IJSEAT International Journal of Science Engineering and Advance Technology

An Unprecedented Approach of Detecting and Reporting System of Earthquakes Using Tweet Analysis

Hemantha. Mareedu[1], Ella. Karunakar[2]

[1] M.Tech (CSE), NRI Institute of Technology, A.P., India.

[2]Asst Professor, Dept. of Computer Science & Engineering, NRI Institute of Technology, A.P., India.

***Abstract*— Social media has got an exponential growth in recent years. One of the most representative examples is Twitter, which allows users to publish short tweets (messages within a 140-character limit) about "what's happening". This paper focuses on detecting those events to have a better understanding of what users are really discussing about in Twitter. Event detection has long been a research topic. The underlying assumption is that some related words would show an increase in the usage when an event is happening. An event is therefore conventionally represented by a number of keywords showing burst in appearance count. In this paper, we investigated the real-time nature of Twitter, devoting particular attention to event detection like earthquake. We developed an earthquake reporting system that extracts earthquakes from Twitter. It is possible to detect an earthquake by monitoring tweets. Our system detects an earthquake occurrence and sends an e-mail, possibly before an earthquake actually arrives at a certain location. This paper is the first of its kind using social media for detecting natural calamities.**

***Keywords*— Disaster, Earthquake, Event detection, Twitter, Tweets.**

I. Introduction

Ever since social media introduced, it has dramatically changed the world. In less than a decade we have gone from being passive media consumers to become active producers of social media content. On top of the social media mountain is Facebook and Twitter. Twitter comes second and has more than 288 million [1] active users every month while the total number of users is almost twice as much [2]. Twitter users create more than 500 million tweets [2] per day. As a microblogging service Twitter lets people share 140 character long texts called tweets. According to a survey from early 2012 [3] 15 % of the American population which is online use Twitter and 8 % use it on a daily basis. Twitter is used by a diverse set of the population where income and education plays a minor role. With smart phones dominating the handset market, Twitter users tweet more and more from their smart phones. As of early 2012 9 % of all American smart phone owners used Twitter on their phone and 5 % used Twitter on a daily basis. [3] Scenarios less explored are monitoring social media during a developing crisis. Detecting events and gaining insights could help government appointed crisis handling teams to get a better overview of the situation and thus improve delegation of aid. During the Haiti earthquake social media was actively used to gain an understanding of the extent of the crisis [4]. Regardless of the scenario, detecting events by monitoring social media data is becoming increasingly important as more and more people find their way to one or more of the social media platforms.

Detecting events by monitoring social media have many difficult aspects. First of all the amount of social media content produced is enormous and coming up with an efficient solution is in many cases a non trivial exercise. Another aspect involves the content of social media and especially Twitter. A study from 2009 revealed that 40 % of all tweets are just babble [5] like, "I am eating a sandwich". Because Twitter users are limited to only 140 characters they often resort to unconventional abbreviations of words. In many cases these abbreviations can be difficult to understand. A hypothetical approach to this challenge of analyzing vast amounts of tweets could be to not use any specialized software. This basic approach to detect events on Twitter could be for a group of analysts to read through tweets and collectively reach consensus about emerging events. There are many difficulties with such an approach. Because the amount of tweets is so great the number of analysts would also have to be great. Often time is limited on such assignments which might again require more analysts. A larger group of analysts might also have problems reaching consensus than a smaller group. A dedicated workforce to coordinate the effort might also be required. It is therefore likely that scaling the analytical team would not be linear. A number of other issues would probably reveal themselves. The difficulty of the task and increased popularity of social media makes the area of event detection on social media a growing field of study.

This paper presents an investigation of the real-time nature of Twitter that is designed to ascertain whether we can extract valid information from it. We propose an event notification system that monitors tweets and delivers notification promptly using knowledge from the investigation. In this research, we take three steps: first, we crawl numerous tweets related to target events; second, we propose probabilistic models to extract events from those tweets and estimate locations of events; finally, we developed an earthquake reporting system that extracts earthquakes from Twitter and sends a message to registered users.

II. Related work

Several studies have been conducted to evaluate the capabilities of micro-blogging services during natural disasters. In [6], Twitter usage is analyzed in the context of an earthquake which occurred off coast of Chile. The authors observed after the event that people tweeted about tsunami alerts, missing or deceased persons, etc. They also surveyed the trustworthiness of tweets and verified that false rumors were much more often questioned than confirmed truths. In [7], the authors evaluate the significance of Twitter as a contribution to the situational awareness picture of two natural disaster events (Oklahoma Grassfires and Red River flood in 2009). They checked for the occurrence of geographical information (geo-tags, street names, etc.) and discovered that about 82% of the users posted at least one tweet containing geo-location information. The authors of [8] propose a micro-blogging-based earthquake reporting system. Tweet analysis was performed to identify messages on Japanese earthquakes in real-time. The system proved to be reliable for earthquakes with a seismic intensity above 3 since 96% of these earthquakes were detected 80% of them while they happened. In [9], the impact on Twitter during and after a deadly shooting killing four police officers in Seattle Tacoma, Washington, has been evaluated. The authors could show that Twitter was used by citizens and news media organizations to share event-related information during and after the event.

Another important characteristic is the near-real-time nature of information provided by users since the situational awareness can only be improved if the information is provided temporally close to the event. Twitcident [10] for example enables searching, and analyzing Twitter information streams during incidents. It listens to an emergency broadcast service which provides information about local incidents. Whenever a message comes in, it searches for related tweets which are semantically extended to allow for effective filtering during a user search. A case study was performed with Twitcident on approx. 97,000 tweets that got published around a storm event at Belgium festival [11]. The results indicated an exponential rise of tweets during the storm event and how the provided information could be helpful for crisis management.

III. Proposed work

A. Semantic Analysis of Tweets

A tweet is a post or status update on Twitter. It is often used as a message to friends and colleagues. A user can follow other users; that user's followers can read her tweets on a regular basis. A user who is being followed by another user need not necessarily reciprocate by following them back, which renders the links of the network as directed. To detect a target event from Twitter, we search from Twitter and find useful tweets. Our method of acquiring useful tweets for target event detection is portrayed in Figure 1.

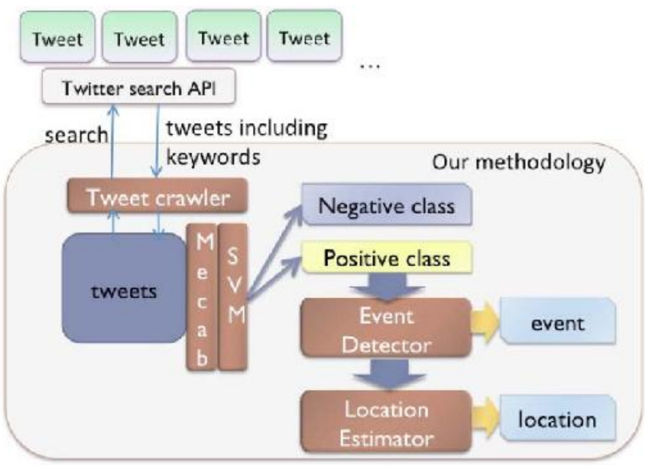

Figure 1: Method to acquire tweets referred to a target event precisely.

First, we crawl tweets including keywords related to a target event. From them, we extract tweets that certainly refer to a target event using devices that have been trained with machine learning. Second, we detect a target event and estimate the location from those tweets by treating Twitter users as "social sensors." To classify a tweet as a positive class or a negative class, we use a support vector machine (SVM), which is a widely used machine-learning algorithm. By preparing positive and negative examples as a training set, we can produce a model to classify tweets automatically into positive and negative categories.

B. Tweets as Sensors

Each Twitter user is regarded as a sensor and each tweet as sensory information. These virtual sensors, which we designate as social sensors, are of a huge variety and have various characteristics: some sensors are very active; others are not. A sensor might be inoperable or malfunctioning sometimes, as when a user is sleeping, or busy doing something else. Figure 2 presents an illustration of the correspondence between sensory data detection and tweet processing. By regarding a tweet as a sensory value associated with location information, the event detection problem is reduced to detection of an object and its location based on sensor readings. Estimating an object's location is arguably the most fundamental sensing task in many ubiquitous and pervasive computing scenarios. In this research field, some probabilistic models are proposed to detect events and estimate locations by dealing appropriately with sensor readings.

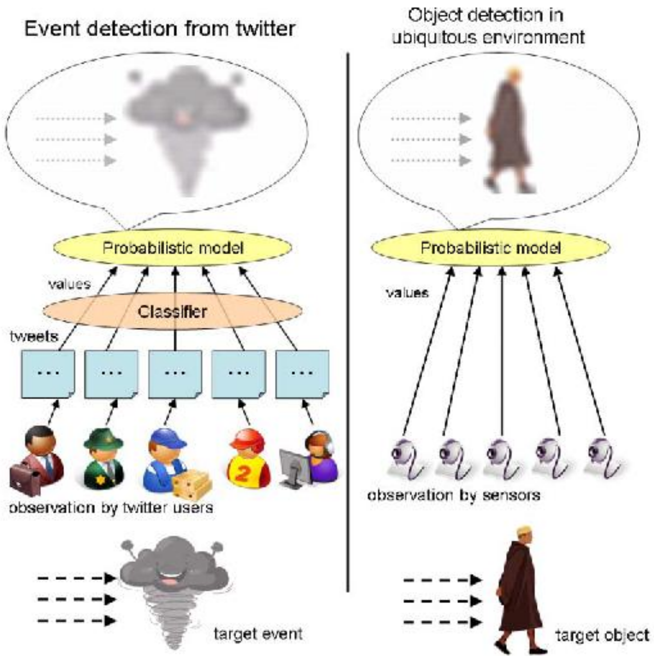

Figure 2: Correspondence between event detection from Twitter and object detection in a ubiquitous environment.

C. Temporal Model and Spatial Model

Each tweet is associated with its post time, and we use this information as the estimated occurrence time of our target event. We use GPS data and the registered location of a user as the location information, and we filtered out all the tweets without location information. We form a temporal model which gives the probability of event occurrence at time t, for a given tweet that is a positive example. If the probability is larger than a predetermined threshold, then it determines an actual occurrence of the target event. This includes choosing an appropriate threshold and the build of temporal mode:

1. The false-positive ratio *Pf* of a sensor is approximately 0.35.

2. Sensors are assumed to be independent and identically distributed.

Each tweet is associated with a location. If the probability given by the temporal model is larger than the threshold, the next step is to determine the event location. We obtained the location information of each tweet using its associated GPS data or the registered location. We then apply particle filter to all set of tweet to obtain the event location .

D. Sequence Importance Sampling

A particle filter is a probabilistic approximation algorithm implementing a Bayes filter, and a member of the family of sequential Monte Carlo methods. The Sequential Importance Sampling (SIS) algorithm is a Monte Carlo method that forms the basis for particle filters. The SIS algorithm consists of recursive propagation of the weights and support points as each measurement is received sequentially. The algorithm is presented below:

Step 1- Generation: Generate and weight a particle set,which means N discrete hypothesis evently on map.

$$S0=(S_0^0, S_0^1, S_0^2, \ldots\ldots, S_0^{N-1})$$

Step 2: Resampling: Resample Nparticles from a particle set t using weights of respective particles and allocate them on the map. (We allow resampling of more than that of the same particles.).

Step 3: Prediction. Predict the next state of a particle set St from Newton's motion equation.

Step 4: Find the new document vector coordinates in this reduced 2-dimensional space.

Step 5: Weighting Re-calculate the weight of St.

Step 6: Measurement: Calculate the current object location o(xt, yt) by the average of s(xt, yt) $\in$ St.

Step 7: Iteration: Iterate Step 3, 4, 5 and 6 until convergence.

IV. CONCLUSION

Recently, Twitter, a popular micro blogging service, has become a new information channel for users to receive and to exchange information. Everyday, nearly 170 million tweets are created and redistributed by millions of active users. People from different places may tweet about the same event (e.g, “Norway shooting”) or the same type of event (e.g, “earthquake”); by collecting tweets over time. We investigate the real-time interaction of events such as earthquakes in Twitter and propose an algorithm to monitor tweets and to detect a target event. First, to obtain tweets on the target event precisely, we apply semantic analysis of a tweet. For example, users

might make tweets such as “Earthquake!” or “Now it is shaking,” for which earthquake or shaking could be keywords,
but users might also make tweets such as “I am attending an Earthquake Conference,” or “Someone is shaking hands with my boss.” We prepare the training data and devise a classifier using a Support Vector Machine (SVM) based on features such as keywords in a tweet, the number of words, and the context of target-event words. After doing so, we obtain a probabilistic spatiotemporal model of an event. We then make a crucial assumption: each Twitter user is regarded as a sensor and each tweet as sensory information. Regarding each Twitter user as a sensor, the event-detection problem can be reduced to one of object detection and location estimation in a ubiquitous/ pervasive computing environment in which we have numerous location sensors: a user has a mobile device or
an active badge in an environment where sensors are placed. Through infrared communication or a WiFi signal, the user location is estimated as providing location-based services such as navigation and museum guides.

REFERENCES

[1] Mari, M. (2013). Infographic: Twitter The Fastest Growing Social Platform, retrieved April 19, 2013, from http://www.globalwebindex.net/twitter-thefastest-growing-social-platform-infographic/

[2] Holt, R. (2013). Twitter in numbers. , . retrieved April 19, 2013, from http://www.telegraph.co.uk/technology/twitter/9945505/Twitter-innumbers.html

[3] Smith, A., & Brenner, J. (2012). Twitter use 2012. Pew Internet & American Life Project.

[4] Merchant, R. M., Elmer, S., & Lurie, N. (2011). Integrating social media into emergency-preparedness efforts. New England Journal of Medicine, 365(4), 289-291.

[5] Analytics, P. (2009). Twitter Study–August 2009. San Antonio, TX: Pear Analytics. Available at: www. pearanalytics. com/blog/wpcontent/uploads/2010/05/Twitter-Study-August-2009. pdf.

[6] Mendoza, M., Poblete, B. and Castillo, C. (2010) Twitter Under Crisis: Can we trust what we RT?, Proceedings of the First Workshop on Social Media Analytics, pp. 71 – 79, ACM.

[7] Vieweg, S., Hughes, A.L., Starbird, K. and Palen, L. (2010) Micro-blogging during two Natural Hazards Events: What Twitter may Contribute to Situational Awareness, Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 1079 – 1088, ACM.

[8] Sakaki, T., Okazaki, M. and Matsuo, Y. (2010) Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors, Proceedings of the 19th International Conference on World Wide Web, pp. 851-860, ACM.

[9] Heverin, T. and Zach, L. (2010) Micro-blogging for Crisis Communication: Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma, Washington, Area, Proceedings of the 7th International ISCRAM Conference, Seattle.

[10] Abel, F., Hauff, C., Houben, G.-J., Stronkman, R. and Tao, K. (2012) Semantics + Filtering + Search = Twitcident. Exploring Information in Social Web Streams, Proceedings of the 23rd Conference on Hypertext and Social Media, pp. 285 – 294, ACM.

[11] Terpstra, T. and Stronkman, R. and de Vries, A. and Paradies, GL. (2012) Towards a realtime Twitter analysis during crises for operational crisis management, Proceedings of the 9th International ISCRAM Conference, Vancouver, Canada.

Hemantha. Mareedu, M.Tech (CSE), NRI Institute of Technology, A.P., India.

Ella. Karunakar, Asst Professor, Dept. of Computer Science & Engineering, NRI Institute of Technology, A.P., India