ISSN: 2321-7782 (Online)
Volume 2, Issue 11, November 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Social Networking for Real Time Earthquake Detection and Reporting System

Swapnali V Kulkarni[1]
Department of computer Engineering.
Vishwabharati Academy's College of Engineering
Ahmednagar, India

Prof Jayapal PC[2]
Department of computer Engineering.
Vishwabharati Academy's College of Engineering
Ahmednagar, India

Priyanka D Navgire[3]
Department of computer Engineering.
Vishwabharati Academy's College of Engineering
Ahmednagar, India

Abstract: Now a days large number of people are moving towards the social networking. Twitter is one of the best examples of social networking. Twitter is used by millions of people in the country. Twitter is generally categorized as a micro-blogging service, which enables users to send brief text updates. These tweets are processed through twitter search API and crawler and used in real time event detection system such as an earthquake.

In proposed system semantic analysis is used to analyze the tweets. Classification of tweets is done by Support Vector Machine algorithm. Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection. The occurrence of event detection is done by spatiotemporal model by treating each twitter user as a sensor value. For the purpose of location estimation, an algorithm is used which is based on the tweet content similarity and the past similarity of twitter user.

Keywords: Twitter, Earthquake, Event detection, Probabilistic model, sensor

I. INTRODUCTION

A **social networking service** (also called **SNS**) is a very good platform to build social networks or social relations among people who share their interests, activities or real-life connections. Social networks are web-based services that allow individuals to create a public profile, to create a list of all users with whom we want to share connections, and view and cross the connections within the system. Most social network services are web-based services and allow users to interact over the Internet, such as through e-mail and instant messaging.

Twitter is one of the widely used social networking services that allow you to answer the question, "What are you doing now?". This can be done by sending short text messages to your friends, or "followers." [11]. These short massages are called as "tweets". Twitter provides to send the message up to 140 characters. Registered users can read as well as post tweets, but unregistered users can only read tweets. Users access Twitter through the website interface, SMS, or mobile device app. Twitter Inc. is based in San Francisco and has more than 25 offices around the world.

We generally categorized twitter as a microblogging service [1]. A microblog differs from a traditional blog in that its content is typically smaller in both actual and aggregated file size. Generally microblogs "allow users to share small elements of content such as short sentences, small messages, individual images, or video links". These small messages are sometimes known as *micro-posts.* An important characteristic that is common among microblogging services is their real-time nature. Although blog users typically update their blogs once every several days, Twitter users may write tweets several times per day. Users can know what 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.

As an application, we develop an earthquake reporting system using tweets. Our system detects an earthquake occurrence and sends an e-mail, possibly before an earthquake actually arrives at a certain location: An earthquake propagates at about 3-7 km/s. For that reason, a person who is 100 km distant from an earthquake is able to communicate and act for about 20 s before the arrival of an earthquake wave.

This paper is organized as described below. In the next section, we explain the proposed system and its structure Section 3 presents our explanation of Twitter search API. Query pre-processing is explained in section 4. Section 5 presents our explanation of semantic analysis and sensory information with subsequent the Probabilistic model for Location estimation in Section 6. Section 7 contains an earthquake reporting system.

II. PROPOSED SYSTEM

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. This event detection system is displayed in fig 1.

Users can see the detection of past earthquakes. They can register their e-mails to receive notices of future earthquake detection reports. Given system alerts users and urges them to prepare for the imminent earthquake. It is hoped that a user receives the e-mail before the earthquake actually affects that area.

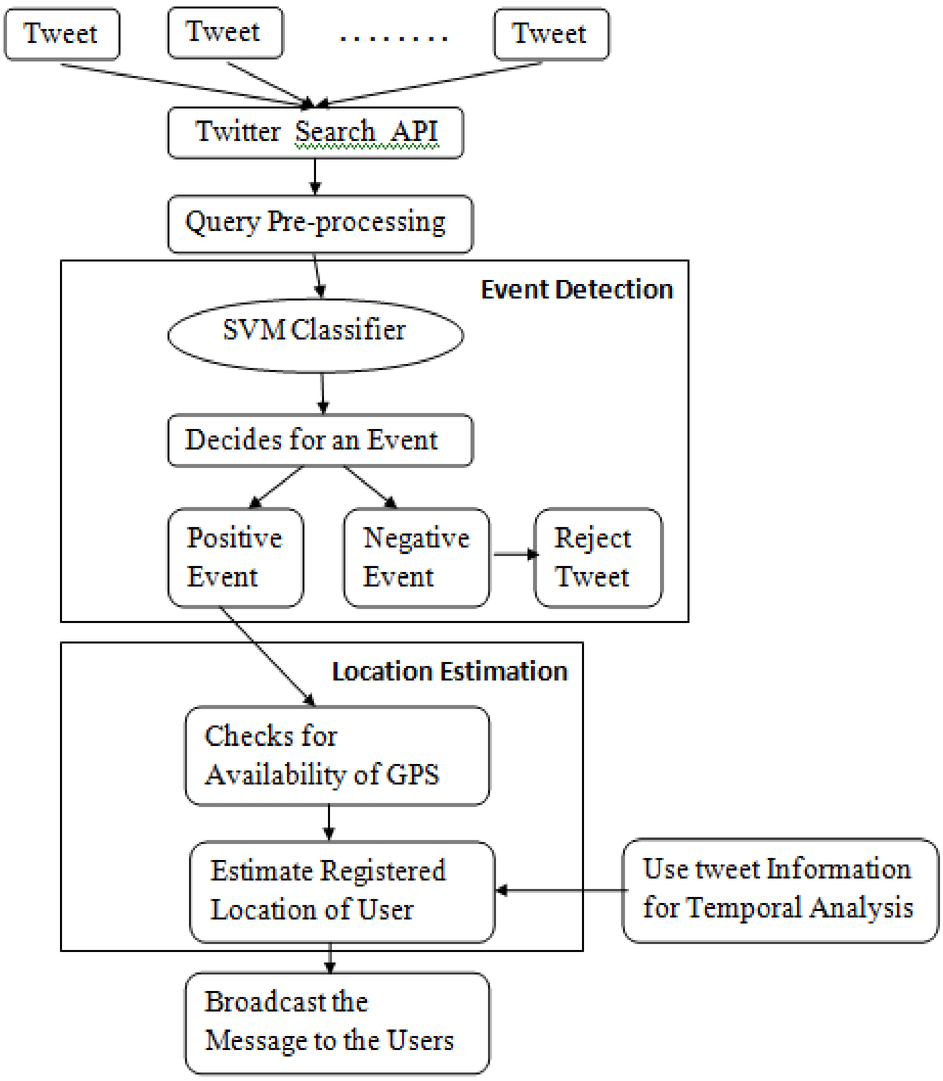

Fig 1. Method to acquire tweets referred to a target event

III. TWITTER SEARCH API

Twitter search API allows user defined queries against the indices of recent or popular Tweets. IT behaves similarly to, but not exactly like the Search feature available in Twitter mobile or web clients. Search API is basically concentrates on relevance and not completeness. This means that some Tweets and users may be missing from search results. The Search API is an interface to this search engine. Our search service is not meant to be an exhaustive archive of public tweets and not all tweets are indexed or returned. Some results are refined to better combat spam and increase relevance. Due to capacity constraints, the index currently only covers about a week's worth of tweets.

IV. QUERY PRE- PROCESSING

Query preprocessing is a technique that prepares the query for optimization purpose. It may change the representation of statement such that the SQL statement Component Integration Services generates is syntactically different from the original statement. Preprocessing performs view expansion of query, so that a query can operate on tables referenced by the view. It also takes steps such as re-ordering expressions and transforming sub-queries to improve processing efficiency. With the help of query pre-processing, we can reduce the time required to extract the queries from the database. Here we are using query pre-processing to extract the tweets more efficiently from twitter database. Hence the query pre-processing technique plays a vital role in the real time system.

V. EVENT DETECTION

In this paper our objective is to detect the event. An event might be anything that happens in the world [2]. Events usually refer to something abnormal, that is, something that rarely happens in normal situation. Event detection aims to find such abnormal phenomenon from collected data. We use twitter to detect the target events.

Semantic Analysis of Tweets

We apply search on twitter and find useful tweets related to our target event. Tweets may include information about target event. For example, user might make tweet as "Now earthquake is happening" or "Now it is shaking". Here earthquake and shaking may be a keyword. However user might make the tweet as "I am watching movie of earthquake" or "Someone is shaking hands with my boss".

Although the tweet refers to the target event, it may not appropriate as an event report. For instance, "Earthquake happened in the last year was dangerous" or "It is an earthquake zone". Such tweets are truly descriptions of the target event, but they are not real-time reports of the events. Therefore it is important to classify the tweets into the classes. One that refers to the actual event occurrence, which we denote as a positive class and remaining as a negative class.

To classify a tweet as a positive class or a negative class, we use a support vector machine as a classifier. SVM is a widely used machine-learning algorithm. These methods are used for *classification, regression* and *outliers detection* [3]. Support vector machines are based on the "Structural Risk Minimization principle" from computational learning theory. The idea of structural risk minimization is to find a hypothesis *h* for which we can guarantee the lowest true error. The true error of h is the probability that *h* will make an error on an unseen and randomly selected test example. Support Vector Machine contains two phases:

1. Training phase
2. Testing phase.

The advantages of support vector machines are:

1. SVMs are Effective in high dimensional spaces.

2. Effective in cases where number of dimensions is generally greater than the number of samples.
3. Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
4. Versatile: different *Kernel functions* can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

Tweet as a Sensory Value

We can detect and classify the tweets into the positive class if user makes the tweet about target event. i.e. user acts as a sensor for target event. If user makes the tweet about an earthquake, then it can consider as he is an "Earthquake sensor" and it returns the positive value. In this paper two assumptions are made. Some researchers describe their efforts to extract place names from tweets as a part of Named Entity Recognition [4], [5].

Assumption 1: Each Twitter user is regarded as a sensor. A sensor detects a target event and makes a report about event probabilistically.

Assumption 2: Each tweet is associated with a particular time and location, which is a set of latitude and longitude co-ordinates.

VI. Probabilistic Model For Location Estimation

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) Temporal Model

Each tweet is associated with its post time. We use the temporal model of event detection. The distribution is apparently an exponential distribution. The probability density function of the exponential distribution is $f(t,\lambda) = \lambda e^{-\lambda t}$ where t>0 and λ>0. The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process. 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 Δt is fixed as λ. 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.

In this system it important to calculate the reliability of multiple sensor values. A user may produce a false alarm by writing a tweet about target event. It is also possible that the classifier may misclassify a tweet into a positive class. We can design the system probabilistically by considering the following two facts.

» The false-positive ratio P_f of a sensor.

» Sensors are assumed to be independent and identically Distributed (i.i.d.)

Assume that an n sensor produces a positive signal, the probability of all n sensors returning a false alarm is P_f^n . Therefore the probability of event occurrence is calculated as 1 - P_f^n .

b) Spatial Model

Each tweet is having its location i.e. from it is sent. In this paper we describe one method which is used to find the destination of user by using twitter post. The proposed method chooses base tweets, which is close to the current user's tweet, and then predict destination using the next set of tweets of base tweet [6]. This method only considers location transition information, but ignoring text data (user's tweets). We assume that if user's preference is different, the destination will change even if user's past location transition is similar. Therefore, in our approach, we consider the user's preference (captured from tweet content) to predict destination and verify if the accuracy of location prediction improves or not.

Proposed method for destination prediction:

Process will start when user u posts tweet u_i with GPS information. First, we choose "base tweet" among all the past tweets, which achieves the best match to the tweet u_i from the view point of its location and tweet content. The base-tweet is selected based on the two criteria:

1. Base user's past similarity to target user *u*.
2. Base tweet's current similarity to target user *u*'s tweet u_i .

According to the above criterion, we define the similarity between user *u*'s current tweet u_i and the base tweet c_j a

$$Similarity_{ui,cj} = \alpha\ PastSimilarity_{u,c} + (1- \alpha)\ CurrentSimilarity_{ui,cj}$$

Where, α is a weighting parameter.

Past similarity is calculated with the help of tweet similarity and location similarity between user *u* and base user *c.*

$$PastSimilarity_{u,c} = \beta\ TweetSimilarity_{u,c} + (1- \beta)\ LocationSimilarity_{u,c}$$

Where, β is a weighting parameter.

The high value of tweet similarity indicates that high similarity of the users in their interests and tendencies. High value of location similarity indicates that the target user and the base user share more information in where they live and where they usually go. In this approach, we introduced two parameters α and β, which means weight to consider past tweets and that to consider tweet content respectively. This method gives us the better result than that of the baseline method. As baseline method [7] β=0 (i.e. only location similarity is considered). This technology will bring benefits to various kinds of services, such as recommendations of real world contents and efficient navigation application.

VII. Earthquake Reporting System

We developed an earthquake-reporting system using the event detection algorithm. Earthquake information is very much valuable if it is received in real time [10]. Given some amount of advanced warning, any person would be able to turn off a domestically used gas or heater at home and then find protection under a desk or table if such a person were to have several seconds notice before an earthquake actually strikes an area. Large amount of work is done on seismology [8]. Also much more efforts have been taken to produce short-term forecasts to realize an earthquake warning system by observing electromagnetic emissions from ground-based sensors and satellites [9].

VIII. Conclusion

In this paper we have studied, the real time nature of twitter and used it in the event detection system. Semantic analysis is applied to the tweets to classify the tweets into the positive and negative class. Semantic analysis is used for detecting the target events. We regard each Twitter user as a sensor, and set the problem as detection of an event based on sensory observations. For location estimation a method is proposed based on tweet content similarity and the past similarity of tweet. As an application, we developed an earthquake reporting system, which is a best approach to notify people promptly of an earthquake event.

Acknowledgement

The heading of the Acknowledgment section and the References section must not be numbered.

Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to the top of file IEEETran.cls in the IEEE LaTeX distribution.

References

1. M. Sarah, C. Abdur, H. Gregor, L. Ben, and M. Roger, “Twitter and the Micro-Messaging Revolution,” technical report, O’Reilly Radar, 2008.
2. Y. Raimond and S. Abdallah, “The Event Ontology,” http:// motools.sf.net/event/event.html, 2007.
3. T. Joachims, “Text Categorization with Suport Vector Machines: Learning with Many Relevant Features,” Proc. 10th European Conf. Machine Learning (ECML ’98), pp. 137-142, 1998.
4. X. Liu, S. Zhang, F. Wei, and M. Zhou, “Recognizing Named Entities in Tweets,” Proc. 49th Ann. Meeting of the Assoc. for Computational Linguistics: Human Language Technologies (HLT ’11), pp. 359-367, June 2011.
5. A. Ritter, S. Clark Mausam, and O. Etzioni, “Named Entity Recognition in Tweets: An Experimental Study,” Proc. Conf. Empirical Methods in Natural Language Processing, 2011.
6. Takuya Shinmura, Dandan Zhu, Jun Ota, Yusuke Fukazawa “ *Destination Prediction Considering Both Tweet Contents and Location Transition Hitstory”,* ICMU, pp95-96, 2014.
7. Naoharu Yamada, Yoshinori Isoda, Masateru Minami and Hiroyuki Morikawa, “Incremental Route Refinement for GPS-enabled Cellular Phones”, ICMU, pp87-93, 2010.
8. E. Scordilis, C. Papazachos, G. Karakaisis, and V. Karakostas, “Accelerating Seismic Crustal Deformation before Strong Mainshocks in Adriatic and Its Importance for Earthquake Prediction,” J. Seismology, vol. 8, pp. 57-70, http://dx.doi.org/10.1023/ B:JOSE.0000009504.69449.48, 2004.
9. T. Bleier and F. Freund, “Earthquake [earthquake warning systems],” IEEE Spectrum, vol. 42, no. 12, pp. 22-27, Dec. 2005.
10. 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.
11. A. Tumasjan, T.O. Sprenger, P.G. Sandner, and I.M. Welpe, “Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment,” Proc. Fourth Int’l AAAI Conf. Weblogs and Social Media (ICWSM), 2010.

