

Visualizing and Mining Social Media Data for Smart Emergency Management

Lei Zou^{a, *}

^a Department of Geography, Texas A&M University, lzou@tamu.edu

* Corresponding author

Keywords: Social Media, Emergency Management, Disaster Resilience, Geospatial Intelligence, Geo-visualization

Abstract:

The ability of a community to prepare for, absorb, recover from, and more successfully adapt to disastrous events is defined as disaster resilience. Disaster resilience can be better understood by investigating human behaviors during the four phases of emergency management – preparedness, response, recovery, and mitigation. However, a major challenge is that data describing communities' behaviors in different phases of emergency management are difficult to access through traditional databases. Social media such as Twitter is increasingly being used as an effective platform to observe human behaviors in disastrous events. These responses and behaviors could be better understood by analyzing real-time social media data through categorizing them into different phases of the emergency management.

This research studies the Twitter use during 2012 Hurricane Sandy and 2017 Hurricane Harvey, which struck the U.S. northeast and south coasts, respectively. The objectives are fourfold: (1) to develop a Twitter data mining and visualization framework and a set of indexes for emergency management and resilience analysis; (2) to visualize the spatial-temporal patterns of disaster-related Twitter activities during the two hurricane events; (3) to examine and compare the social-geographical disparities of disaster-related Twitter activities during Sandy and Harvey; and (4) to build applications using social media data for smart management, including surveying human behaviors and emergency rescue.

A total of 200 million records of Twitter data were processed in this research. First, a Twitter data mining framework is developed (Figure 1). The data collection is implemented in a Hadoop Distributed File System (HDFS) with eight nodes. All selected background and disaster-related tweets were stored in a sharded MongoDB database. Each collected tweet was geocoded to a location using a local Nominatim server, which uses OpenStreetMap planet data. Each tweet was also assigned a sentiment score through a lexicon and rule-based sentiment analysis module. Second, we calculated three Twitter indexes to examine the social-geographical disparities. Three indexes were used in this study, including Ratio, Normalized Ratio, and Sentiment at different spatial and temporal scales. Third, we use a collection of sample tweets to survey and compare online behaviors of flood victims who requested for rescue on social media during two hurricane events. Finally, a deep neural network trained tool, Victim Locator, is developed to quickly identify disaster victims and deliver their location information to nearby first responders.

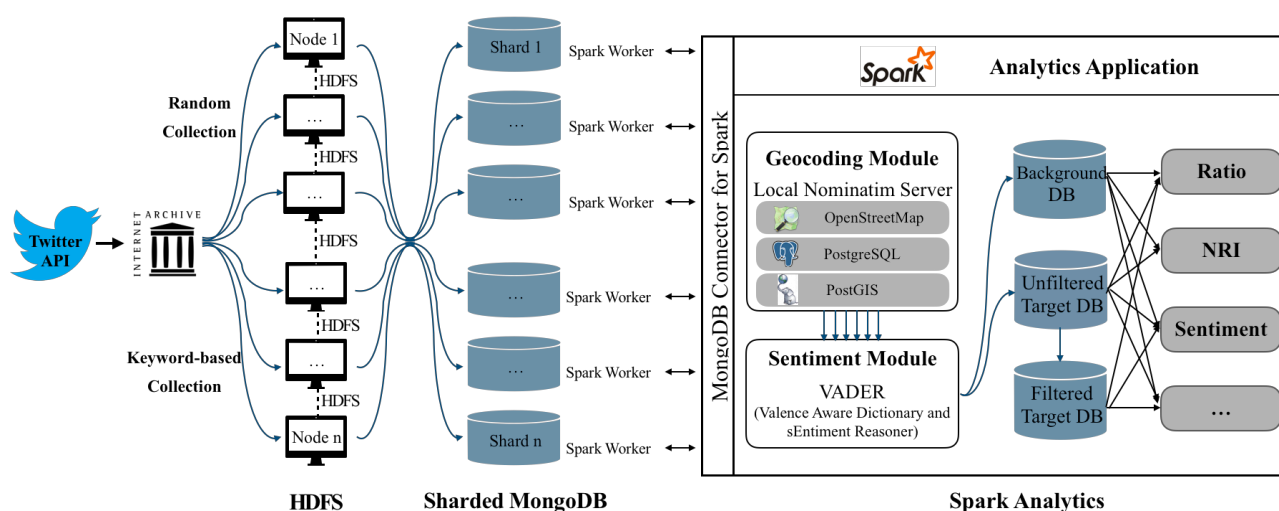


Figure 1. Twitter data mining framework

Results show that common indexes derived from Twitter data, including ratio, normalized ratio, and sentiment, could enable comparison across regions and events (Figure 2). Social and geographical disparities in disaster-related Twitter use existed in both Hurricane events during the preparedness, response, and recovery phases, with higher disaster-related Twitter use communities generally being communities of higher socioeconomic conditions. Although social media has

become more ubiquitous in 2017 than in 2012, the digital divide in social media use during disasters still existed in Harvey. In addition, social media data provide a fast, low cost, and efficient method to survey public behaviors towards any topics or during any events, such as how flood victims requested for rescue online during Harvey. Although 79% victims indicated their addresses in the tweet contents, more than 60% of them did not provide important rescue information, including number of victims, contact, and flooding status. Finally, a victim locator based on tweets requesting for rescue is developed, which can successfully locate 81% of flood victims.

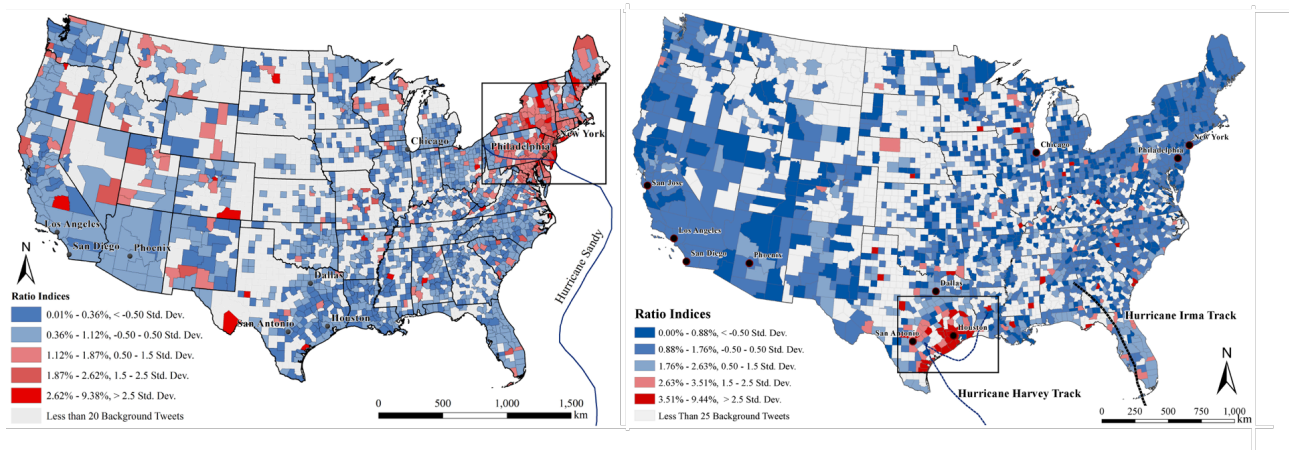


Figure 2. County-level ratio indexes during Hurricanes Sandy (2012) and Harvey (2017)

This research establishes a framework to quantitatively extract valid information from Twitter data at different spatial-temporal scales. Comparisons between different locations under the same hazard type in different years uncover the underlying geographical and social factors influencing the disparities of Twitter activities. Findings from this project shed lights on the “Big Data” challenges of developing methods for mining social media data as well as using social media data to improve understanding of resilience and smart emergency rescue.