International Journal On Engineering Technology and Sciences – IJETS™
ISSN(P): 2349-3968, ISSN (O): 2349-3976
Volume III, Issue II, February- 2016

EARTHQUAKE ANALYSIS SYSTEM IN TWITTER NETWORK

L.S.Maheswari
PG Scholar,
Department of computer applications,
Hindusthan College Of Arts And Science,
E-Mail- mahisubbu1994@gmai.com

Dr.A.V.Senthil Kumar MCA.,M.Phil.,PGDCA.,Ph.D.,
Director,
PG & Research Department of computer applications,
Hindusthan College Of Arts And Science,
E-Mail- avsenthilkumar@yahoo.co.in

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 microblogging 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. Earthquake Reporting systems are an approach to earthquake hazard mitigation which takes advantage of the rapid availability of earthquake information to quantify the hazard associated with an earthquake and issue a prediction of impending ground motion prior to its arrival in populated or otherwise sensitive areas. One such method, Earthquake Reporting System is a methodology designed to provide warnings to the peoples those are located around the world through Twitter network. Twitter is an online social networking and micro blogging service that enables social users to update their nearby location status by composing tweet and read other's "tweets", which are text messages limited to 140 characters. Registered users can read and post tweets. In order to detect the earthquake event, we use support vector machine learning approach to extract and analyze the tweet messages which are posted by twitter users .Itcan provide a few seconds to tens of seconds warning prior to ground shaking during an earthquake and also can be used to reduce the damage, costs and casualties in an earthquake.

Index Terms: **Twitter, Earthquake, Event detection, Probabilistic model, sensor**

1. INTRODUCTION

Twitter, a popular microblogging service, has received much attention recently. 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 [18]. Twitter asks one question, "What are you doing?" Answers must be fewer than 140 characters. A status update message, called a tweet, is often used as a message to friends and colleagues. A user can follow other users; and her followers can read her tweets. A user who is being followed by another user need not necessarily have to reciprocate by following them back, which renders the links of the network as directed. After its launch on July 2006, Twitter users have increased rapidly. They are currently estimated as 44.5 million worldwide1. Monthly growth of users has been 1382% year-on-year, which makes Twitter one of the fastest-growing sites in the world2. 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.

2. RELATED WORK

2.1. Micro blogging: A Semantic and Distributed Approach

After blogging that let people put their thoughts online to an open audience,podcasting where people record it and even videoblogging (also known as vlog-ging) where they deliver messages in video, microblogging enabled anyone toexchange short messages within their community or simply to write in brief tothe general public on the Web. This new form of blogging allows individuals topublish brief text updates using a multitude of various communication channelssuch as text messages from cell phones, instant messaging, e-mail and the Web.The simplicity of publishing such short updates in various situations and in auid social network based on subscriptions and response posts makes microblogging a groundbreaking communication method that can be seen as a hybridof blogging, instant messaging and status notifications, and that some alreadystudied from a social point of view Microblogging can be characterized by rapid (almost real-time) knowledge exchange and fast propagation of new information. For a company, this canmean real-time Q&As and improved informal learning and communication, as well as status notifications, e.g. about upcoming meetings and deliveries. Yet, potential for microblogging in corporate environments still has to be demonstratedwith real use cases, which we hope to happen in the next years, as already wasthe case for blogging, wikis and other Enterprise 2.0 services.

2.2. Social networks that matter: Twitter under the microscope

Scholars, advertisers and political activists see massive online social networks as a representation of social interactions that can be usedto study the propagation of ideas, social bond dynamics and viral marketing, among others. But the linked structures of social networks donot reveal actual interactions among people. Scarcity of attention andthe daily rythms of life and work make people default to interactingwith those few that matter and that reciprocate their attention. Astudy of social interactions within Twitter reveals that the driver ofusage is a sparse and hidden network of connections underlying thedeclared" set of friends and followers.

2.3 A Probabilistic Approach to Spatiotemporal Theme Pattern Mining on Weblogs

Scholars, advertisers and political activists see massive online social networks as a representation of social interactions that can be usedto study the propagation of ideas, social bond dynamics and viral marketing, among others. But the linked structures of social networks donot reveal actual interactions among people. Scarcity of attention andthe daily rythms of life and work make people default to interactingwith those few that matter and that reciprocate their attention. Astudy of social interactions within Twitter reveals that the driver ofusage is a sparse and hidden network of connections underlying thedeclared" set of friends and followers.

2.4. Can Collective Sentiment Expressed onTwitter Predict Political Elections

Research examining the predictive power of social media (especiallyTwitter) displays conflicting results, particularly inthe domain of political elections. This paper applies methodsused in studies that have shown a direct correlation betweenvolume/sentiment of Twitter chatter and future electoral resultsin a new dataset about political elections. We show thatthese methods display a series of shortcomings that makethem inadequate for determining whether social media messagescan predict the outcome of election.

2.5 From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series

Twitter is a popular microblogging service in which userspost messages that are very short: less than 140 characters,averaging 11 words per message. It is convenient for researchbecause there are a very large number of messages,many of which are publicly available, and obtaining themis technically simple compared to scraping blogs from theweb. In this paper, we connect measures of public opinionderived from polls with sentiment measured from analysisof text from the popular microblogging site Twitter.We explicitly link measurement of textual sentiment in microblogmessages through time, comparing to contemporaneouspolling data.

3 . PROBLEM DEFINITION

The research question of our study is, “can we detect such event occurrence in real-time by monitoring tweets?” 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 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.

4. PROPOSED MODEL

We developed an earthquake-reporting system using the event detection algorithm. Users can see the detection of past earthquakes. They can register their e-mails to receive notices of future earthquake detection reports. It 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. We evaluate various conditions under which alarms might be sent to choose better parameters for our proposed system. We set alarm conditions as N_{tweet} positive tweets come in 10 minute. We evaluate those methods by Precision ¼ $N_{earthquake,}$ $N_{alarmsand}$ Recall ¼ $N_{earthquake}$ $All_{earthquake}$ ($N_{earthquake}$: Number of earthquakes detected correctly, N_{alarms}: number of alarms, $All_{earthquake}$: number of all earthquakes that occurred).

5 METHODOLOGIES

Methodologies are the process of analyzing the principles or procedure for making tweet monitoring service with event analyzer to detect earthquake information from twitter network using Support Vector Machine and Particle Filtering Algorithm.

5.1 MODULES

- Authentication
- Profile Manager
- Tweet Subscription
- Tweet Analysis
- Alert Sender

5.2 MODULE DESCRIPTION

- **Authentication**

If you are the new user going to consume the service then they have to register first by providing necessary details. After successful completion of sign up process, the user has to login into the application by providing username and exact password. The user has to provide exact username and password which was provided at the time of registration, if login success means it will take up to main page else it will remain in the login page itself.

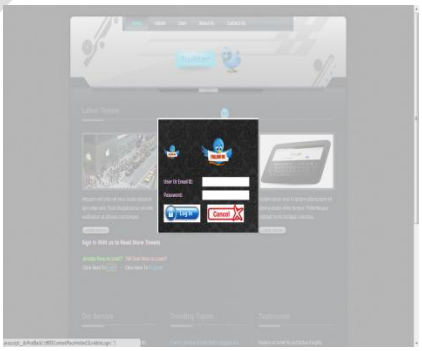


Fig .5.2.1 Authentication

- **Profile Manager**

Under this module, user can create a profile to maintain their personal information about him/her and also can able to customize the tweet subscription process.

Fig .5.2.2 Profile Manager

- **Tweet Subscription**

Under this module, registered user can able to subscribe the tweets from different source by following the genuine tweet maker. Once if we subscribe then all future updates are sent to the subscriber through electronic communication.

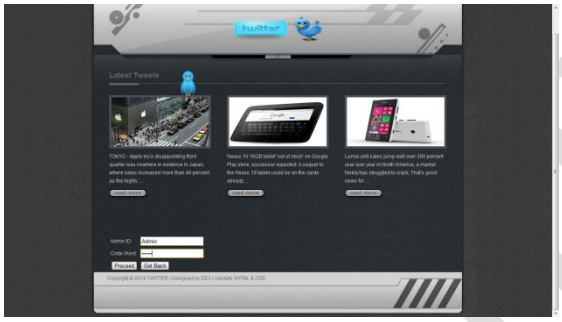


Fig .5.2.3 Tweet Subscription

- **Tweet Analysis**

Tweet Analysis Module helps to identify the tweet which having information about earthquake and will detect the target event using support vector machine learning algorithm and particle filtering algorithm for location estimation.

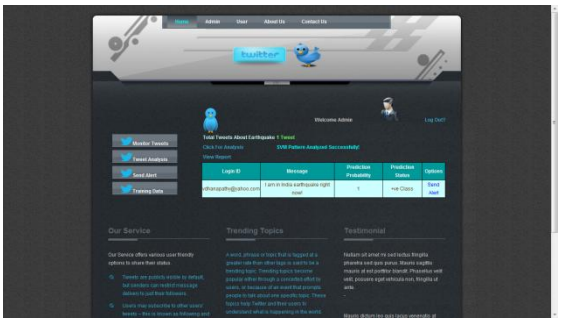


Fig .5.2.4 Tweet Analysis

- **Alert Sender**

This Module helps to prepare and distribute the alert message to twitter recipients through electronic communication along with thelatitude and longitude details.

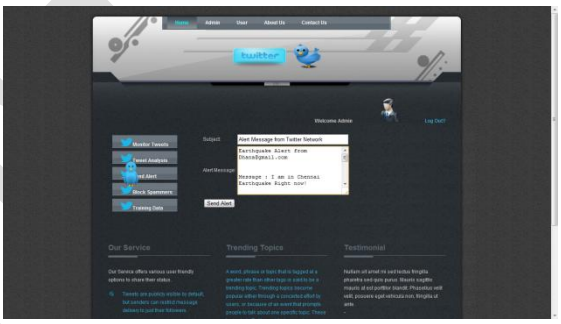


Fig .5.2.5 Alert Sender

5.3. ALGORITHM

A particle filter is a probabilistic approximation algorithm implementing a Bayes filter, and a member of the family ofsequential Monte Carlo methods. For location estimation, it maintains a probability distribution for the location estimationat time t, designated as the belief $Bel(x_t)=\{x_t^i, w_t^i\}$, i=1 . . . n. Each $x_t^i$is a discrete hypothesis related to the objectlocation. The w_t^i are nonnegative weights, called importance factors, which sum to one. 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.

STEP 2:**Re-sampling**. Resample N particles from a particle set St using weights of respective particles and allocate them on the map. (We allow re-sampling of more than that of the same particles.).

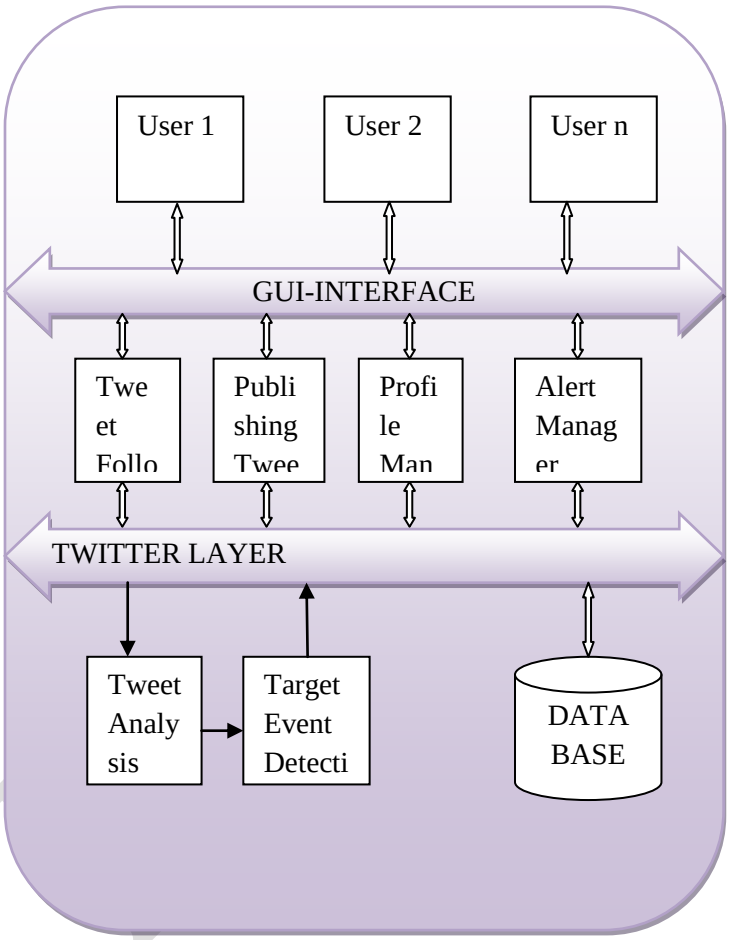
STEP 3: **Prediction**. Predict the next state of a particle set St from Newton's motion equation.

STEP 4: **Weighing**. Recalculate the weight of St by measurement.

STEP 5: **Measurement**. Calculate the current object location.

5.3. SYSTEM ARCHITECTURE

The users or nodes involved in our projects are Sender, Intermediate and Receiver. In order to send file, the sender has to find out the list of nodes which are connected with the sender. From that available list he can choose receiver. Then the sender has to analyze the performance of each and every node which is connected with the sender. The performance analysis list will return the priority based result so that sender can choose the intermediate to send the file. The Intermediate will receive the file from sender then it will analyze the performance so that it can send data to another intermediate or receiver. In the receiver side, the receiver has to select the file path to receive the file from sender or intermediate. Then the receiver can view the file received file.



6. CONCLUSION AND FUTURE ENCHNCEMENT

We investigated 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. As an application, we developed an earthquake reporting system, which is a novel approach to notify people promptly of an earthquake event.

Microblogging has real-time characteristics that distinguish it from other social media such as blogs and collaborative bookmarks. In this paper, we presented an example using the real-time nature of

Twitter. It is hoped that this paper provides some insight into the future integration of semantic analysis with microblogging data.

Reference

1. H. Lee, R.T. Ng, and K. Shim, "Extending Q-Grams to EstimateSelectivity of String Matching with Low Edit Distance," Proc. 33rdInt'l Conf. Very Large Data Bases (VLDB '07), pp. 195-206, 2007.

2. M. Hadjieleftheriou, A. Chandel, N. Koudas, and D. Srivastava,"Fast Indexes and Algorithms for Set Similarity SelectionQueries," Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE '08),pp. 267-276, 2008.

3. S. Chaudhuri, K. Ganjam, V. Ganti, and R. Motwani, "Robust andEfficient Fuzzy Match for Online Data Cleaning," Proc. ACMSIGMOD Int'l Conf. Management of Data (SIGMOD '03), pp. 313-324, 2003.

4. V. Hristidis and Y. Papakonstantinou, "Discover: Keyword Searchin Relational Data Bases," Proc. 28th Int'l Conf. Very Large DataBases (VLDB '02), pp. 670-681, 2002.

5. T. Tran, H. Wang, S. Rudolph, and P. Cimiano, "Top-KExploration of Query Candidates for Efficient Keyword Searchon Graph-Shaped (RDF) Data," Proc. IEEE Int'l Conf. Data Eng.(ICDE '09), pp. 405-416, 2009.

6. Y. Luo, X. Lin, W. Wang, and X. Zhou, "Spark: Top-K KeywordQuery in Relational Data Bases," Proc. ACM SIGMOD Int'l Conf.Management of Data (SIGMOD '07), pp. 115-126, 2007.

7. F. Liu, C.T. Yu, W. Meng, and A. Chowdhury, "Effective KeywordSearch in Relational Data Bases," Proc. ACM SIGMOD Int'l Conf.Management of Data (SIGMOD '06), pp. 563-574, 2006.

8. S. Ji, G. Li, C. Li, and J. Feng, "Efficient Interactive Fuzzy KeywordSearch," Proc. 18th ACM SIGMOD Int'l Conf. World Wide Web(WWW), pp. 371-380, 2009.

9. C. Xiao, W. Wang, X. Lin, and H. Shang, "Top-K Set SimilarityJoins," Proc. IEEE Int'l Conf. Data Eng. (ICDE '09), pp. 916-927,2009.

10. B. Ding, J.X. Yu, S. Wang, L. Qin, X. Zhang, and X. Lin, "FindingTop-K Min-Cost Connected Trees in Data Bases," Proc. IEEE 23rdInt'l Conf. Data Eng. (ICDE '07), pp. 836-845, 2007.