IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 2, Issue 2, Apr-May, 2014
ISSN: 2320 – 8791 (Impact Factor: 1.479)
www.ijreat.org

Real Time Social Data Analysis for Forecasting Nature Disasters

P.Balaji[1], G.Ilancheziapandian[2], A.Appandairaj[3]

[1]PG Student, Department of Computer Science and Engineering, Ganadipathy Tulsi's Jain Engineering College, Kaniyambadi, Vellore-632 102, Tamil Nadu, India.

[2]Professor, Department of Computer Science and Engineering, Ganadipathy Tulsi's Jain Engineering College, Kaniyambadi, Vellore-632 102, Tamil Nadu, India.

[3]Assistant Professor, Department of Computer Science and Engineering, Ganadipathy Tulsi's Jain Engineering College, Kaniyambadi, Vellore-632 102, Tamil Nadu, India.

Abstract

The online social network has millions of users all over the areas in the world, where the people collaborate and should share their information related to real world events. The status update which almost specifies what is happening around an individual and also around the individual's location. This small content with real world information when processed with some statistical tool may help us to predict a real world event. Inspired by the concept we propose a method to predict the cyclone formation in India by processing the tweets messages posted by the social users. Our system consist of two levels, first we create a classifier to filter out the messages which specify the occurrence of our event, in our case cyclone formation. In second level we apply statistical method over the generated data model to predict the occurrence of the cyclone in the future. We also create an alarm system for our registered users to deliver a notification during cyclone which helps saving many lives

Keywords: *Cyclone, prediction, Tweets, svm*

1. Introduction

To create an event notification or reporting system along with an event forecasting system using real world social data analysis. Twitter, a popular social networking service, has received much attention recently. This online social network is used by millions of people around the world to remain socially connected to their friends, family members, and coworkers through their computers and mobile phones. Twitter asks one question, "What's happening?" 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; 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. Since its launch on July 2006, Twitter users have increased rapidly. The number of registered Twitter users exceeded 100 million in April 2010. The service is still adding about 300,000 users per day.(1) currently, 190 million users use Twitter per month, generating 65 million tweets per day.(2)many researchers have published their studies of Twitter to date, especially during the past year. Most studies can be classified into one of three groups: first, some researchers have sought to analyze the network structure of Twitter. Second, some researchers have specifically examined characteristics of Twitter as a social medium. Third, some researchers and developers have tried to create new applications using Twitter. Twitter is categorized as a micro blogging service. Social networks are a form of blogging that enables users to send brief text updates or micro media such as photographs or audio clips. Social networking services other than Twitter include Tumbler, Plurk, identi.ca, and others.(3) our study, which is based on the real-time nature of one social networking service, is applicable to other social networking services, but we specifically examine Twitter in this investigation because of its popularity and data volume. An important characteristic that is common among social networking services is their real-time nature. Although blog users typically update their blogs once every several days, Twitter users write tweets several times in a single day. Users can know how 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. Several

important instances exemplify their real-time nature: in the case of an extremely strong earthquake in Haiti, many pictures were transmitted through Twitter. People were thereby able to know the circumstances of damage in Haiti immediately. In another instance, when an airplane crash-landed on the Hudson River in New York, the first reports were published through Twitter and tumbler. In such a manner, numerous update results in numerous reports related to events. They include social events such as parties, games, and presidential campaigns. They also include disastrous events such as cyclones, fires, traffic jams, thunder, heavy rainfall, and cyclones. Actually, Twitter is used for various real-time notifications such as that necessary for help during a large-scale fire emergency or live traffic updates.

I propose an event notification system that monitors tweets and delivers notification promptly using knowledge from the investigation. In this research, we take three steps: first, we crawl numerous tweets related to target events; second, we propose probabilistic models to extract events from those tweets and estimate locations of events; finally, we developed a cyclone reporting system that extracts cyclones from Twitter and sends a message to registered users.

2. System Description and Methodology

Twitter users message tweets several times in a single day. Users can know how 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. Several important instances exemplify their real-time nature: in the case of an extremely strong earthquake in Haiti, many pictures were transmitted through Twitter. People were thereby able to know the circumstances of damage in Haiti immediately. In another instance, when an airplane crash-landed on the Hudson River in New York, the first reports were published through Twitter and tumbler. In such a manner, numerous update results in numerous reports related to events. They include social events such as parties, games, and presidential meetings. They also include disastrous events such as cyclones, fires, traffic jams, riots, heavy rainfall, and cyclones. Actually, Twitter is used for various real-time notifications such as that necessary for help during a large-scale fire emergency or live traffic updates.

Our system crawl numerous tweets related to target events and use a probabilistic model to extract events from those tweets and estimate locations of event Finally, we developed an event notification or reporting system that extracts event from Twitter and sends a message to registered users. To obtain tweets on the target event precisely, we apply semantic analysis over a tweet to understand the concept of the posted tweets. We also use this tweet information to forecast a particular event by applying statistical tools over the extracted data model.

2.1 Methodology

- **Linear Regression**
- **Support vector machine**

2.1.1 Linear regression (LR)

Linear regression is mathematical statistical model is used for processing our predictive data model and finds the behavior outcomes. Modeling is done by finding the relationship between a scalar dependent variable y and explanatory variables denoted x which are extracted from the model and form a matrix or table. After finding the relationship, regression coefficient a and b is calculated using the following equation

$$a = \frac{n\sum_{t=1}^{n} x_t y_{t-} \sum_{t=1}^{n} x_t \sum_{t=1}^{n} y_t}{n\sum_{t=1}^{n} x_t^2 - \left(\sum_{t=1}^{n} x_t\right)^2} \quad (1)$$

$$b = \frac{1}{n}\left(\sum_{t=1}^{n} y_t - a\sum_{t=1}^{n} x_t\right) \quad (2)$$

And the a and b value is used to predict the behavior outcomes by applying them into the equation y=ax+b

Example:

The sales of a company (in million dollars) for each year are shown in the table below. Use the least squares regression line as a model to estimate the sales of the company in 2012.

Table 1: sample

x (year)	2005	2006	2007	2008	2009
y (sales)	12	19	29	37	45

We first change the variable x into t such that t = x - 2005 and therefore t represents the number of years after 2005. Using t instead of x makes the numbers smaller and therefore manageable.

The table of values becomes.

Table 2: new values

t (years after 2005)	0	1	2	3	4
y (sales)	12	19	29	37	45

We now use the table to calculate a and b included in the least regression line formula

Table 3: regression values

t	y	t y	t 2
0	12	0	0
1	19	19	1
2	29	58	4
3	37	111	9
4	45	180	16
Σx = 10	Σy = 142	Σxy = 368	Σx2 = 30

We now calculate a and b using the least square regression formulas for a and b.

a = (nΣt y - ΣtΣy) / (nΣt2 - (Σt)2) = (5*368 - 10*142) / (5*30 - 102) = 8.4

b = (1/n)(Σy - a Σx) = (1/5)(142 - 8.4*10) = 11.6 In 2012, t = 2012 - 2005 = 7 y=ax+b

The estimated sales in 2012 are: y = 8.4 * 7 + 11.6 = 70.4 million dollars

2.1.2 Support Vector Machine

A Support Vector Machine (SVM) performs classification by constructing an N-dimensional hyper plane that optimally separates the data into two categories. SVM models are closely related to neural networks. In fact, a SVM model using a sigmoid Kernel function is equivalent to a two-layer, perception neural network. N the parlance of SVM literature, a predictor variable is called an attribute, and a transformed attribute that is used to define the hyper plane is called a feature. The task of choosing the most suitable representation is known as feature selection. A set of features that describes one case (i.e., a row of predictor values) is called a vector. So the goal of SVM modeling is to find the optimal hyper plane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyper plane are the support vectors.

3. System Design

3.1. Architectural Diagram

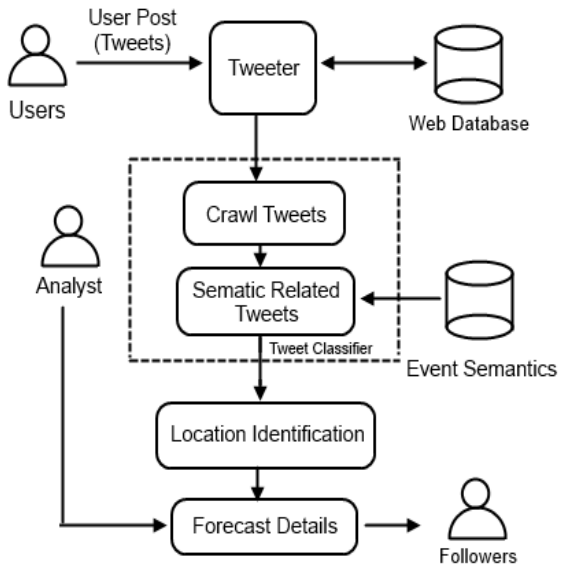

Fig 1.System Architecture

3.2 Crawl Tweets

Crawling the tweets provides the functionalities to completely search tweeter for post relative to the specified event. Crawler search posts with given semantic relationship relative to our specific event to understand the context of the users posted message. Crawler uses support vector machine algorithm for classification of our event relevant posted message from the generally posted messages.

3.3 Semantic Related Tweets

From the content containing the post messages can classify the semantic features related to our event. This can be achieved through support vector machine (SVM). It can be used to find the localization and the related event messages.

3.4 Location Identification

This provides the functionality to identify the location of our event by analyzing the data extracted from our event relevant posted message classified from the generally posted messages.

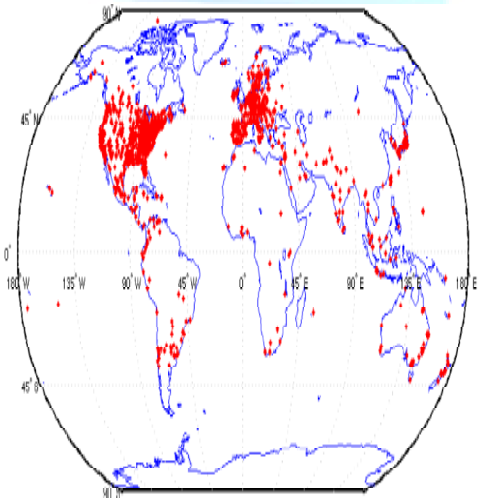

Fig 2.Twitter user map

4. Forecasting

We have to focus on the future prediction in this forecasting process by the past event reports which we collected from the Indian metrological Centre. It is used to predict a future event from the extracted event from the post along with their location and time. This is achieved through the statistical mathematic model for predicting the future event.

4.1 Event Notification System

For producing the event notification system we have to monitor all the past event reports which we collected from the Indian metrological centre. We have to focus the time, date, event and the location in the past report and apply the linear regression method to predict the future event and send such notification to all the users to prevent from such nature disasters.

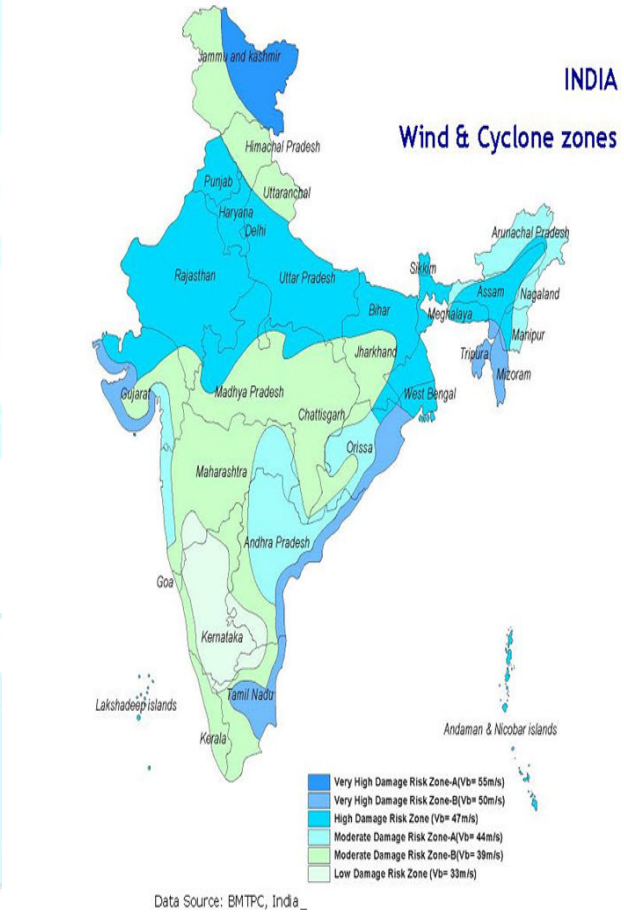

Fig 3.Cyclone zones in India

5. Conclusion

As described in this paper, we analyzed the real-time nature of Twitter; devoting particular attention based on sensory observations Location predicted methods such as support vector machine are used to estimate

the locations of events. As an application, we developed a cyclone reporting system, which is a unique approach to notify people promptly of a cyclone event. Social network has the real-time characteristics that distinguish it from other social media such as blogs and collaborative stickers. As described in this paper, we presented an example that leverages the real-time nature of Twitter to make it useful in solving an important social problem: Cyclones (natural disasters). I hoped it is given a sufficient and confident that this paper will provide some insight into the future integration of semantic analysis with useful data.

References

[1] Regional Metrological Center Cyclone and Research Department (MAUSAM) Chennai.

[2] A. Java, X. Song, T. Fining, and B. Tseng, "Why We Twitter: Understanding Micro blogging Usage and Communities,".

[3] B. Huber man, D. Romero, and F. Wu, "Social Networks that Matter: Twitter Under Microscope," ArXiv E-Prints, http:// arxiv.org/abs/0812.1045, Dec. 2008.

[4] H. Kawaka, C. Lee, H. Park, and S. Moon, "What is Twitter, A Social Network or A News Media?"

[5] G.L. Danah Boyd and S. Golder, "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter,"

[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] M.wieser,"The computer for the twenty first century,"scientific AM.1991e Twitter:Understanding Microblogging Usage and Communities,".

[9] Indian metrological centre and .www.indianmetrological.com mumbai, "Social Networks that Matter: Twitter Under Microscope," ArXiv E-Prints, http:// arxiv.org/abs/0812.1045, Dec. 2008.

[10] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, A Social Network or A News Media?"