Extracting Non-Situational Information from Twitter During Disaster Events

Poonam Sarda, Department of Computer Science and Engineering, Government Engineering College Bikaner, Bikaner, India

Ranu Lal Chouhan, Department of Computer Science and Engineering, Government Engineering College Bikaner, Bikaner, India

ABSTRACT

Micro blogging sites have become important forums for discussion during disaster events in which Twitter has become one of the important source of real time information. Millions of tweets are posted during disasters, which include not only information about the present situation or relief efforts, but also the emotions or opinion of the masses. Much research has been conducted on extracting situational information from tweets during disaster. However, according to current knowledge, there has not been any prior attempt to study the non-situational tweets posted during disasters, such as those which express the emotions/opinions of the people, political and governmental views, raising charities and event analysis. In this study, the authors characterized the non-situational tweets posted during recent disaster events, the Nepal Earthquake and the Gurudaspur Terrorist attack. They developed a classifier to categorize various types of non-situational tweets into a set of fine-grained classes utilizing state-of-the-art machine learning technique. This system also helps in filtering out communal tweets which can make worst the situation by disrupting communal harmony during certain disaster events.

1. INTRODUCTION

Twitter, a micro-blogging social networking website has a large and rapidly growing user base. Rich bank of data is provided by the Twitter in form of 'tweets' which must be written within 140 characters. The experiment in Go et al. (2009) found that the average length of tweets is 14 words or 78 characters. Some of the applications which rely on Twitter data are analysis of disasters (Sen et al., 2015; Brynielsson et al., 2013), detection of diseases (Dai and Bikdash, 2015; Grover and Aujla, 2015; Grover et al., 2014; Aramaki et al., 2011), political elections (Asur and Huberman, 2010), movie review (Bollen et al., 2011) and stock market (Tumasjan et al., 2010). In tweets, abbreviations, orthographic mistakes, emoticons and hash tags are frequently used to express the message in few words. There have been a lot of researches on analyzing tweets posted during disasters, and most of

the prior studies have focused on extracting situational information, that is, information which helps to gain a high-level understanding of the circumstance (Sarter and Woods, 1991; Vieweg et al., 2010). For instance, several studies have been done to develop classifiers for differentiating situational tweets from other non-situational tweets (Sen et al., 2015; Imran et al., 2013), while some studies not only attempted to summarize situational tweets (Sen et al., 2015; Nguyen et al., 2015) in English but also in different languages like Hindi (https://en.wikipedia.org/wiki/April_2015_Nepal_earthquake).

On the other hand, non-situational tweets, such as those which reflect the opinion and sentiments of the masses, have received much lesser attention. Since the non-situational tweets comprise majority of the tweets posted during disasters, the lack of studies on non-situational tweets imply that most of the tweets posted during disasters remain to be analyzed. This motivates us to conduct the first systematic characterization of non-situational tweets in this study. Specifically, we characterized the non-situational tweets posted during recent two distractive events in the Indian sub-continent which are Nepal Earthquake and Gurudaspur attack (https://en.wikipedia.org/wiki/2015_Gurdaspur_attack; Sharma et al., 2015).

We observed different types of non-situational tweets which have different applications, for instance, tweets expressing opinion can be mined to gauge the public opinion on the adequacy of relief operations, while tweets about charity can be promoted such that more people can contribute to charitable efforts. For such applications, it is necessary to separate out different types of non-situational tweets, a task that needs to be automated considering the high rate at which tweets are posted. Hence, we developed a Stochastic Gradient Boosted Trees (SGBT) based classifier for automatically separating different categories of non-situational tweets. The classifier relies on a proposed set of low-level lexical and syntactic features of tweets, and performs significantly better than a baseline bag-of-words classifier. Because of these features, proposed classifier functions accurately in cross domain scenarios (that is, the classifier is trained during one disaster event and then deployed on other disaster event). Study has been attempted on extraction of situational information during disaster events (Sen et al., 2015; Brynielsson et al., 2013). Explicitly, Verma *et al.* (2011) examine that situational tweets are more objective, formal and impersonal linguistic style compared to non-situational tweets and the classification of tweets is based on bag-of-words classifier models. Verma *et al.* themselves reported that this approach does not work well in cross-domain scenario because of large dependency on the vocabulary on particular event. Therefore, this study develops event independent classifier for situational and non-situational tweets using lexical and syntactical features and hence, overcomes the limitations of bag-of-words model.

This study, to our observation, is the first on characterizing non-situational tweets gives insight in how people use social media platforms during disasters, for purposes such as expressing their sympathies or organizing charities.

2. DATASET

This section describes the datasets collected, and the various types of tweets observed in the dataset.

2.1. Disaster Events

Considered tweets posted during the following disaster events are:

- **Nepal Earthquake (NE):** A devastating earthquake in Nepal (https://en.wikipedia.org/wiki/2015_Gurdaspur_attack);
- **Gurudaspur Shoot (GS):** A terrorist attack killing seven people in Gurudaspur, India (Sharma et al., 2015).

Note that one of the events is a natural disaster, while the other is a man-made. Both of these events also occurred in developing regions (in the Indian sub-continent).

7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/extracting-non-situational-information-from-twitter-during-disaster-events/178468

Related Content

An Introduction to Kernel Methods

Gustavo Camps-Valls, Manel Martínez-Ramónand José Luis Rojo-Álvarez (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1097-1101).*

www.irma-international.org/chapter/introduction-kernel-methods/10958

Distance-Based Methods for Association Rule Mining

Vladimír Bartík (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 689-694).*

www.irma-international.org/chapter/distance-based-methods-association-rule/10895

Projected Clustering for Biological Data Analysis

Ping Deng, Qingkai Maand Weili Wu (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1617-1622).*

www.irma-international.org/chapter/projected-clustering-biological-data-analysis/11035

Statistical Metadata Modeling and Transformations

Maria Vardaki (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1841-1847).*

www.irma-international.org/chapter/statistical-metadata-modeling-transformations/11069

Visualization of High-Dimensional Data with Polar Coordinates

Frank Rehm, Frank Klawonnand Rudolf Kruse (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 2062-2067).*

www.irma-international.org/chapter/visualization-high-dimensional-data-polar/11103