International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 4 | ISSN : 2456-3307

Real-Time Event Recognition and Earthquake Reporting System Development by Using Tweet Analysis

M. Vijay Kumar
MCA Sri Padmavathi College Of Computer Sciences And Technology Tiruchanoor, Andhra Pradesh , India

ABSTRACT

Twitter has received a lot of attention recently. A very important characteristic of Twitter is its period of time nature. Abstraction event foretelling from social media is probably very helpful however suffers from essential challenges, like the dynamic patterns of options (keywords) and geographic non uniformity (e.g., abstraction correlations, unbalanced samples, and totally different populations in numerous locations). Most existing approaches (e.g., LASSO regression, dynamic question enlargement, and burst detection) address some, however not all, of those challenges. We tend to investigate the period of time interaction of events like earthquakes in Twitter Associate in Nursing propose a rule to observe tweets and to observe a target event. To observe a target event, we tend to devise a classifier of tweets supported options like the keywords during a tweet, the quantity of words, and their context. Later, we tend to turn out a probabilistic spatiotemporal model for the target event which will realize the middle of the event location. We tend to regard every Twitter user as a sensing element and apply particle filtering, that area unit wide used for location estimation. The particle filter works higher than different comparable strategies for estimating the locations of target events.

Keywords: Twitter, Event Detection, Earthquake, LASSO

I. INTRODUCTION

Today Social Networking Sites (SNS) have become a part of our day to day life. We share a lot of data on these sites. They helped us to make the world smaller and integrated with each other. There are many SNS available today and many more are piling each day. Thus users use many SNS each day and communicate and share data with friends and family. This communication medium gave rise to complex structure whether a user really like the SNS which he uses more or he needs another SNS other than he uses more. Thus one of the most famous SNS is TWITTER which is used to share data and post our thoughts and latest buzz upon the internet. The users using TWITTER have increased dramatically in the recent years. So the analysis of this SNS may help in answering and predicting many answers. This online social network (twitter, Face book, etc.) is used by millions of people around the world to remain socially connected to their friends, family members, and work met through their computers and mobile phones. Answer is less than 140 characters when Twitter asks the question, "What's happening?". A status update message, called a tweet, is often used as a tweet to friends and colleagues. A one user can follow other users; that user's followers can read her tweets on a daily 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. Since its launch on July 2006, Twitter users have increased fast. The number of registered Twitter users added 100 million in April 2010. The service(twitter) is still adding about 300,000 users per day.1 now a days, 190 million users

use Twitter per month, generating 65 million tweets per day.

TWITTER is the red hot tool for micro blogging and social networking these days. Started in late march of 2006 and twitter's off-the-wall the features makes twitter stand tall in this cyber world. As it is era of blogging, micro blogging and people connecting through social sites hence one cannot overlook online blogging and social networking site named TWITTER which differs from traditional blogging and has vital add inns. It is a web application which gives users features like Direct Messaging, Following People & Trending Topics, Links, Photos, Videos message, image, or video links to share with their peers/colleagues and with followers such as personal online diaries or news on particular subject also one important aspect to notice is the small message refers to only `140 characters. These short messages are called tweets. Hash tags are those which starts with special characters # and which is meant to group similar micro blog topics such as #economics and #amazing. The information flow, i.e tweet flow from author (source) to follower (subscriber) and is bidirectional. Generally when a user post tweets, the tweets are displayed on both the user and the author home page. As reported in august 2011,twiteer has attracted 200 million users and produced 8.3 million tweets per hour tweets it ranks 10th among the top 500 site list as per Alexa in December 2011[1][2].

Other side of this is malicious bots have been greatly misused by spammers to spread spam. Spam can be defined as spreading malicious, phishing, or unsolicited commercial content in tweets. Bot adds users as their friends and expects users to follow back. In this way homepages are displayed with such spam's tweets. The content attracts users by appealing text content, accidently users may visit such link by clicking which gets rerouted to spam or malicious sites. Experience of those users (i.e. human users). who are in such situation wherein they are surrounded by malicious bots and spam tweets become progressively worse and at the end there is risk of whole community getting affected by these bots and gets hurt. The ultimate approach of this paper is identify and classify automation feature of Twitter accounts into 3 categories, human, bots, and cyborgs which we will manage. This will help Twitter to have healthy community tweets and also human users to recognize the real tweets. An automated Classification system proposed here consists of four major components. 1. Entropy component: tweeting interval as a measure of behavior complexity, and detects the periodic and regular timing that is an indicator of automation which is used by the entropy component. 2. Spam detection component: tweet content to check whether text patterns contain spam or not is used by the spam detection component 3. Account properties component: this component utilize useful account properties, such as tweeting device makeup, URL ration, to detect deviations from normal tweet behavior; and finally 4. The decision maker: The decision maker is based on Random Forest, and it uses the combination of the features generated by the above three parts to group an unknown user as human, bots, or cyborgs.

II. PROPOSED ALGORITHM

EVENT DETECTION:

As described in this paper, we target 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 [13]. We 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 (many users experience the event).
2. They particularly influence the daily life of many people (for that reason, people are induced to tweet about it).

3. They have both spatial and temporal regions (so that real-time location estimation is possible).

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. 1. 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.

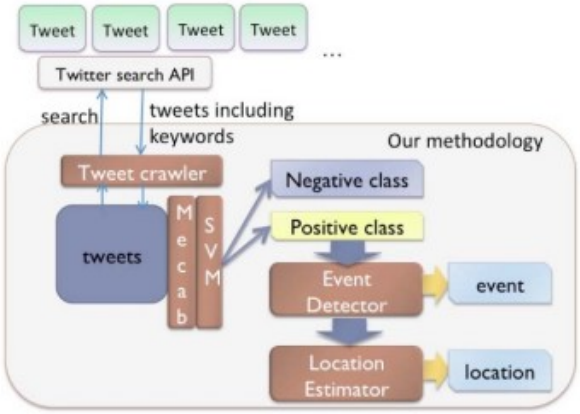

Figure 1. Method to acquire tweets referred to a target event precisely

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.

Table 1. 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

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. If she makes a tweet about an earthquake occurrence, then it can be considered that she, as an "earthquake sensor," returns a positive value. A tweet can therefore be regarded as a sensor reading. This crucial assumption enables application of various methods related to sensory information.

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

Assumption 2. Each tweet is associated with a time and location, which is a set of latitude and longitude, coordinates.

III. MODEL

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.

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. Actually, the data fit an exponential distribution very well. We get ¼ 0:34 on average,

Assuming that we have n sensors, which produce positive signals, the probability of all n sensors returning a false alarm is p_f^x therefore, the probability of event occurrence can be estimated as 1- p_f^x.

Therefore, the number of sensors we expect at time t is

$$\sum_{t_k=0}^{t} n_0 e^{-\lambda t_k} = n_0(1 - e^{-\lambda(t+1)})/(1 - e^{-}$$

Consequently, the probability of an event occurrence at time t is

$$p_{occur}(t) = 1 - p_f^{n_0(1-e^{-\lambda(t+1)})/(1-e^{-\lambda})}.$$

We can calculate the probability of event occurrence if we set λ =0:34 and Pf= 0:35.

Spatial Model:-

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

1. Generation: - Generate and weight a particle set, which means N discrete hypothesis.

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

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. (We allow resampling of more than that of the same particles.).

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

$$(x_t^k, y_t^k) = \left(x_{t-1}^k + v_{x_{t-1}}\Delta t + \frac{a_{x_{t-1}}}{2}\Delta t^2,\ y_{t-1}^k + v_{y_{t-1}}\Delta t + \frac{a_{y_{t-1}}}{2}\Delta t^2\right)$$

$$(v_{x_t}, v_{y_t}) = (v_{x_{t-1}} + a_{x_{t-1}}, v_{y_{t-1}}, a_{y_{t-1}})$$

$$a_{x_t} = \mathcal{N}(0;\sigma^2), \quad a_{y_t} = \mathcal{N}(0;\sigma^2).$$

4. Weighing. Recalculate the weight of St by measurement m(m_x ,m_y)as follows:

$$dx_t^k = m_x - x_t^k, \quad dy_t^k = m_y - y_t^k$$

$$w_t^k = \frac{1}{(\sqrt{2\pi}\sigma)} \cdot exp\left(-\frac{(dx_t^{k2} + dy_t^{k2})}{2\sigma^2}\right).$$

5. Measurement. Calculate the current object location

$$o(x_t, y_t) \text{ by the average of } s(x_t, y_t) \in S_t.$$

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

IV. CONCLUSION

As represented during this paper, we tend to investigate the period nature of Twitter, devoting explicit attention to event detection. Linguistics analyses were applied to tweets toclassify them into a positive and a negative category. We regard every Twitter user as a detector, and set the matter as detection of an occasion supported sensory observations.

Location estimation ways like particle filtering square measure used to estimate the locations of events. As Associate in nursing application, we developed Associate in nursing earthquake reportage system, which is a novel approach to inform individuals promptly of Associate in nursing earthquake event. Microblogging has period characteristics that distinguish it from alternative social media like blogs and collaborative bookmarks. As represented during this paper, we presented Associate in nursing example that leverages the period nature of Twitter to form it helpful in determination a crucial social problem: natural disasters.

V. 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]. 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.

[3]. 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.

[4]. H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, A Social Network or A News Media?" Proc. 19th Int'l Conf. World Wide Web (WWW '10), pp. 591-600, 2010.

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

[6]. 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.

[7]. P. Galagan, "Twitter as a Learning Tool. Really," ASTD Learning Circuits, p. 13, 2009.

[8]. K. Borau, C. Ullrich, J. Feng, and R. Shen, "Microblogging for Language Learning: Using Twitter to Train Communicative and Cultural Competence," Proc. Eighth Int'l Conf. Advances in Web Based Learning (ICWL '09), pp. 78-87, 2009.

[9]. J. Hightower and G. Borriello, "Location Systems for Ubiquitous Computing," Computer, vol. 34, no. 8, pp. 57-66, 2001.

[10]. M. Weiser, "The Computer for the Twenty-First Century," Scientific Am., vol. 265, no. 3, pp. 94-104, 1991.

[11]. 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.

[12]. 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.

[13]. Y. Raimond and S. Abdallah, "The Event Ontology," http:// motools.sf.net/event/event.html, 2007. 14T. Joachims, "Text Categorization with Suport Vector Machines: Learning with Many

Relevant Features," Proc. 10th European Conf. Machine Learning (ECML '98), pp. 137-142, 1998.

[14]. 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.

[15]. 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.

[16]. M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/NonGaussian Bayesian Tracking," IEEE Trans. Signal Processing, vol. 50, no. 2, pp. 174-188, Feb. 2002.

[17]. J. Leskovec, L.A. Adamic, and B.A. Huberman, "The Dynamics of Viral Marketing," Proc. Seventh ACM Conf. Electronic Commerce (EC '06), pp. 228-237, 2006.

[18]. Y. Matsuo and H. Yamamoto, "Community Gravity: Measuring Bidirectional Effects by Trust and Rating on Online Social Networks," Proc. 18th Int'l Conf. World Wide Web (WWW '09), pp. 751-760, 2009.

[19]. W. Zhu, C. Chen, and R.B. Allen, "Analyzing the Propagation of Influence and Concept Evolution in Enterprise Social Networks Through Centrality and Latent Semantic Analysis," Proc. 12th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD '08), pp. 1090-1098, 2008.