International Journal of Information Management 44 (2019) 25–37

Contents lists available at ScienceDirect

International Journal of Information Management

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

Social media data and post-disaster recovery

Mehdi Jamali[a,*], Ali Nejat[b], Souparno Ghosh[c], Fang Jin[d], Guofeng Cao[e]

[a] Department of Civil, Environmental, and Construction Engineering, Texas Tech University, Office: 104, Lubbock, TX 79409, United States
[b] Department of Civil, Environmental, and Construction Engineering, Texas Tech University, Office: 123, Lubbock, TX 79409, United States
[c] Department of Mathematics & Statistics, Texas Tech University, Office: 234, MATH, Lubbock, TX 79409, United States
[d] Department of Computer Science, Texas Tech University, Box 43104, Lubbock, TX 79409-3104, United States
[e] Department of Geosciences, Texas Tech University, MS 1053, Lubbock, TX 79409-1053, United States

ARTICLE INFO

Keywords:
Temporal–spatial patterns
Post-disaster recovery
Social media
Twitter

ABSTRACT

This study introduces a multi-step methodology for analyzing social media data during the post-disaster recovery phase of Hurricane Sandy. Its outputs include identification of the people who experienced the disaster, estimates of their physical location, assessments of the topics they discussed post-disaster, analysis of the tract-level relationships between the topics people discussed and tract-level internal attributes, and a comparison of these outputs to those of people who did not experience the disaster. *Faith-based*, *community*, *assets*, and *financial* topics emerged as major topics of discussion within the context of the disaster experience. The differences between predictors of these topics compared to those of people who did not experience the disaster were investigated in depth, revealing considerable differences among vulnerable populations. The use of this methodology as a new Machine Learning Algorithm to analyze large volumes of social media data is advocated in the conclusion.

1. Introduction

A natural disaster negatively impacts all aspects of one's life. It can not only devastate the physical settings of a community by destroying infrastructure, the landscape, residential and businesses properties, it can also affect one's emotional well being after witnessing loss of life and suffering the disruption of established social interactions. Aside from such immediate mental and physical harm, disasters also have long-term consequences, such as job losses, financially insurmountable property damage, and post-traumatic stress disorder (PTSD). Since the routines of daily life are tightly interwoven with the stability of both physical settings and social interactions, disasters upend the tranquility of people's lives for short, and sometimes long periods of time.

A return to normalcy is the ultimate goal of post-disaster recovery policies. A robust understanding of the patterns and types of damage common to disasters in general is crucial in the process of formulating effective post-disaster recovery policies and programs. Disasters are complex. They impact survivors' quality of life through the damage they inflict on natural and manmade landscapes. The existing literature distinguishes five categories of impacts (Lindell & Prater, 2003): *social impacts*, such as the appearance of conflicts and the loss of social capital (Lindell & Prater, 2003); *psychosocial impacts*, such as post-traumatic stress disorder (PTSD) (Gleser, Green, & Winget, 2013; Steinglass & Gerrity, 1990); *demographic impacts*, such as changes in population distribution (Kaniasty & Norris, 1993; Smith & McCarty, 1996); *socio-economic impacts*, such as job loss and business closures (Okuyama & Chang, 2013); and *political impacts* (Drury & Olson, 1998; Toya & Skidmore, 2014). These impacts are obviously interconnected. For instance, improvements in the socioeconomic condition of a community will influence psychosocial attributes (generally in positive ways), or, abrupt disruption of pre-established social and communal interactions may bring about adverse psychosocial and political reactions. Moreover, the relative importance of these categories can differ among communities and even individuals in the same community who experience the same disaster event, based on the innate characteristics of that community or individual. These characteristics can comprise a many different parameters, including job/income (Fothergill & Peek, 2004; Masozera, Bailey, & Kerchner, 2007), ethnicity (Bolin & Bolton, 1986), and age/gender (Nakagawa & Shaw, 2004), among others.

The recovery priorities of disaster survivors (Nejat, Brokopp Binder, Greer, & Jamali, 2018; Quarantelli, 1999; Wold, 2006) play a major role in the success of disaster recovery policies and programs (Ragini, Anand, & Bhaskar, 2018). Individual priorities are strongly influenced by income, age, and social capital. Hence, the design of a post-disaster recovery plan is a dilemma complicated by a diversity of parameters, internal attributes, unique impacts of a given disaster, and survivors'

* Corresponding author.
E-mail addresses: mehdi29j@gmail.com (M. Jamali), ali.nejat@ttu.edu (A. Nejat), souparno.ghosh@ttu.edu (S. Ghosh), fang.jin@ttu.edu (F. Jin), guofeng.cao@ttu.edu (G. Cao).

https://doi.org/10.1016/j.ijinfomgt.2018.09.005
Received 13 December 2017; Received in revised form 15 September 2018; Accepted 15 September 2018
0268-4012/

priorities. A major objective of recovery plans is the swift return of impacted communities to normalcy. Researchers and policymakers must therefore be able to quickly arrive at an accurate understanding of the relationship between community characteristics, individual personal internal attributes, and survivors' post-disaster priorities design the most effective recovery plans.

Modern social media applications, which have achieved considerable penetration into the everyday life of many users, provide an invaluable source of data regarding user thoughts, beliefs, and opinions. Social media consists of users with diverse backgrounds, and have ability to encourage aggregation of the users, can provide unique substrate for researchers to understand behavioral patters of communities (Dhir, Kaur, & Rajala, 2018; Kapoor et al., 2018; Liu, Lee, Liu, & Chen, 2018; Liu, Shao, & Fan, 2018). Research shows that heavy social media users seek out contacts, content boosts, favorable information, requirement inquiries, stress discharge, "emotional support", and sense of belonging (Gilbert & Karahalios, 2009; Kaplan & Haenlein, 2010; Liu, Shao et al., 2018). As an example, Grover, Kar, and Davies (2018) showed that providing emotional support, creating awareness, and sharing information are important users' motivations that health related industries seeking in social media platforms. Disaster cause severe, instantaneous distress on individuals, who as a result seek to mend their emotional traumas through social media outlets (Gao, Barbier, & Goolsby, 2011; Hughes, Palen, Sutton, Liu, & Vieweg, 2008). Therefore, these applications can provide invaluable data with which to study peoples' priorities after disaster strikes (Li, Zhang, Tian, & Wang, 2018). Currently, Facebook, Twitter and Instagram are major examples of worldwide social media applications with millions of everyday users scattered around the world. Fortunately, their data are relatively publicly available for research purposes. However, as discussed by Stieglitz, Mirbabaie, Ross, and Neuberger (2018), volume and variety of the social media data is the most common cited challenge in social media studies. Unfortunately, besides the volume and variety of data, the complex nature of a given disaster's consequences, makes said data less applicable in post-disaster recovery studies. Also, due to variety of users with wide range of purposes it is necessary to better understand the role of social media data in emergency management (Kim, Bae, & Hastak, 2018; Martínez-Rojas, del Carmen Pardo-Ferreira, & Rubio-Romero, 2018). To better utilize social media data in post-disaster recovery studies, we introduce a new methodology to analyze social media data (specifically Twitter data) and scrutinize community reactions in the aftermath of a disaster, based on social media statements. The methodology answers three questions: 1) What are the priorities of people who have experienced a disaster? 2) What are the tract-level relationships between these priorities and survivors' internal attributes? 3) How do these priorities differ between people who experienced the disaster and those who did not?

Definitely, understanding the priorities of people who have experienced the disaster is an indispensible part of designing an effective post-disaster recovery policy. This perception can assist policy makers to optimize the distribution of federal resources, and enhance the planning for reconstruction process. In order to design an effective post-disaster recovery policy it is not only important to understand the overall priorities of disaster victims, but also it is important to figure out regional priorities of victims which may be differing based on communal internal attributes. Therefore, understating the tract-level relationship between these priorities and survivors' internal attributes can assist policy makers to figure out regional priorities of disaster victims. Finally, in this study we will not only reveal overall and regional variations of priorities, but also we will study how these priorities may affect for people who did not experience the disaster. This outcome may again help policy makers to design better policies for affected and non-affected zones.

Twitter, created in 2006, is a micro-blog social media tool where registered users read and write messages, called "tweets," of up to 140 characters and unregistered users can read such messages (Twitter, 2016). Twitter is available via website, short message services (SMS), and mobile app (Twitter, 2016). As of December 2016, Twitter had more than 300 million monthly active users, with more than 1 billion monthly visits to the website (Twitter). According to estimates by InternetLiveStats (2017), Twitter publishes around 200 billion tweets each year, or approximately 6500 tweets per second. Additionally, as of December 2016, Twitter had more than 67 million active users in the United States, which made up about 20% of the service's active users (InternetLiveStats, 2017). Fortunately, when Twitter, Inc. promoted the Twitter API, tweet data became available for research purposes. This data found wide applications in several research studies in political science where (Tumasjan, Sprenger, Sandner, & Welpe, 2010) analyzed tweet contents for sentiments like anxiety, anger, sadness, and compared their correspondence with subjects' political parties; in human studies where (Bakshy, Hofman, Mason, & Watts, 2011) used graph analysis of followers among 1.6 M Twitter users to define the impact of users on the whole community of users; in human mobility and geography where (Leetaru, Wang, Cao, Padmanabhan, & Shook, 2013) illustrated the year to year growth of social media and visualized the impact of social media on human communication by mapping the available world-wide geo-tagged tweets; in public health where (Cao et al., 2015) developed a spatiotemporal model to understand the movements of Twitter users within specific geographic boundaries; in economics where (Bollen, Mao, & Zeng, 2011) defined public mood as the percetnage of positive tweets and the time series analysis of Dow Jones Industiral Avearage (DJIA) was shown to be predictable by the public mood index; and in education where (Junco, Heiberger, & Loken, 2011) showed engagment in social media can lead to an increase in student and faculty communication and improve the performance of the education process. As another interesting example, Nisar, Prabhakar, and Patil (2018) discussed how sports clubs absorb fans' attention by strategizing their activity in social media applications. As such, social media provides unique substrate to study human behavior, their sentiments and their decision making in everyday situations of life (Dong & Wang, 2018; Jeong, Yoon, & Lee, 2017; Lee & Hong, 2016).

Although the methodology discussed in this study can be generalized to all disasters, Hurricane Sandy was chosen for this case study due to the abundance of available data and the totality of damage it inflicted. Hurricane Sandy formed on October 22, 2012 and faded on November 1. It affected 24 states in all, including those of the United States Eastern Seaboard from Florida to Maine (Force, 2013). Hurricane Sandy was the second most destructive natural disaster in United States history. It caused more than $85 billion in damage and more than 200 fatalities (Force, 2013). The hurricane and its associated floods and fires affected millions of people in both urban and suburban communities, caused power outages, impeded transportation systems, destroyed residential properties, and produced more than an estimated $32 billion in economic losses (Bloomberg, 2013).

In the remaining sections of this manuscript, we will discuss a literature review on manuscripts related to post-disaster recovery indicators, post-disaster studies on Twitter data, suitable text mining algorithms and statistical descriptive method. Following literature review we will discuss theoretical basis of our study in which we will provide a framework for our expected results. And after that, we will have methodology in which will discuss in detail the thirteen steps of the algorithm. Following the methodology we will discuss the findings and at the end we have conclusion and future work.

2. Literature review

2.1. Disaster recovery and related indicators

According to Chang (2010), post-disaster recovery ought to be judged in terms of either returning environments to their pre-disaster conditions, building communities up to where they would have progressed if disaster had not struck, or finding a middle ground between the two. Chang's study suggests a framework for measuring the success