Journal of Advanced Engineering Research
ISSN: 2393-8447
Volume 1, Issue 1, 2014, pp.41-47

Real Time Nature Disasters of Data Analysis Using Tweet

K. Sundharivel[1], D. Udaya Suriya Rajkumar[2,] *

[1] Department of Computer Science and Engineering, Arulmigu Meenakshi Amman College of Engineering, India.
[2] Department of Computer Science and Engineering, GGR College of Engineering, Vellore, India.

* Corresponding author email: u_suriya@yahoo.com, Tel.: +91 9944925121

ABSTRACT

Twitter is an online social network in which the people can share the real time activity. Here, we discuss the real-time interaction of events such as Weather Forecasting report in Twitter and propose a method to monitor tweets. To finding an event, there are two methods tweets based classifier such as the tweet keywords, and their context. For that, we implement a probabilistic spatiotemporal model for the target event that can find the information and location of the event. We apply particle filtering and twitter uses as a sensor, which are widely used for location estimation. For estimating the location particle filter methods are used. As an application, we develop a weather reporting system for use in Tamil Nadu to know the weather throughout the country. We can detect information about the cyclone with high probability. Our system identifies the information about the weather and detects the information about cyclone and change in the climate and notification is delivered much faster to the register user than other broadcast announcements.

Keywords - Event detection, Twitter, location Estimation, Weather.

1. INTRODUCTION

Twitter is an online social network is used by more numbers of peoples throughout the world to remain socially connected to their friends, members in a family and Colleagues through their computers and mobile phones [1]. Tweet is nothing but a status update message, is often used as a message to friends, relatives and others. A user can follow other users; that user's followers can read her tweets regularly. A user need not necessarily to follow again, and the links of the network as directed. Since its launch on July 2006. The number of registered Twitter users exceeded more than 100 million in April 2010. The service is currently, 190 million users can use a Twitter per month, generating 65 million tweets per day. Most studies can be classified into one of three groups: first, some researchers have analyzed the network structure of Twitter [2]. Second, some researchers have specifically examined characteristics of Twitter as a social medium Third, some developers creating a new application [12, 13].

Micro blogging services are their real-time nature. Users can know how other users are doing and thinking now and again and again return to the site and check to see what other done. The real time nature: in the case changing of weather and climate were transmitted through Twitter. People were able to know about the weather in particular place immediately. In such a manner, numerous update and reports related to events and disastrous such as storms, fires, traffic jams, riots, rainfall, and earthquakes [6].

1. This paper investigates real-time nature of Twitter and event notification system to monitors tweets and delivers notification promptly from the investigation using knowledge. Here, we take three steps: first, we crawl tweets related to event target; second, we estimate the location to target events finally developed a cyclone reporting system that extracts climate changes from Twitter and sends a message to registered users [7]. Here, we explain our methods using a weather reporting using semantic analysis as a target event over tweet.

2.We prepare the training data using a support vector machine and devise a classifier based on features such as tweet in a, the number of words, and the target-event word context and a probabilistic spatiotemporal model of an event. For that virtual sensors are used to designate a social sensor, and some sensors are very active; others are not [3]. When a user is sleeping, or busy doing something else a sensor might be inoperable or malfunctioning. Compared to ordinary physical sensors social sensors are very noisy [11]. The event-detection problem can be reduced using a sensor to location estimation and object detection in a ubiquitous pervasive computing environment in which we have numerous location sensors: a user has an active badge or mobile device in an environment where sensors are placed. The user location is estimated and provides location-based services such as museum guides and navigations [4].

We apply particle filters, which are widely used for location estimation in pervasive computing [5].As an application, we develop a weather reporting system. Twitter users are similarly numerous and

geographically dispersed through all over country. So it is possible to detect sometimes a cyclone by monitoring tweets. Our system detects a cyclone occurrence and sends an e-mail, possibly before a cyclone actually arrives at a certain location. Therefore, prompt notification of cyclone occurrences is extremely important to decrease damage by natural disaster. For cyclone prediction and warning, many studies have been done in the geographic field. This paper is organized as described below. We explain an investigation of Twitter users and cyclone in the real world. Section 2 Previous Work Section 3 contribution of the work Section 4. We describe the evaluation and experiments of event detection in Section 5. Finally, we conclude the paper.

2. PREVIOUS WORK

We choose target event as cyclone, based on the primary investigations. We choose cyclone as target events for the following reasons:

- Meteorological observations are conducted worldwide, which facilitates acquisition of cyclone information and also makes it easy to validate the event detection methodology

- It is valuable to detect cyclone in prone regions, we choose India as the target based on investigation. Other regions such as Pakistan, Bangladesh, Turkey, SriLanka and pacific coastal also roughly intersects, but their respective densities are much lower than that in India. Many cyclone events occur in India and many Twitter users observe cyclone in India, which means that social sensors are distributed throughout the country.

2.1 Event Detection

In this paper, an arbitrary classification of a space-time region for event detection. An event might have passive factors, products, estimation time and actively participating agents [17]. We target events such as cyclones, 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 in tweet. 3. They have both spatial and temporal regions (so that real-time location estimation is possible). We design an event to detect a target event using Twitter. In this section, we explain how to detect a target event. First, we crawl tweets related to keywords of target event. Then extract tweets refer to a target event using devices that have been trained with machine learning. Second, Twitter users a social sensors to detect a target event and estimate the location. The method to acquire tweets referred to a target event precisely is shown in figure-1

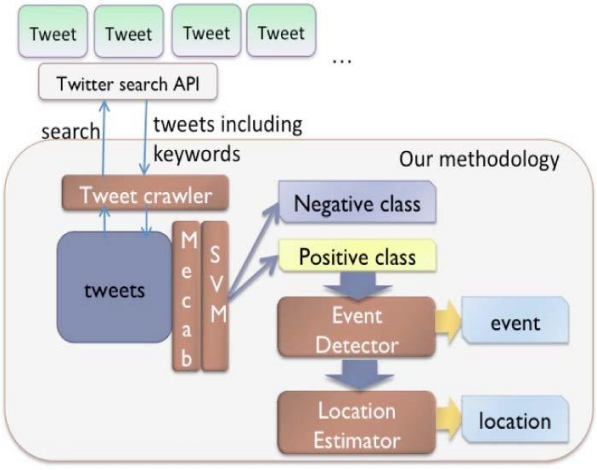


Fig. 1 Method to acquire tweets referred to a target event precisely.

2.2 Spatial Temporal Model

Each tweet is associated with a location and estimate of the event from sensor readings. The problem of location estimation and evolution of the problem to solve Bayesian filters are proposed such as kalman filters, multi-hypothesis tracking, grid-based and topological approaches, and particle filters [10]. We use particle filters which are widely used in location estimation. Each tweet has its own message post time using sensors to event detect. The temporal model of event detection. First, we examine the actual data. Specifically regarding a cyclone, more than 10 cyclones occurred during the period. In the Twitter case, a user detects an event, and assume that the probability of his posting a tweet from t. Then, the time to produce a tweet can be regarded as exponential distribution. If a user detects an event, a tweet immediately not online or something differ [16]. After resolved the problems resolved she might make a post. 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 average, based on alarm to calculate the reliability of multiple sensor values. For example, a user might produce an alarm false by writing a tweet. The alarms are design using two facts. The false-positive ratio pf of a sensor is approximately 0.35. Sensors are assumed to be independent and identically distributed. Assume that n sensors, which produce positive signals and the probability of all n sensors, return a false alarm.

2.3 Particle Filters

A particle filter is a probabilistic approximation algorithm for estimate the location, and it maintains a probability distribution for the location estimation at time t, designated as the belief measurement. We must consider the sensor geographic distribution to treat readings of social sensors more precisely. In location estimation the physical sensors are located evenly in many cases. Social sensors are not placed evenly in

many cases because social media users are concentrated in different areas [15]. The spatial models are used in social sensors to get geographic distribution. The few social sensors in various areas where fewer Twitter users live, those sensors have lower probabilities to response value, if a sensor in a less-populated area produces a positive value to one cyclone, then it can be inferred to the center of the cyclone is nearer to that sensor. Hence the values of sensor are more important in less populated areas are more important than those in densely populated areas. We calculate weights of respective particles based on the geographic distribution of social sensors based on the assumptions. We use a more advanced algorithm with re-sampling [9]. To examine the biases of user locations weight distribution as obtained from the Twitter user distribution. We customize the algorithm related to particle filters as follows:

1. We collect Twitter users randomly along with their location information.

2. In the Generation step, we weight each particle based on weight distribution after they are allocated.

3. In the Weighing step, we calculate the weights of each particle.

2.4 Techniques to Speed up the Process

In this paper, we want to estimate location of events quickly as soon as possible to develop a real-time cyclone detection system. Therefore, we must decrease the time complexity of methods used for location estimation. The time complexity of a normal particle filter is expressed as p, number of particles Nm, and number of observations. The time complexity of the weighted particle filter is expressed as Ns, number of sensors to calculate the geographic distribution. It takes less than 1 s to estimate the location of a cyclone center using as usual particle filter. The weighted particle filter can takes less from 1 minute to 3 minute to estimate the location of a cyclone. Therefore, we want to decrease Ns to calculate the location of cyclone centers more quickly. As described in this paper, we sample some users from all users to calculate the sensor geographic distribution and produce a new set of S users. We apply the following three approaches.

3. CONTRIBUTION OF THE WORK

The proposed system, users can see the detection of past cyclones occurs. They can register their e-mails to receive notices of future cyclone detection reports. It alerts users and urges them to prepare for the imminent cyclone. It is hoped that a user receives the e-mail before the cyclone actually affects that area. We evaluate various conditions under which alarms might be sent to choose better parameters for our proposed system to detecting multiple errors. The first tweet of a cyclone is usually made within a minute. That delivery is far faster than the rapid broadcasts of announcements of IMA, which are widely broadcast on TV; on average, IMA announcement is broadcast 6 minutes after a cyclone occurs. Based on these results, we infer that our system probably has a high recall rate and medium precision. When several cyclones occur during a day sometimes the system can produce a false alarm. The static condition for giving an alarm for current system "Ntweet" in every 10 minutes". We must change the condition dynamically to increase the precision of the system and in terms of the repetition and intensity of cyclone.

3.1 Cyclone Reporting System

We developed a cyclone-reporting system using the event detection algorithm. Cyclone information is much more valuable in real time and gives prior warning. Vast amounts of work have been done on intermediate term cyclone prediction in the meteorological field. Various attempts have also been undertaken to produce short-term forecasts to realize a cyclone warning system by observing electromagnetic emissions cyclone location estimation based on tweets. Balloons show the tweets related to a cyclone. The cross shows the cyclone epicenter. Red represents early tweets; blue shows later tweets. Location estimation, accuracy of weather and a typhoon trajectory from Tweets for each method and present the estimation latitude and longitude of actual ones, and their Euclidean distance for better performance. The sampled particle filter mean values are used for sampling method of sampled 300 users. In India, the government has allocated a considerable amount of its budget to mitigating cyclone damage. A cyclone warning service has been operated by IMA since 2007 it estimated the arrival times and estimated the changing of weather intensities.

3.2 Semantic Analysis of Tweets

The classification performance we use two query words: cyclone and thunder Performance results obtained. The linear kernel is used to obtain the highest F-value when we use feature A and all features using SVM. Surprisingly, features B and C do not contribute much to the classification performance. When a cyclone occurs, a user becomes surprised and might produce a very short tweet. That result is attributable to the usage of query words in a different context than we had intended. In Sometimes it is difficult to judge whether a tweet is reporting an actual cyclone or not. Some examples are that a user might write "Is this and cyclone or a truck passing?" Overall, the classification performance is good considering that we can use multiple sensor readings as evidence for event detection. To detect a target event from Twitter and find useful tweets. Tweets might include the target event and users might make tweets such as storm, thunder and cyclone. Apart from that users might also make tweets such as "I am attending a cyclone Conference." or "Someone is shaking hands with my boss." Moreover, it

might not be appropriate as an event report and it is referring to the target event. For instance, a user makes tweets such as "The cyclone yesterday was scary. These tweets are truly descriptions of the target event, but it is not a real report of the event. To classify a tweet as a positive class or a negative class, we use a support vector machine [14], which is a widely used machine-learning algorithm.

3.3 System Architecture

System architecture is used to define the structure and behavior of a conceptual system. An architecture description is a formal description of a system, supports reasoning organized in a way about the structural properties of the system. This may enable the business needs to manage investment. Here explain methods using a weather reporting as a target event. First obtain tweets on the target event precisely and apply semantic analysis of a tweet. The fundamental organization of a system embodied in its components their relationships to each other and the environment and the principle governing its evolution and design. The composite of the design architectures for products and their life cycle processes. An allocated arrangement of physical elements which provides the design solution for a consumer product or life-cycle process intended to satisfy the requirements of the functional architecture and the requirements baseline. The most important, pervasive, top-level, strategic inventions, decisions and overall structure (i.e., essential elements and their relationships) and associated characteristics and behavior and proposes an event notification system that monitors tweets and delivers notification promptly using knowledge from the investigation. In this paper there are three steps: first, crawl numerous tweets related to target events, second probabilistic models to extract events from those tweets and estimate locations of events; finally, developed a weather reporting system that extracts information about weather from Twitter and sends a message to registered users. The system architecture diagram will show how the process may be done. At first, the user can post the message in the tweeter and the message can be stored in the b database. Through b data base the analyst can analyze the message. Generally, analyst can crawl the message from the tweeter the information presented in b database and classify the message using event based on semantics messages related to the target event. At first generally can crawl the message related to event when using the semantic can get the information not only the related event and also various information about the event. By having this information the analyzer can analyze the message about the target event and also the same event can be occur before or not. If the same event may be occur in before by having that information like where the event may be occur and also how it will occur can be analyzed by the analyzer. After analyze these the analyzer can easily identify the location of the event and forecast the message to the followers. The following figure-2 shows the system architecture diagram.

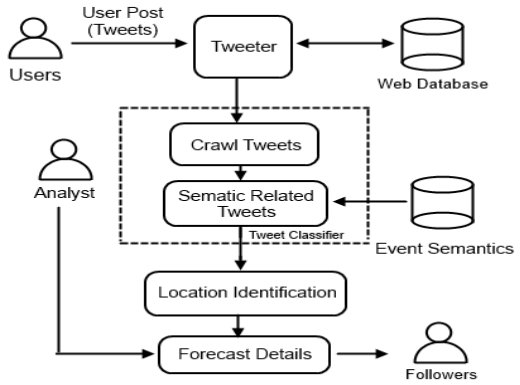


Fig. 2 System Architecture

There are various types of modules involved are

- Tweeter Environment Construction
- Tweet Crawler
- Location Identification
- Event Prediction

3.4 Tweeter Environment Construction

This module provides the basic functionalities of twitter like Twitter users write tweets several times in a single day other thinking now. Users come again and again to the site and check the status of other. The real-time nature is numerous update of results reports related to events. The basic things of this module are registration from, login form, Create profile, Post message, View message, Follow friends.

3.5 Tweet Crawler

This module provides the functionalities to completely search tweeter for post relative to the specified event. Crawler search posts with given semantic relationship relative to specific event to understand the context of the users posted message. Crawler uses support vector machine algorithm for classification of event relevant posted message from the generally posted messages. To detect a target event from Twitter, search from Twitter and find useful tweets. The method of acquiring useful tweets for target event detection is portrayed. Tweets might include the target event. Even if a tweet is referring the target event, it is not an appropriate event

report. These tweets are truly descriptions of the target event, but the events are not real-time reports.

3.6 Location Identification

This module provides functionality to identify the location of event by analyzing the data extracted from event relevant posted message classified from the generally posted messages. A tweet can be associated with a time and location of the tweet post and it can be obtained using application programming interface. Sometimes GPS data are attached to a tweet when user is using a phone. Alternatively, each Twitter user makes a registration on their location in the user profile. The registered location and current location might not be the same. I use GPS data and the registered location of a user and not use tweets for spatial analysis if a location is not available; however, use the tweet information for temporal analyses. A particle filter is a probabilistic approximation algorithm implementing a Bayes filter, and a member of the family known as Monte Carlo methods. For estimate the location, it maintains a probability distribution for the location estimation at time t. The Sequential Importance Sampling (SIS) algorithm is a Monte Carlo method that forms the basis for particle filters.

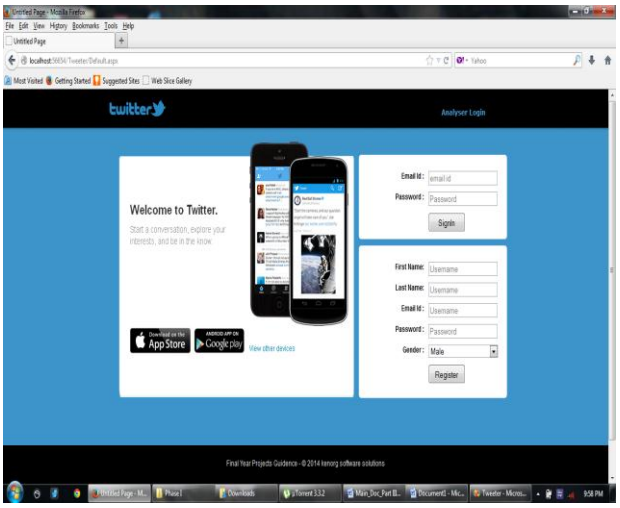
3.7 Event Prediction

This module is used to predict a future event from the extracted event from the post along with their location and time. This module uses a statistical mathematic model for predicting the future event. An arbitrary classification of event is a time space region. In target events such as earthquakes, typhoons, and traffic jams are examined in tweets. These events have several properties are large scale (many users experience the event).They particularly influence the daily life of many people (for that reason, people are induced to tweet about it).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 and political campaigns.

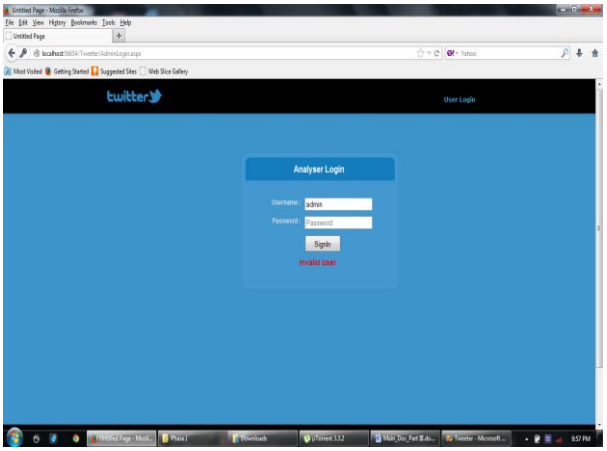
4 EXPERIMENTS AND EVALUATION

Here, we describe the experiment and evaluation results of tweet classification and location estimation.

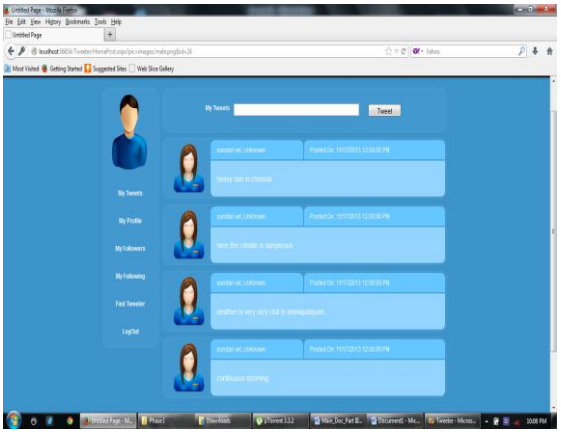
User Login form:



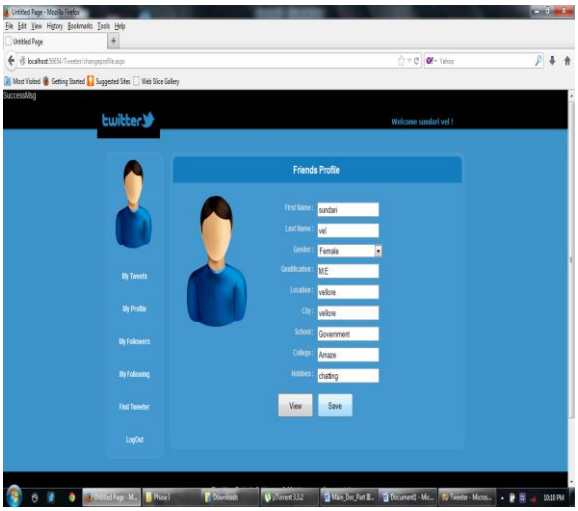
Analyser Login Form:



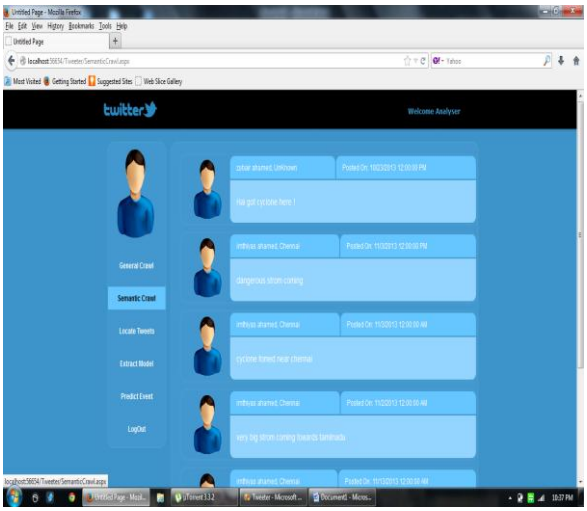
Tweet:



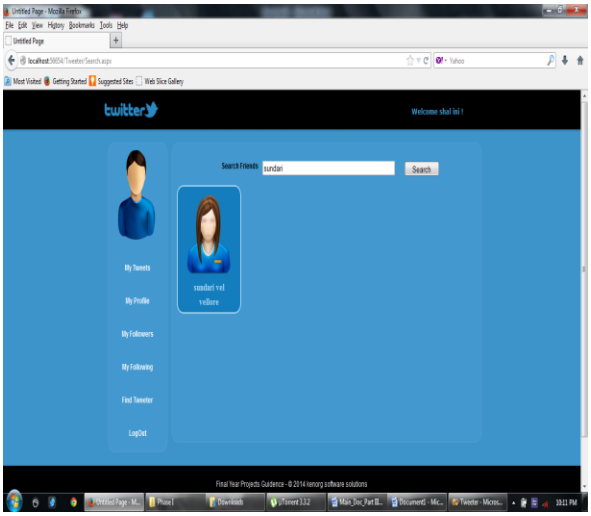
Tweet Profile:



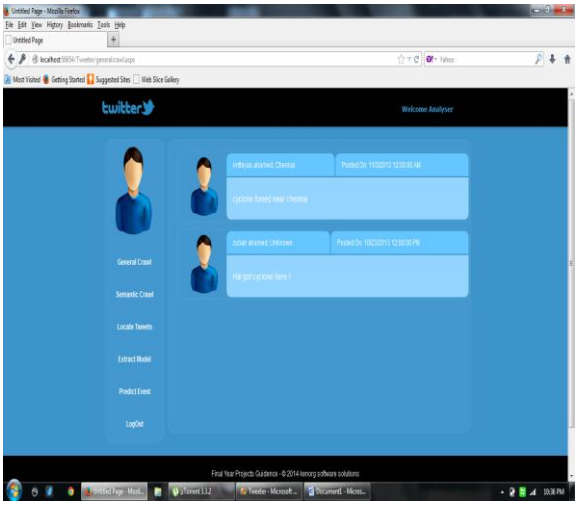
Semantic Crawl:

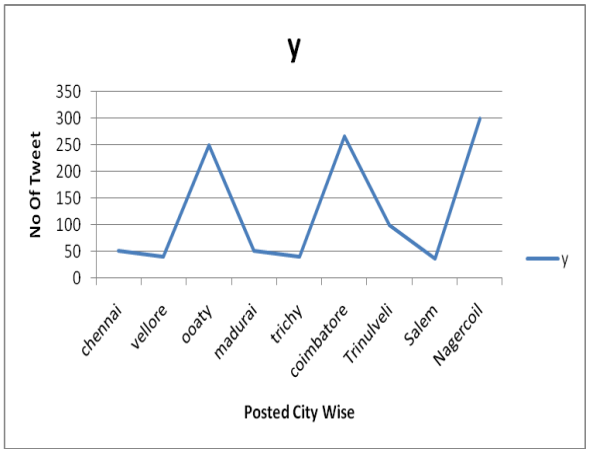


Search Friends:



General Crawl:





Graph-1 Relate Message Posted Region Wise

The above graph-1 show that number of tweet posted relate to the target event from the various location. This is used to identify where the more number of tweet is posted that will be consider as event may occur more.

5. CONCLUSIONS

As described in this paper, we investigated the nature of real time Twitter, devoting particular attention to event detection. Semantic analysis was applied to tweets to classify them into a negative and positive class. We regard each Twitter sensors are used to set the problem and detect an event based on observation of sensor. To estimate the location particle filtering events are used. As an application, we developed a cyclone reporting system, used to notify promptly about the climate changes in the country and send alert information about the cyclone.

5.1 Future Enhancement

In future we plan to solve an important social problem: natural disasters. It is hoped that this paper will provide some insight into the future integration of semantic analysis with micro blogging data.

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] P. Galagan, Twitter as a learning tool, *ASTD Learning Circuits*, 2009, 13.

[3] J. Hightower and G. Borriello, Location systems for ubiquitous computing, *Computer*, 34 (8), 2001, 57-66.

[4] M. Weiser, The Computer for the Twenty-First Century, *Scientific Am.*, 265 (3), 1991, 94-104.

[5] V. Fox, J. Hightower, L. Liao, D. Schulz and G. Borriello, Bayesian filtering for location estimation, *IEEE Pervasive Computing*, 2 (3), 2003, 24-33.

[6] M. Arulampalam, S. Maskell, N. Gordon and T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, 50 (2), 2002, 174-188.

[7] 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*), 2008, 1090-1098.

[8] J. Leskovec, L.A. Adamic and B.A. Huberman, The dynamics of viral marketing, *Proc. Seventh ACM Conf. Electronic Commerce (EC '06)*, 2006, 228-237.

[9] Y. Matsuo and H. Yamamoto, Community gravity: measuring bidirectional effects by trust and rating on online social networks, *Proc. 18th International Conference on World Wide Web (WWW '09*), 2009, 751-760.

[10] E. Scordilis, C. Papazachos, G. Karakaisis and V. Karakostas, Accelerating seismic crustal deformation before strong main shocks in adriatic and its importance for earthquake prediction, *J. Seismology*, 8, 2004, 57-70.

[11] T. Bleierand and F. Freund, Earthquake warning systems, *IEEE Spectrum*, 42 (12), 2005, 22-27.

[12] A. Java, X. Song, T. Finin and B. Tseng, Why we Twitter: understanding micro blogging usage and communities, *Proc. Ninth Web KDD and First SNA-KDD Workshop Web Mining and Social Network Analysis (WebKDD/SNA-KDD '07)*, 2007.

[13] G.L. Danah Boyd and S. Golder, Tweet, Tweet, Retweet: conversational aspects of Retweeting on Twitter, *Proc. 43rd Hawaii International Conference on System Sciences (HICSS-43)*, 2010.

[14] 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.

[15] B.O. Connor, R. Balasubramanyan, B.R. Rutledge and N.A. Smith, From Tweets to polls: linking text sentiment to public opinion time series, *Proc. International AAAI Conference Weblogs and Social Media*, 2010.

[16] M. Ebner and M. Schiefner, Microblogging - More than Fun?, *Proc. IADIS Mobile Learning Conference*, 2008, 155-159.

[17] A. Passant, T. Hastrup, U. Bojars and J. Breslin, Micro blogging: A semantic web and distributed approach, *Proc. Fourth Workshop Scripting for the Semantic Web (SFSW'08)*, 2008.