© March 2016 | IJIRT | Volume 2 Issue 10 | ISSN: 2349-6002

Tweet Analysis for Earthquake Reporting System Development

B. Kishore
M. Tech Student, Dept. of CSE, B V C Engineering College Odalarevu, A.P, India

***Abstract*- Twitter has gotten much consideration as of late. An essential normal for Twitter is its ongoing nature. We explore the constant collaboration of occasions, for example, tremors in Twitter and propose a calculation to screen tweets and to recognize an objective occasion. To identify an objective occasion, we devise a classifier of tweets in view of components, for example, the watchwords in a tweet, the quantity of words, and their connection. In this way, we produce a probabilistic spatiotemporal model for the objective occasion that can locate the focal point of the occasion area. We see every Twitter client as a sensor and apply molecule separating, which are generally utilized for area estimation. The molecule channel works superior to anything other similar techniques for assessing the areas of target occasions. As an application, we build up a tremor reporting framework for use in Japan. Due to the various seismic tremors and the expansive number of Twitter clients all through the nation, we can identify a quake with high likelihood (93 percent of quakes of Japan Meteorological Agency (JMA) seismic force scale 3 or more are recognized) simply by observing tweets. Our framework recognizes tremors quickly and warning is conveyed much quicker than JMA show declarations**

***Index Terms*— Twitter, occasion location, social sensor, area estimation, quake**

I. INTRODUCTION

Ever since Social media introduced, it has 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 micro blogging 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. 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, many studies have been made in the seismology field. This paper presents an innovative social approach that has not been reported before in the literature.

II. RELATED WORKS

We pick quakes in Japan as target occasions, taking into account the preparatory examinations. We clarify them in this segment. In the first place, we pick quakes as target occasions for the accompanying reasons:

Seismic perceptions are led around the world, which encourages obtaining of tremor data, which likewise makes it simple to accept the exactness of our occasion discovery system; and It is entirely important and significant to recognize quakes in quake inclined locales. Second, we pick Japan as the objective range taking into account the accompanying examination.

Fig. 1 depicts a guide of Twitter clients around the world (got from UMBC eBiquity Research Group); Fig. 2 delineates a guide of seismic tremor events around the world (utilizing information from Japan Meteorological Agency (JMA)). It is clear that the main crossing point of the two maps, those areas with numerous seismic tremors and huge Twitter clients, is Japan. Different locales, for example, Indonesia, Turkey, Iran, Italy, and Pacific beach front US urban communities, for example, Los Angeles and San Francisco likewise generally converge, however their individual densities are much lower than that in Japan. Numerous seismic tremor occasions happen in Japan and numerous Twitter clients watch quakes in Japan, which implies that social sensors are dispersed all through the nation. We exhibit a brief diagram of Twitter in Japan: the Japanese adaptation of Twitter was propelled on April 2008. In February 2008, Japan was the No. 2 nation as for Twitter traffic.5. At the season of this written work, Japan has the second biggest number of tweets (18 percent of all tweets are posted from Japan) on the planet. Along these lines, we pick seismic tremors in Japan as an objective occasion due to the high thickness of Twitter clients and quakes in Japan.

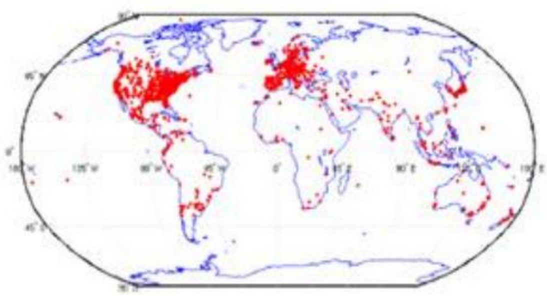

Fig. 1. Twitter User Map

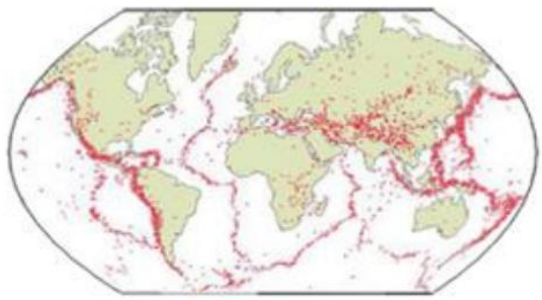

Fig. 2. Earth quake Map

III. PROPOSED METHODS

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 Fig.3.

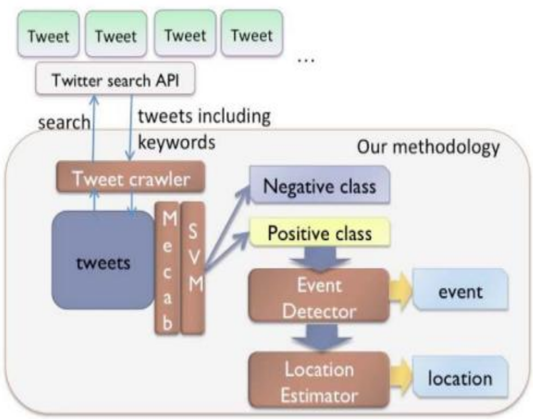

Fig. 3: Method to acquire tweets referred to a target event precisely.

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. Fig. 4 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.

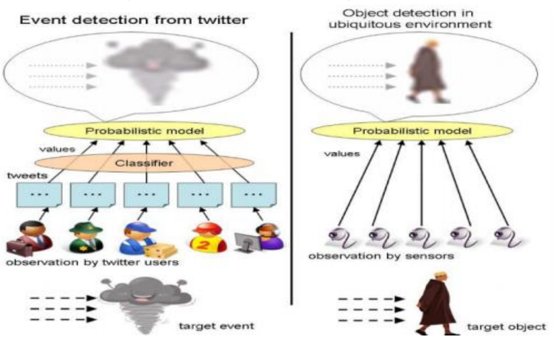

Fig.4: Correspondence between event detection from Twitter and object detection in a ubiquitous environment.

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 evenly on map.

$S_0= (S_0^0 ,S_0^1,S_0^2 ,......,S_0^{N-1})$

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) ∈St.

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

IV. CONCLUSION

As of late, Twitter, a prominent miniaturized scale blogging administration, has turned into another data channel for clients to get and to trade data. Ordinary, about 170 million tweets are made and redistributed by a huge number of dynamic clients. Individuals from better places might tweet about the same occasion (e.g, "Norway shooting") or the same kind of occasion (e.g, "quake"); by gathering tweets after some time. We research the constant communication of occasions, for example, quakes in Twitter and propose a calculation to screen tweets and to identify an objective occasion. To start with, to get tweets on the objective occasion correctly, we apply semantic examination of a tweet. For instance, clients may make tweets, for example, "Quake!" or "Now it is shaking," for which tremor then again shaking could be catchphrases, however clients may likewise make tweets, for example, "I am going to an Earthquake Conference," then again "Somebody is shaking hands with my supervisor."

REFERENCES

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

[2] Holt, R. (2013). Twitter in numbers. , retrieved April 19, 2013, from http://www.telegraph.co.uk/technology/twitter/9945505/Twitterinnumbers.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) Microblogging 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 real time Twitter analysis during crises for operational crisis management, Proceedings of the 9th International ISCRAM Conference, Vancouver, Canada.

BIODATA

Author

B.Kishore has received her B.E in Computer Science and Engineering, from Saraswathi Velu College of engineering in 2011. Presently he is pursuing his M.Tech in Software Engineering from BVC Engineering College Odalarevu, A.P, India. Her research interests include Software testing Data structures and web services