Exploit Social Relations in Sentiment Analysis of Social Media Content for Disaster Management

Emergent Research Forum (ERF)

Pratik Shivarkar
University of Houston-Clear Lake
ShivarkarP7939@uhcl.edu

Wei Wei
University of Houston-Clear Lake
Wei@uhcl.edu

Abstract

The world has witnessed the prevailing usage of social media for communication during disasters. Being able to monitor and predict public opinions on social media during disasters allows us to design more efficient and effective communication mechanisms during critical times. However, this potential is yet to be materialized due to difficulties in sentiment analysis of social media content. We propose to augment the effectiveness of such analysis by incorporating social relations in sentiment classification models. The proposed study extends previous work substantially by looking at a larger set of social relations and focusing on different communication goals at each stage of disaster management. In addition, we look at sentiments at both individual and community levels. The study can help formulate a better understanding of how opinions are formed and propagated during disasters, thus allow stakeholders to strategize for better communication.

Keywords

Disaster Management, Social Media, Sentiment Analysis, Social Relation, Network Analysis.

Introduction

Disasters, natural or man-made, are events that adversely affect human lives due to inability to handle hazards, accidents or attacks. Unpredictability, uncertainty, unfamiliarity, and velocity of hazardous incidents often elevate the severity of the disasters. Individuals, organizations, and societies strive to reduce their exposure to the consequences of disasters, which constitutes disaster management. Such efforts entail identifying vulnerabilities, organizing and managing resources to reduce risks, mitigating the impact, and facilitating relief to disaster victims (Haddow et al. 2003). Effective disaster management requires us to strategize communication of disaster-related information. The efforts should address how to collect, exchange, and disseminate the information using a wide range of communication tools. These tools have to address many of the disaster communication challenges including technological, sociological, and organizational ones (Ryan 2017).

Studies show that social media is a viable platform for disaster communications due to its ease of access, availability, and scalability (Austin et al. 2012; Fraustino et al. 2012). Recently, government agencies start to use social media for disaster response and recovery purposes such as situational awareness and citizen communication, emergency warnings, and recovery efforts (Kavanaugh et al. 2012; Lindsay 2011). The adoption of social media for disaster management has generated a tremendous volume of content. This content is often emotion-rich because crisis triggers strong emotions. Sentiment analysis is a discipline that focuses on mining emotions, attitudes, and opinions expressed towards a subject or a topic in natural language text (Liu 2012). Sentiments can influence human behavior and decision-making, making it a crucial factor affecting how information is perceived and interpreted (Beigi et al. 2016). If sentiments in disaster-related social media content can be properly retrieved/predicted, stakeholders can modify existing or design new disaster communication mechanisms. A better understanding of sentiments in communications help achieve better community engagement, information dissemination, and situational awareness. In addition, exploring sentiments also enable organizations involved in disaster management.
to assess reputational threats, evaluate the effectiveness of their disaster response and long-term and short-term recovery efforts.

Conventional sentiment analysis techniques such as lexicon-based techniques and supervised machine learning-based techniques have limitations. The effectiveness of lexicon-based techniques depends on the quality and coverage of constructed lexicons which are hard to compose. Supervised machine learning techniques offer the better performance of sentiment classification than lexical techniques. However, they require sizeable labeled dataset for training and are application-oriented.

In our study, we propose to incorporate social relations into machine learning-based sentiment analysis (Tan et al. 2011) in the context of disaster management. The expected contribution is improved performance of sentiment classification of disaster-related social media content, by incorporating extracted social relations.

**Literature Review**

Our proposed study is interdisciplinary and builds on several streams of research as follows.

**Analyze Social Media Sentiments for Disaster Management**

Effective disaster management could benefit from collection and interpretation of data/information generated on social media. The nature and magnitude of sentiments in social media content vary at different disaster stages. Attribution Theory states that people look for the causes and attributes the responsibility of unexpected events such as disasters. According to this theory, the disaster-affected public is motivated to assign causes to their distrust and hostility towards an organization. Based on Attribution Theory, Situational Crisis Communication Theory (SCCT) (Coombs 2007) is an evidence-based framework that provides a mechanism for anticipating how stakeholders will react to a crisis in terms of the reputational threat posed by the crisis. This framework remains highly relevant for government agencies and organizations that deal with disaster response, to maintain reputation, which is essential in trust building and public engagement. Therefore, to gauge and plan for effective communication with citizens during the crisis, we could monitor and predict public opinions by analyzing social media sentiments, with a different focus at a different stage of a disaster.

Large-scale sentiment analysis has been made possible by the advent of machine learning techniques, natural language processing, and access to large datasets. Methods for sentiment analysis can be categorized as language processing based and application-oriented methods. The former focuses on the lexical and linguistic analysis of the text to assign a sentiment polarity score to a piece of content, based on the presence of opinion words such as ‘helpful’ or ‘frustrating’. The coverage of the corpus is essential to the quality of the analysis. Lexical analysis can be combined with linguistic analysis such as part-of-speech tagging. Application-oriented methods are usually applied to text documents from a specific domain. These methods use supervised machine learning techniques to predict the polarity of the sentiment of the text. This approach requires a labeled training dataset. Many features based on unigrams and bigrams, as well as metrics such as star ratings, are included in the training datasets. These methods often yield better results. However, they are only effective with text documents that more homogeneous in nature (Beigi et al. 2016). High variance in social media content during disasters renders existing sentiment analysis methods less effective.

**Use Social Relation to Improve Sentiment Analysis**

Social media is not just a place for people to publish, but also to connect. The connections can be established explicitly (user-based) or implicitly (content-based). Those networks depict some types of social relations. The dynamics of peoples’ opinions and their social relations are bi-directional. People who are socially related are more likely to share similar opinions about certain topics, and similar minds tend to flock together. This is what we call emotional contagion (Hatfield et al. 2014). More specifically, homophily (McPherson et al. 2001) and social influence (Marsden and Friedkin 1993) are the two processes that enable emotional contagion. The former suggests that people befriend those who are alike, and the latter says friends tend to become similar over time. Based on these theories, we could speculate that connected social media users may hold similar opinions (Thelwall 2010). Studies (Hu, L. Tang, et al. 2013; Tan et al. 2011)
have shown that when social relations are incorporated into the learning model, performance of sentiment classification can be improved significantly. Both studies used Twitter data, taking into consideration of the user-user networks formed on Twitter.

**Research Opportunity**

Though sentiment analysis of social media content has made progress in recent years, work on how to exploit social relations is still sparse. Furthermore, applying this approach for disaster management is a void. In this study, we propose to fill the gap by extending previous work: (1) Extend the social media platforms of study beyond Twitter. This is important for two reasons. First, the size limit of tweets poses difficulty in analyzing emotions. Second, other platforms are widely used for disaster communication and the content is richer. (2) Extend the type and scope of social relations that could be exploited to improve the performance of sentiment analysis techniques. In addition to the explicitly formed following networks, there are other implicit networks. Such networks can be studied and understood using social theories (Tang et al. 2014) to get a better perspective on how they could be incorporated into sentiment analysis. (3) Take into consideration of variation of emotions during different phases of disaster communication. (4) Apply communication theories/models to interpret the results of sentiment analysis. This helps establish better practices in strategizing how to effectively utilize social media for disaster-related communications. (5) Study sentiments at a different level. Most of the studies so far focus on document-level analysis such as sentiment of a tweet. However, the social relations formed on social media could help aggregate sentiments at a higher and meaningful level.

In this study, we define the research scope as a quadruple \((p, c, r, o)\). \(p\) is the phases in disaster communication, \(c\) is the collected social media content, \(r\) is the set of social relations, and \(o\) is the set of opinions. Based on previous work (Liu 2012), an opinion is a quintuple \((e_i, a_i, s_{ijkl}, h_k, o_i)\). The opinion measures the classification/magnitude of the sentiment that an opinion holder has at a specific time toward an aspect of an entity. In Table 1, we provide some typical use cases as instantiations of the quadruple.

<table>
<thead>
<tr>
<th>Phase ((p))</th>
<th>Social Media Content ((c))</th>
<th>Social Relation ((r))</th>
<th>Disaster Related Opinion ((o))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Disaster Warning</td>
<td>Tweets issued by a government agency, plus all comments and retweets</td>
<td>The follower-followee network</td>
<td>What type and magnitude of emotions residents hold toward the timeliness of the mandatory evaluation updates at a specific time?</td>
</tr>
<tr>
<td>2 Disaster Impact</td>
<td>Discussions taking place in emergent citizen groups from online discussion forums</td>
<td>The network of groups-topics</td>
<td>What emotions define collective goals and attitude of emergent citizen groups during disasters?</td>
</tr>
<tr>
<td>3 Disaster Relief</td>
<td>Requests for goods, information, and emergency services rendered on Facebook</td>
<td>The network of expressed needs-requester</td>
<td>What type and magnitude of emotions residents hold towards their needs of different things/services?</td>
</tr>
</tbody>
</table>

**Table 1. Use Cases of Sentiments Analysis of Disaster Communication Content**

The high-level research problem we want to address is “How to improve the classification of sentiments presented in social media content related to disaster communication, by incorporating network data into the model?” The above proposed quadruple suggests there is a wide range of interesting research questions to answer such as: (1) Which type(s) of social relation can help improve sentiment classification more? (2) Do we observe different networking patterns in different phases of disaster communication?

**Research Plan**

**Data**

From Facebook and Twitter, we collected content from official pages of local news platforms, emergent citizen groups, and organizations such as Federal Emergency Management Agency (FEMA), and American Red Cross. The Facebook dataset records contain data about the individual posts, comments, and their metrics such as likes, shares, and reactions. Dataset from Twitter is a collection of Tweets from relevant topic discussions, and conversations taking place around official handles of government and non-profit
organizations and news platforms. In addition, the collected data allows us to extract networks such as follower-followee, commenter-commented, @-network, and #-network. We are planning to further extend our dataset by collecting relevant topic discussions from Reddit, and Disqus commenting platform.

**Techniques**

In order to extract content-based networks, we need to extract issues and topics being discussed over social media and form linkages between content, replies, and users. There are similarity and ranking algorithms for this purpose. However, while dealing with large, heterogeneous data simple text retrieval techniques might fall short. Other approaches to model our topics include employing statistical and semantic analysis techniques such as but not limited to Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), Latent Semantic Analysis (LSA). Extracted topic information can be used to model social relations formed around them. Various techniques allow us to reduce the noise in generated topic models. For example, to remove irrelevant features, we can tweak commonly used text-representation techniques based on vector space model and statistical language model (Martineau et al. 2008; Wallach 2006). Or we can employ supervised machine learning techniques to improve topic modeling results (Xie and Xing 2013).

To classify sentiments, we have a plethora of supervised machine learning algorithms and methods that can be utilized. However, use of graph analysis techniques for sentiment analysis is an emerging field of research, such techniques can be witnessed in researches series of research (Pang and Lee 2004; Wang et al. 2011). Modification of the algorithms such as SANT (Hu, L. Tang, et al. 2013) or ESSA (Hu, J. Tang, et al. 2013) is needed based on social relations to be incorporated. In order to evaluate the quality and performance of our proposed models, we will evaluate the accuracy of the sentiment classification method using conventional metrics. In case of multi-class classification micro-average of precision and recall and macro-average of precision and recall can be used. The problem of lacking labeled data could be resolved using semi-supervised approach, or use human-labeled data.

**Conclusion and Future Work**

Sentiment analysis of social media content suffers from “dirty” data and lack of training datasets. These issues hindered the effectiveness of supervised techniques such as classification of sentiments. An option is to adopt novel supervised and semi-supervised machine learning techniques. We reviewed studies that employ explicit social relations to improve sentiment classification. We formally defined the research space in the context of disaster management, based on which we propose to extend relevant work in various aspects with the goal to facilitate disaster communication using social media. Preliminary studies proved the feasibilities of the proposed study and the next step is to design ways to incorporate social relations extracted into learning models, followed by the evaluation of the performance of the augmented models.

**REFERENCES**


