International Journal of Information Systems for Crisis Response and Management
Volume 8 • Issue 3 • July-September 2016

Predicting Tweet Retweetability during Hurricane Disasters

Venkata Kishore Neppalli, University of North Texas, Computer Science and Engineering, Denton, TX, USA
Cornelia Caragea, University of North Texas, Computer Science and Engineering, Denton, TX, USA
Doina Caragea, Kansas State University, Department of Computer Science, Manhattan, KS, USA
Murilo Cerqueira Medeiros, University of North Texas, Computer Science and Engineering, Denton, TX, USA
Andrea H Tapia, Pennsylvania State University, University Park, PA, USA
Shane E. Halse, Pennsylvania State University, University Park, PA, USA

ABSTRACT

Twitter is a vital source for obtaining information, especially during events such as natural disasters. Users can spread information on Twitter either by crafting new posts, which are called "tweets," or by using the retweet mechanism to re-post previously created tweets. During natural disasters, identifying how likely a tweet is to be retweeted is crucial since it can help promote the spread of useful information in a social network such as Twitter, as well as it can help stop the spread of misinformation when corroborated with approaches that identify rumors and misinformation. In this paper, we present an analysis of retweeted tweets from two different hurricane disasters, to identify factors that affect retweetability. We then use these factors to extract features from tweets' content and user account information in order to develop models that automatically predict the retweetability of a tweet. The results of our experiments on Sandy and Patricia Hurricanes show the effectiveness of our features.

KEYWORDS

Disaster Events, Hurricane Patricia, Hurricane Sandy, Retweetability Analysis, Retweetability Prediction, Twitter

INTRODUCTION

In response to increased online public engagement and the emergence of digital volunteers, emergency responders have sought to better understand how they too can use online media to communicate with the public and collect intelligence (Denef, Bayerl, & Kaptein 2013; Latonero & Shklovski 2011; Sutton et al. 2014; St. Denis, Palen, & Anderson 2014). Many emergency decision makers see the data produced through crowdsourcing as ubiquitous, rapid and accessible - with the potential to contribute to situational awareness (Vieweg et al. 2010). As the use of public social media in crisis increased, emergency responders started to take notice of the way citizens engaged in social media and the information exchanges that took place there (Hughes & Tapia, 2015). Consequently, responders began to consider if social media might be a useful tool for their practice. Research revealed that social media could be used to distribute information quickly to a wide-spread audience (Kodrich & Laituri 2011) and to engage more directly in a two-way conversation with members of the public (Hughes & Palen 2012). However, incorporating the products of digital volunteer activity into professional emergency practice has proven to be challenging due to issues with credibility, liability, training, and organizational process and procedure (Hughes & Palen 2012; Tapia et al. 2011).

The information that the public produced looked to be useful, as researchers showed that it could contribute to situational awareness during a crisis event (Cameron et al. 2012; Ireson 2009).

DOI: 10.4018/IJISCRAM.2016070103

According to Starbird, Munzy and Palen (2012), social media data that can be identified as coming from local bystanders to a disaster can be extremely important to emergency responders. Most of the social media data surrounding a disaster are derivative in nature: information in the form of reposts or pointers to information available elsewhere (Starbird et al., 2010). These derivative data are abundant, as a form of noise that must be filtered out to arrive at the signal of good data (Anderson & Schram, 2011). A small subset of the data comes from locally affected populations in the form of citizen reports (Starbird et al., 2010). Starbird et al. (2010) assert that bystanders "on the ground are uniquely positioned to share information that may not yet be available elsewhere in the information space … and may have knowledge about geographic or cultural features of the affected area that could be useful to those responding from outside the area."

Hughes and Palen (2012) examined the role of social media in emergency management and found that emergency managers see the potential of social media as means of engaging the public quickly and widely during a crisis. Vieweg et al. (2010) showed that retweeted tweets are likely to contain information that contributes to situational awareness and are likely to be actionable compared with non-retweeted tweets. In addition to the information that creates awareness to the responders, people also post information related to relief efforts during disasters (such as offering shelters, donations, and food), for which the target consumers are the victims who need aid. We believe that understanding how likely a tweet is to be re-tweeted seconds after it is posted has the potential to help responders to influence the speed and spread of messages, which could make substantial improvements in the relief efforts and can positively impact people who are badly affected by a disaster. However, the retweetability of a tweet is influenced by many factors including the aspects of a user who posted the information and the content present in it. In this paper, we focus on identifying factors that affect retweetability of a tweet during mass emergencies. These factors could be used in a real-time system to promote relevant tweets that convey useful information as well as to stop the spread of misinformation when corroborated with approaches that identify rumors and misinformation. Our research questions are: "In a social media (e.g., Twitter) stream of messages, what features (or factors) affect the spreading (retweetability in our case) of a message? How well do these features help in automatically predicting a message retweetability during disasters?" We specifically address these questions with our research agenda using Twitter datasets collected during Hurricane Sandy and Hurricane Patricia. Precisely, we present a supervised learning approach with a wide range of feature exploration to identify the retweetability of a tweet. Our approach, originally introduced in Neppalli et al. (2016), can help increase situational awareness and can inform emergency response organizations on how to reach the widest audience in the fastest way during disaster events. In this extended work, we augment our contributions to retweetability analysis and prediction in disaster events with our findings from a larger spectrum of experiments using state-of-the-art classification approaches and two different hurricanes: Sandy and Patricia.

Contributions and Organization

The contributions of this paper are as follows: (i) we present an analysis of retweeted tweets during Sandy and Patricia Hurricanes to determine several aspects affecting retweetability; (ii) we design features from tweets' content and user account information for learning models that predict the tweets' retweetability during disaster events, and study the predictive power of these features; (iii) we experimentally show that the models trained using the designed features perform better than strong baselines and previous approaches; and (iv) we find that the quantitative user features (e.g., #friends, #followers, #statuses) when normalized with the user age show better performance than the unnormalized user features.

City Networks in Cyberspace and Time: Using Google Hyperlinks to Measure Global Economic and Environmental Crises

Andrew Boulton, Lomme Devriendt, Stanley D. Brunn, Ben Derudder and Frank Witlox (2014). *Crisis Management: Concepts, Methodologies, Tools, and Applications (pp. 1325-1345).*

www.igi-global.com/chapter/city-networks-in-cyberspace-and-time/90780?camid=4v1a