ISSN 23499842(Online), Volume 1, Special Issue 2(ICITET 15), March 2015
International Journal of Innovative Trends and Emerging Technologies

AUTOMATIC EXTRACTION OF SOCIAL INTERFERENCE SYSTEM USING BIG DATA ANALYSIS WITH EMERGENCY ALERT

M.SHARMILA AFIREEN [1] K.RAJESH KAMBATTAN [2]

[1] PG Student, Department of Computer Science and Engineering
Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India
[2] Assistant Professor, Department of Computer Science and Engineering
Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India
[1] sharmilaafrin@gmail.com
[2] rajesh.kambattan@gmail.com

***Abstract*—Emergencies often involve escalating and evolving events that demand high performance and flexibility from the systems. Message prioritization, automation of communication, fast message delivery, and other capabilities are often required by each unique emergency situation. The traditional communication channel lack proper alert system to report about natural disaster such as the earthquake, Tsunami, and typhoon to the people. Thus, An automatic approach to Twitter demonstrated its value as a viable substitute to traditional communication channels. In this strategy, particle filter technique is used to extract the important keywords from tweets using stemming algorithm and genia tagger. Each twitter user will consider as a sensor and apply particle filtering to find location and time estimation. If the system interfaces reaches the maximum peak of the particular keyword like "Earthquake / Typhoon / Tsunami" at a particular time and at particular location, then a peak is generated immediately and an auto alert notification send to the nearest rescue team and the nearest people residing in that locality. Hence by using emergency alert, SMS and email alert will be automatically send for the registered tweet users as well as to the nearest rescue team. The disaster and notification are delivered faster than broadcast announcements. It is less time consuming process.**

Keywords— Twitter, Sensor, earthquake, location estimation, micro blogging .

I. INTRODUCTION

Twitter, a popular microblogging service, has become a new information channel for users to receive and to exchange information. It is an online social network used by millions of people around the world to stay connected to their friends, family members and co-workers through their computers and mobile phones [1].Everyday, nearly 170 million tweets are created and redistributed by millions of active users. Twitter has several unique advantages that distinguish it from news web sites, blogs, or other information channels. First, tweets are created in real-time. The tweets are in 140-character-message limit and the popularity of Twitter's mobile applications; users tweet and retweet instantly [2]. For example, we could detect a tweet related to a shooting crime 10 minutes after shots fired, while the first news report appeared approximately 3 hours later.

An important common characteristic among micro blogging services is its real-time nature. Although blog users typically update their blogs once every several days, Twitter users write tweets several times in a single day. Users can know how other users are doing and often what they are thinking about now, users repeatedly return to the site and check to see what other people are doing. The large number of updates results in numerous reports related to events. They include social events such as parties, baseball games, and presidential campaigns. They also include disastrous events such as storm, fire, traffic jam, riots, heavy rainfall, and earthquakes. Actually, Twitter is used for various real times notification such as that necessary for help during a large-scale fire emergency and live traffic updates. This work 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; and developed an earthquake reporting system that extracts earthquakes from Twitter and sends an alert message to registered users as well as to the rescue team. Here, we explain our methods using an earthquake as a target event.

A. The contributions of the paper are summarized as follows:

- The paper provides an example of integration of semantic analysis and real-time nature of Twitter, and presents potential uses for Twitter data.
- For earthquake prediction and early warning, many studies have been made in the seismology field. This paper presents an

innovative social approach, which has not been reported before in the literature.

II. EXISTING WORKS

Earthquake is a disastrous event in which many people loss their life and property. Hence detection and bringing awareness about the event is very important to prevent the hazardous environment. Daily newspaper, TV broadcast channels paves way for it. But these systems are slower when it comes for the real time occurrence as they are time consuming in sensing, reporting or publishing. Also in today's running world people does not pay much attention all time in media.

Proposed system provides social networks as they have lots of followers as microblogging attraction bloomed over them. These social networks served to be both fun time and useful .It helped to bridge the gap between the friends of different community and get connected all over the world[4][3].At the same time some of the facts and important messages were also known by the people by their friends and followers updates, shares and likes. Their like gives rating to some of the ads or some blogs or groups. Considering these facts they decided to develop the reporting system for earth quake. They choose twitter because of its popularity .They considers users as the social sensors. These users posts report on earth quake such as "I am attending Earthquake right now" or "I happened to feel the shaking of the earth". These tweets are analyzed based on features such as statically [2][3]. They separate keywords and words and confirm the event and sent to positive class. At time there may be tweets such as "I read an article on earthquake" or the tweets such as "Earthquake before three days was heart stolen. These tweets were analyzed and found that it is not under the criteria and hence sent to the negative class.

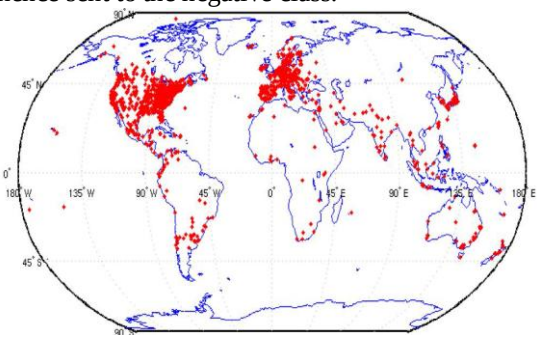

Fig 1. Twitter user map

III. PROPOSED WORK

This paper proposes an algorithm to monitor tweets and to detect a target event and producing a probabilistic spatiotemporal model for target event that can find the center of event location. The proposed scheme uses a particle filter for location estimation. It can be achieved by regarding each user as a sensor, and analyzes a spatial and temporal pattern of an event and applies particle filtering which used for location estimation.

We manually give a set of queries for the target event. For example, we will search for "earthquakes" and "shaking" if our target event is earthquakes. By searching every second, we can obtain a subset of tweet that all contains the keyword in our searching queries. Then, for each tweet in our searching result, obtain semantic features as its feature vector, then apply classifier to the tweet and get a value positive and negative.

A. *Application Creation*

While creating the application, we'll assign the design fields like username, password, phone and other information. The user will enter the tweets through this application.

B. Server

The Server will analyze the user's contents. So, the server will extract the keywords from the tweets. Also the server will be retrieving the user information like access time and location which is used to find the user's location.

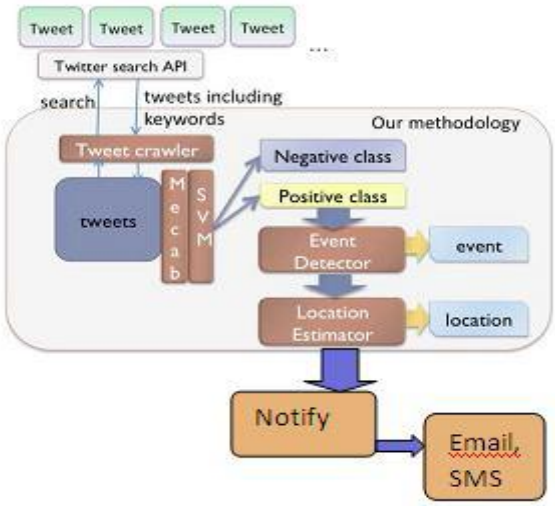

Fig 2. Proposed System Architecture

C. Extracting the Keyword Using Particle Filter

Successful creation of an application, the server will analyze the tweets and extract the keywords using Particle Filter. The Particle Filter will extract the keywords and filter the other words using the Stemming

D. Automatic Alert to Rescue Team

After extracting the keywords, we send the SMS alert and Email to the rescue team once we attains the Maximum Peak of the extracted Keyword.

IV. PROPOSED MODEL EVALUATION

In order for event detection and location estimation, we use probabilistic models. In this section, we first describe event detection from time-series data. Then, we describe the location estimation of a target event.

A. *Event Detection*

An event is an arbitrary classification of a space/time region. An event might have actively participating agents, passive factors, products, and a location in space/time. We target events such as earthquakes, tsunami, and traffic jams, which are visible through tweets.

B. *Semantic Analysis*

To detect a target event from Twitter, we search from Twitter and find useful tweets. target event precisely, we a tweet: For example, users might make tweets such as "Earthquake!" thus earthquake could be keywords, but users might also make tweets such as "I am attending an Earthquake Conference". Moreover, even though a tweet refers to the target event, it might not be appropriate as an event report; for example a user makes tweets such as "The earthquake yesterday was scaring", or "Three earthquakes in four days. Japan scares me." These tweets are truly the mentions of the target event, but they are not real-time reports of the events. Therefore, it is necessary to clarify that a tweet is actually referring to an actual earthquake occurrence, which is denoted as a positive class. By preparing positive and negative examples as a training set, we use a SVM to classify tweets automatically into positive and negative categories [5].

C. *Temporal Model*

Each tweet has its post time. When a target event occurs, how can the sensors detect the event? We describe the temporal model of event detection. First, we examine the actual data.

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.

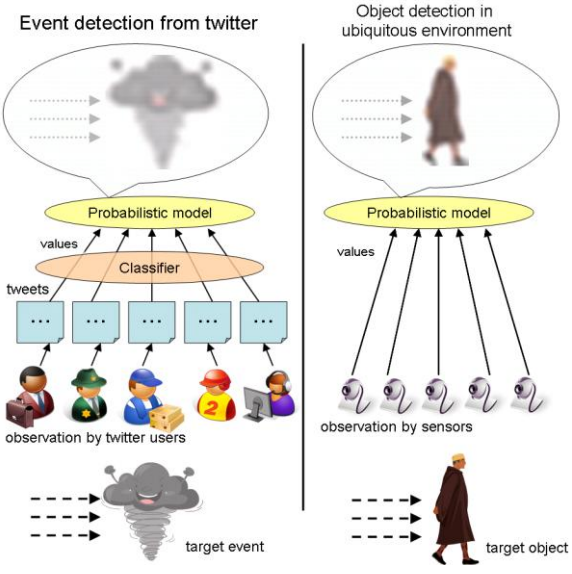

Fig 3. Correspondence between event detection from Twitter and object detection in a ubiquitous environment.

a) Particle filter Algorithm

1. Initialization: Calculate the weight distribution D w(x, y) from twitter users geographic distribution.
2. Generation: Generate and weight a particle set, which means N discrete hypothesis.
3. Re-sampling: Re-sample N particles from a particle set St using weights of each particles and allocate them on the map.
4. Prediction: Predict the next state of a particle set St from the Newton's motion equation.
5. Weighing: Re-calculate the weight of St by measurement m(mx, my) .
6. Measurement: Calculate the current object location
   o(xt, yt) by the average of s(xt, yt) ∈ St.

D. *Spatial Model*

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.

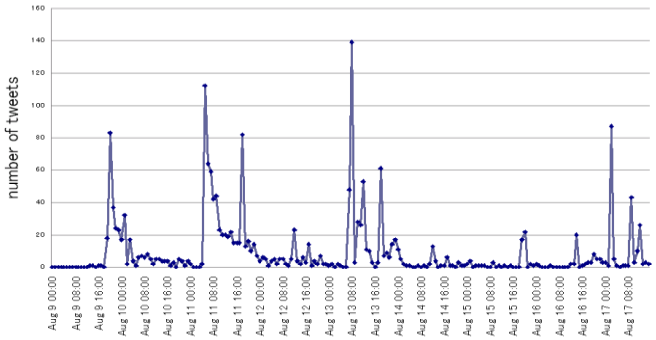

Fig 4. Number of tweets related to earthquakes

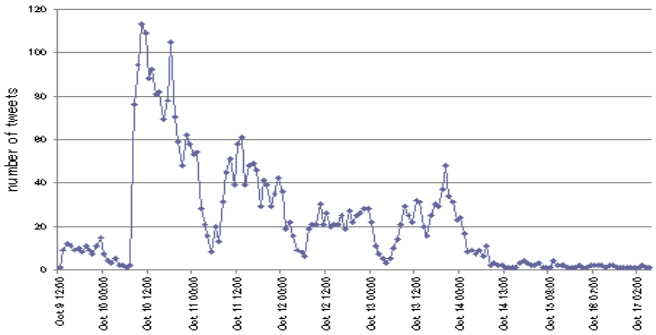

Fig 5. Number of tweets related to typhoons

a) Event detection and location estimation Algorithm

1. Given a set of queries Q for a target event.

2. Put a query Q using search API every s seconds and obtain tweets T.
3. For each tweet t ∈ T, obtain features A, B, and C. Apply the classification to obtain value vt = {0, 1}.
4. Calculate event occurrence probability p occur using vt, t ∈ T; if it is above the threshold p there occur, then proceed to step 5.
5. For each tweet t ∈ T, we obtain the latitude and the longitude lt by i) utilizing the associated GPS location, ii) making a query to Google Map the registered location for user ut. Set lt = null if both do not work.
6. Calculate the estimated location of the event from lt, t ∈ T using Kalman filtering or particle filtering.
7. (optionally) Send alert e-mails to registered users.

V. CONCLUSION AND FUTURE WORKS

As we discussed in this paper, Twitter user is used as a sensor, and set the problem as detection of an event based on sensory observations. Location estimation methods such as particle filtering are used to estimate the locations of events. As an application, we developed an earthquake reporting system.

Microblogging helps us to spread the news at faster rates to the other subscribers of the page, it also distinguishes it from other media in the form of blogs and collaborative bookmarks. The algorithm used in the proposed system helps us to identify the keyword that is required help us to gather more information from various places.

Finding the people of specific areas through particle filtering and sending alarms to the people is a new concept of intimating people which will pay a way for spreading the news at faster rate rather than intimating the people in the whole world and spreading the news at slower rate.

REFERENCES

[1]. S. Milstein, A. Chowdhury, G. Hochmuth, B. Lorica, and R. Magoulas. Twitter and the micro-messaging revolution: Communication, connections, and immediacy.140 characters at a time. O'Reilly Media, 2008

[2]. V. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello,"Bayesian Filtering for Location Estimation," IEEE Pervasive Computing, vol. 2, no. 3, pp.24-33, July-Sept. 2003.

[3]. T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors," Proc. 19th Int'l Conf. World Wide Web (WWW '10), pp. 851-860, 2010.

[4]. M. Cataldi, L. Di Caro, and C. Schifanella, "Emerging Topic Detection on Twitter Based on Temporal and Social Terms Evaluation," Proc.10th Int'l Workshop Multimedia Data Mining (MDMKDD '10), pp. 1-10, 2010.

[5]. M. Ebner and M. Schiefner, "Microblogging - More than Fun?"Proc. IADIS Mobile Learning Conf., pp. 155-159, 2008.

[6]. L. Backstrom, J. Kleinberg, R. Kumar, and J. Novak, "Spatial Variation in Search Engine Queries," Proc. 17th Int'l Conf.World Wide Web (WWW '08), pp. 357-366, 2008.

[7]. T. Joachims. Text categorization with support vector machines. In Proc. ECML'98, pages 137–142, 1998Xiaoyan Zhu, Shunrong Jiang, Liangmin Wang, and Hui Li, "Efficient Privacy-Preserving Authentication for Vehicular Ad Hoc Networks," VOL. 63, NO. 2, 2014

[8]. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. IEEE Transactions on Signal Processing, 2001.

[9]. L. Backstrom, J. Kleinberg, R. Kumar, and J. Novak. Spatial variation in search engine queries. In Proc. WWW2008, 2008.

[10]. T. Bleier and F. Freund. Earthquake warning system. Spectrum, IEEE, 2005.

[11]. K. Borau, C. Ullrich, J. Feng, and R. Shen. Microblogging for language learning: Using twitter to train communicative and cultural competence. In Proc. ICWL 2009, pages 78–87, 2009.

[12]. d. boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In Proc. HICSS-43, 2010.

[13]. Y. Matsuo and H. Yamamoto. Community gravity: measuring bidirectional effects by trust and rating on online social networks. In Proc. WWW2009, 2009.

[14]. Q. Mei, C. Liu, H. Su, and C. Zhai. A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In Proc. WWW'06, 2006.

[15]. S. Milstein, A. Chowdhury, G. Hochmuth, B. Lorica,and R. Magoulas. Twitter and the micro-messaging revolution: Communication,connections, and immediacy.140 characters at a time. O'Reilly Media,2008.