Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 6, June 2014, pg.320 – 327

RESEARCH ARTICLE

Earthquake Reporting System by Using Real Time Nature of Twitter

SK.Haseena, P.Sandeep Reddy, Dr. M.V Siva Prasad

Dept. of CSE & Anurag Engineering College, Kodad, India
haseenask.h@gmail.com, sandeepcse@anurag.ac.in, principal@anurag.ac.in

Abstract -TWITTER, a popular microblogging service, an important characteristic of Twitter is its real-time nature. We analyze the real-time interaction of events such as earthquakes in Twitter and propose an algorithm to monitor tweets and to detect a target event. To detect a target event, we devise a classifier of tweets based on features such as the keywords in a tweet, the number of words, and their context. Subsequently, we propose a probabilistic spatiotemporal model for the target event that can find the center of the event location. We regard each Twitter user as a sensor and apply particle filtering, which are widely used for location estimation. The particle filter works better than other comparable methods for estimating the locations of target events.

Keywords— Twitter, event detection, social sensor, location estimation, earthquake

I. INTRODUCTION

TWITTER, a popular microblogging service, has received much attention recently. This online social network is bused by millions of people around the world to remain socially connected to their friends, family members, and coworkers through their computers and mobile phones. Twitter, has seen a lot of growth since its launch on July 2006, Twitter users have increased rapidly. The number of registered Twitter users exceeded 100

© 2014, IJCSMC All Rights Reserved

million in April 2010. The service is still adding about 300,000 users per day.1 Currently, 190 million users use Twitter per month, generating 65 million tweets per day. Microblogging is a new form of communication in which users can describe their current status in short posts in the form of text updates or macromedias such as photographs or audio clips distributed by instant messages, mobile phones, email or the Web. Twitter - a microblogging service that enables users to post messages ("tweets") of up to 140 characters - supports a variety of communicative practices; participants use Twitter to converse with individuals, groups, and the public at large, so when conversations emerge, they are using service that enables them to post messages ("tweets") of length up to 140 characters - supports a variety of communicative practices; participants use Twitter to converse with individuals, groups, and the public at large, so when conversations emerge, they are often experienced by broader audiences than just the interlocutors.

Our study, which is based on the real-time nature of one social networking service, is applicable to other micro blogging services, but we specifically examine Twitter in this study because of its popularity and data volume. An important characteristic that is common among micro blogging services is their 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. Several important instances exemplify their real-time nature: in the case of an extremely strong earthquake in Haiti, many pictures were transmitted through Twitter. People were thereby able to know the circumstances of damage in Haiti immediately. The contributions of this 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,

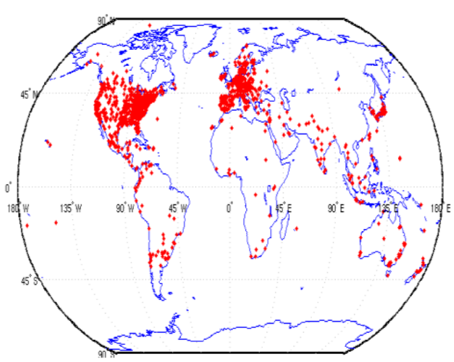

Fig. 1. Twitter user map

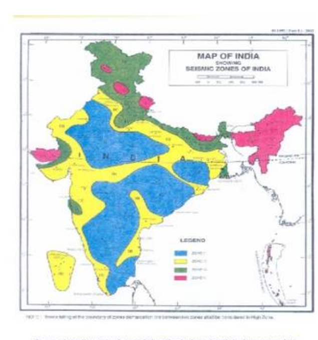

Fig.2.Earthquake map of India

II. INVESTIGATION

We choose earthquakes as target events. Seismic observations are conducted worldwide, which facilitates acquisition of earthquake information, which also makes it easy to validate the accuracy of our event detection methodology. fig 1&2 representing Twitter user map and Earthquake map of India.

© 2014, IJCSMC All Rights Reserved

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. The target events such as earthquakes, typhoons, and traffic jams, which are readily apparent upon examination of tweets. These events have several properties.

1. They are of large scale.
2. They particularly influence the daily life of many people.
3. They have both spatial and temporal regions.

Such events include social events such as large parties, sports events, exhibitions, accidents, and political campaigns. They also include natural events such as storms, heavy rains, tornadoes, typhoons/hurricanes/cyclones, and earthquakes. We designate an event we would like to detect using Twitter as a target event.

SEMANTIC ANALYSIS OF TWEETS

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 Fig. 3. Tweets might include mention of the target event. For example, users might make tweets such as “Earthquake!” or “Now it is shaking.” Consequently, earthquake or shaking might be keywords (which we call query words). However, users might also make tweets such as “I am attending an Earthquake Conference.” or “Someone is shaking hands with my boss.” Moreover, even if a tweet is referring to the target event, it might not be appropriate as an event report. For instance, a user makes tweets such as “The earthquake yesterday was scary.” or “Three earthquakes in four days. Japan scares me.” These tweets are truly descriptions of the target event, but they are not real-time reports of the events. Therefore, it is necessary to clarify that a tweet is truly referring to an actual contemporaneous earthquake occurrence, which is denoted as a positive class. To classify a tweet as a positive class or a negative class, we use a support vector machine 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.

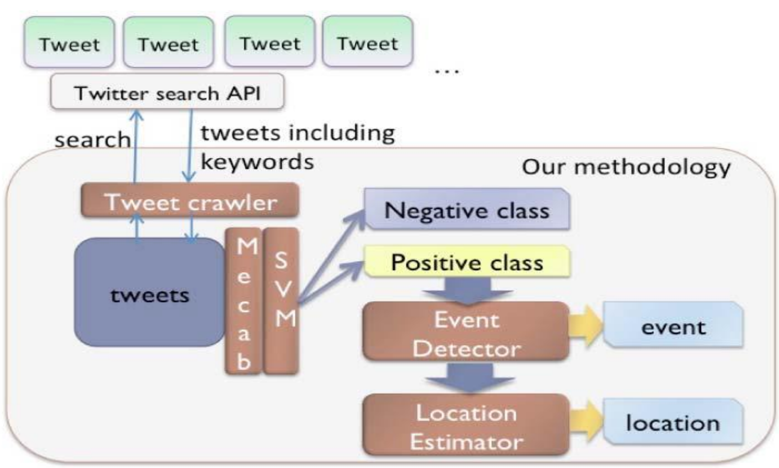

Fig. 3. Method to acquire tweets referred to a target event precisely.

SVM Features of an Example Sentence

Feature Name	Features
Features A	7 words, the fifth word
Features B	I, am, in, Japan, earthquake, right, now
Features C	Japan, right

We prepare three groups of features for each tweet as described below.

. Features A (statistical features): the number of words in a tweet message, and the position of the query word within a tweet.

. Features B (keyword features): the words in a tweet.6

. Features C (word context features): the words before and after the query word.

TWEET AS A SENSORY VALUE

We can search the tweet and classify it into a positive class if a user makes a tweet about a target event. In other words, the user functions as a sensor of the event. A tweet can therefore be regarded as a sensor reading. This crucial assumption enables application of various methods related to sensory information.

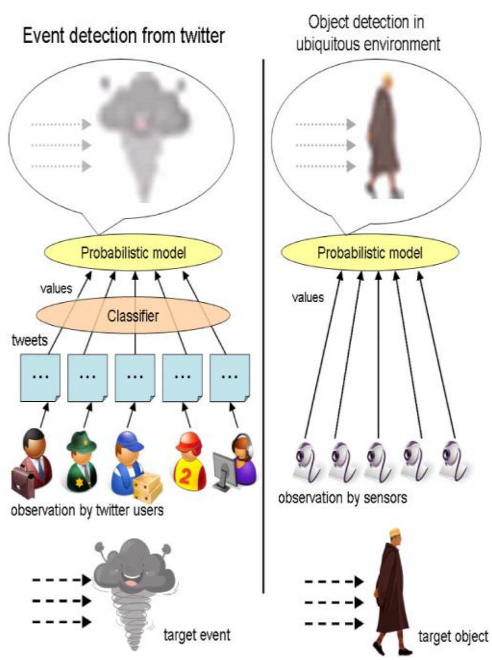

Fig. 4. Correspondence between event detection from Twitter and object detection in aubiquitous environment.

Fig. 4 presents an illustration of the correspondence between sensory data detection and tweet processing. The motivations are the same for both cases: to detect a target event. Observation by sensors corresponds to an observation by Twitter users. They are converted into values using a classifier. The virtual sensors (or social sensors) have various characteristics: some sensors are activated) only by specific events, although others are activated by a wider range of events. The sensors are vastly numerous: there are more than 100 million 'Twitter sensors' worldwide producing tweet information around the clock. A sensor might be inoperable or operating incorrectly sometimes (which means a user is not online, sleeping, or is busy doing something else). For that reason, this social sensor is noisier than ordinary physical sensors such as location sensors, thermal sensors, and motion sensors. Therefore, a probabilistic model is necessary to detect an event, as described in the next section. A tweet can be associated with a time and location: each tweet has its post time, which is obtainable using a searchAPI. In fact, GPS data are attached to a tweet sometimes, such as when a user is using an iPhone. Alternatively, each Twitter user makes a registration on their location in the user profile. The registered location might not be the current location of a tweet. However, we infer it that a Person is probably near the registered location. Some tweets include place names in those bodies. Some researchers describe their efforts to extract place names from tweets as a part of Named Entity Recognition.

III. MODEL

We propose a probabilistic spatiotemporal model for the target event that can find the center of the event location.

TEMPORAL MODEL

Each tweet has its own post time. When a target event occurs, how do the sensors detect the event? We describe the temporal model of event detection.

In the Twitter case, we can infer that if a user detects an event at time 0, then we can assume that the probability of his posting a tweet from t to _t is fixed as . Then, the time
to produce a tweet can be regarded as having an exponential distribution. Therefore, even if a user detects an event, she might not make a tweet immediately if she is not online or if she is doing something else. She might make a post only after such problems are resolved. Therefore, it is reasonable that the distribution of the number of tweets follows an exponential distribution.

SPATIAL MODEL

Each tweet is associated with a location. We describe a method that can estimate the location of an event from sensor readings.

Particle Filters

A particle filter is a probabilistic approximation algorithm implementing a Bayes filter, and a member of the family of sequential Monte Carlo methods. For location estimation, it maintains a probability distribution for the location estimation at time t, designated as the belief
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.
1. Generation. Generate and weight a particle set, which means N discrete hypothesis

$$S_0 = \left(s_0^0, s_0^1, s_0^2, \ldots, s_0^{N-1}\right),$$

© 2014, IJCSMC All Rights Reserved

and allocate them evenly on the map:

$$particle\ s_0^k = (x_0^k, y_0^k, w_0^k)$$
$$x : longitude, y : latitude, w : weight.$$

2. Resampling. Resample N particles from a particle set St using weights of respective particles and allocate them on the map.

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

$$\begin{aligned}(x_t^k, y_t^k) &= \Big(x_{t-1}^k + v_{x_{t-1}}\Delta t + \frac{a_{x_{t-1}}}{2}\Delta t^2, \\ &\qquad y_{t-1}^k + v_{y_{t-1}}\Delta t + \frac{a_{y_{t-1}}}{2}\Delta t^2\Big) \\ (v_{x_t}, v_{y_t}) &= (v_{x_{t-1}} + a_{x_{t-1}}, v_{y_{t-1}}, a_{y_{t-1}}) \\ a_{x_t} &= \mathcal{N}(0;\sigma^2), \quad a_{y_t} = \mathcal{N}(0;\sigma^2).\end{aligned}$$

4. Weighing. Recalculate the weight of St by measurement m(m $_{x,}$m $_{y)}$ as follows:

$$\begin{aligned}dx_t^k &= m_x - x_t^k, \quad dy_t^k = m_y - y_t^k \\ w_t^k &= \frac{1}{(\sqrt{2\pi}\sigma)} \\ &\quad \cdot exp\left(-\frac{(dx_t^{k2} + dy_t^{k2})}{2\sigma^2}\right).\end{aligned}$$

5. Measurement. Calculate the current object location $o(x_t,y_t)$by the average of $s(x_t,y_t)€\ s_t$

6. Iteration. Iterate Steps 2, 3, 4, and 5 until convergence.

IV. SYSTEM DESIGN

By using real time nature of tweets and semantic analysis over the tweets we propose a system.

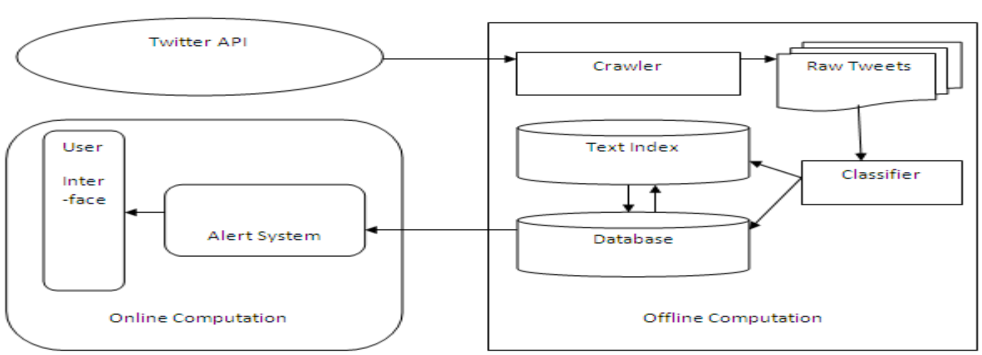

Fig.5.system architecture

© 2014, IJCSMC All Rights Reserved

Fig 5 is representing the architecture of the proposed system which detects the event notification and reports the event to the users.

Tweet collection

It is necessary to collect tweets referring to an earthquake from Twitter. This process includes two steps: crawling tweets from Twitter and filtering out tweets that do not refer to the earthquake. For crawling and filtering tweets, we recommend using script programming languages.

Crawling tweets from Twitter

To collect tweets or some user information from Twitter, one must use the Twitter Application Programmers Interface (API). Twitter API is a group of commands that are necessary to extract data from Twitter. Twitter has APIs of three kinds: Search API, REST API, and Streaming API. Search API and Streaming API, which are necessary to crawl tweets from Twitter. REST API is necessary to extract location information from Twitter information. Additionally, it is known that Twitter API specifications are subject to change. When using Twitter API, it is necessary to know the latest details and requirements. They are obtainable from Twitter API documentation

Twitter Search API

The Twitter Search API extracts tweets from Twitter, including search keywords or those fitting other retrieval conditions, in chronological order. It is possible to use language, date, location and other conditions as retrieval conditions. Some points must be considered when using Twitter Search API:

• It is possible to collect tweets posted only during the prior five days. It is not possible to search tweets posted six days ago.

• It is only possible to collect the latest 1500 tweets at one time.

Filtering tweets using machine learning

We collected data from tweets including keywords related to earthquakes, such as earthquake, shake. Those tweets include not only tweets that users posted immediately after they felt earthquakes, but also tweets that users posted shortly after they heard earthquake news. we must filter tweets to extract those posted immediately after the earthquake. We designate tweets posted by users who felt earthquakes as positive tweets, and other tweets as negative tweets. Here, we describe the creation of a classifier to categorize crawled tweets into positive tweets and negative tweets, using Support Vector Machine: a supervised learning method.

© 2014, IJCSMC All Rights Reserved

Semantic Analysis on Tweets

Semantic Analysis on Tweet Search tweets including keywords related to a target event Example: In the case of earthquakes "shaking", "earthquake" Classify tweets into a positive class or a negative class Example: "Earthquake right now!!" ---positive "Someone is shaking hands with my boss" --- negative Create a classifier

Earthquake Reporting System

The users will be altered if the earthquake occurs based on their location and the tweets. Effectiveness of alerts of this system Alert E-mails urges users to prepare for the earthquake if they are received by a user shortly.

V. CONCLUSION

As described in this paper, we analyze the real-time nature of Twitter, devoting particular attention to event detection. Semantic analyses were applied to tweets to classify them into a positive and a negative class. We regard each Twitter user 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 and to propose earthquake reporting system.

REFERENCES

1. A. Java, X. Song, T. Finin, and B. Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities," Proc. Ninth WebKDD and First SNA-KDD Workshop Web Mining and Social Network Analysis (WebKDD/SNA-KDD '07), pp. 56-65, 2007.

2. B. Huberman, D. Romero, and F. Wu, "Social Networks that Matter: Twitter Under the Microscope," ArXiv E-Prints, http:// arxiv.org/abs/0812.1045, Dec. 2008.

3. G.L. Danah Boyd and S. Golder, "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter," Proc. 43rdHawaii Int'l Conf. System Sciences (HICSS-43), 2010.

© 2014, IJCSMC All Rights Reserved