing news and allow users to generate self-organized networks on popular social topics and celebrities (Gao, Abel, Houben, & Yu, 2012). As microblogging platforms have millions of users and billions of new messages every day, they have become an important channel to evaluate and predict the trends of public opinions on high-impact news and social events. But the high volume, velocity and complexity of microblogging messages also impose challenges on the processing and analysis under big data context.

The mechanism of news propagation is similar to the epidemiology of infectious disease. Disease diffusion starts with a small group of infected, and transmits along the paths of daily contact and interaction. Compartmental models in epidemiology are a state-of-practice approach to model the dynamics of disease propagation. The population is classified into several compartments, such as susceptible (*S*), exposed (*E*), infected (*I*) and recovered (*R*), to represent different epidemic states (Bettencourt, Cintrón-Arias, Kaiser, & Castillo-Chávez, 2006). People in susceptible compartment are healthy and have the opportunity to be infected; the susceptible will be infected if they have sufficient contact with the infected (Hethcote, 2000). There is a prolonged incubation period (i.e. exposed compartment) for some diseases, where people who have contracted the disease need a longer time to develop symptoms and become contagious. For some other diseases, the infected can fully recover and obtain immunity. Classical compartmental models for biological epidemiology include SI, SIS, SIR, and SEIR models. Different models are suitable for diseases with different characteristics.

Information diffusion has been widely studied by previous literatures (Canneyt, Schockaert, & Dhoedt, 2015; Escamilla et al., 2016; García-Silva, Rodríguez-Doncel, & Corch., 2013; Gruhl, Guha, Liben-Nowell, & Tomkins, 2004; Kwon & Lee, 2013; Lerman & Ghosh, 2010; Matsubara, Sakurai, Prakash, Li, & Faloutsos., 2012; Romero, Meeder, & Kleinberg., 2011; Shuai et al., 2012; Trpevski, Tang, & Kocarev, 2010; Weng, Lim, Jiang, & He, 2010; Jaewon Yang & Leskovec, 2010). Previous studies primarily focus on the source and mechanism of information diffusion in social media and its impact on public society. Epidemiological models have been used in some of these studies (Cha, Haddadi, Benevenuto, & Gummadi, 2010; Jin, Dougherty, Saraf, Cao, & Ramakrishnan, 2013; Jin et al., 2014; Khelil, Becker, Tian, & Rothermel, 2002; Newman, 2002; Xiong, Liu, Zhang, Zhu, & Zhang, 2012; Jiang Yang & Counts, 2010). Among all compartmental models in epidemiology, *SIZ* and *SEIZ* models are two most popular models for information diffusion, as they include a skeptical compartment not often seen in biological epidemiological process.

Some researchers have used *SEIZ* model to model news diffusion using Twitter microblogging data (Jin et al., 2013; Jin et al., 2014). They assumed that microblogging users who posted messages on popular news have a positive opinion on the news (i.e. being infected), and users who posted no message have a negative opinion on the news (i.e. being skeptical). With this treatment, the researchers only observed data of the infected compartment. However, the microblogging users can in fact post messages to express their both positive and negative opinions, so previous assumptions and treatments may not be completely reasonable.

The objective of this study is to simultaneously model the propagation of positive and negative public opinions using a large volume of Sina Weibo microblogging data. We used natural language processing and sentiment analysis to process the big microblogging data and extract the positive and negative