ISSN 2395-1621

Earthquake Reporting System Development by Tweet Analysis using Naive Bayes Algorithm

#1Nimbalkar Priyanka, #2Autade Poonam, #3Sawant Pranjali, #4Mirge Akanksha, #5Prof. A.C. Lomte

1priyankanimbalkar27@gmail.com
2punamautade@gmail.com
3pranjali99sawant@gmail.com
4mirgeakanksha@gmail.com

#1234Department of Computer Engineering
#5Prof. Department of Computer Engineering

JSPM's, BSIOTR, Wagholi, Pune.
Savitiribai Phule Pune University India.

ABSTRACT

We investigate the real-time interaction of events such as earthquakes in Twitter and propose an algorithm to monitor tweets and to detect a target event. To detect a target event, we devise a classifier of tweets based on features such as the keywords in a tweet, the number of words, and their context. We develop an earthquake reporting system for use for users. Our system detects earthquakes promptly and notification is delivered much faster to all connected user. Here we use the Navies Bayes mining algorithm for analysis of users twits if any emergency situation, if any emergency then we also generates the alert system to the all connected users.

Keywords: Earthquakes, Twitted Analysis, Navies Bayes, Mining

ARTICLE INFO

Article History

Received: 19th October 2016

Received in revised form : 20th October 2016

Accepted: 24th October 2016

Published online : 25th October 2016

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 active users every month while the total number of users is almost twice as much. Twitter users create more than 500 million tweets per day. As a micro blogging service Twitter lets people share 140 character long texts called tweets. According to a survey from early 2012 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. 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. 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 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.

II. LITERATURE SURVEY

M. Ebner and M. Schiefner, "Microblogging - More than Fun?" , 2008, In this paper he describe how we used Twitter with students of English at the Distant College of Shanghai Jiao Tong University. We analyze the students' messages and show how the usage of Twitter trained communicative and cultural competence. [1]

T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors," , 2010., The KNN based classifier conducts pre-labeling to collect global coarse evidence across tweets while the CRF model conducts sequential labeling to capture fine-grained information encoded in a tweet. [2]

T. Bleier and F. Freund, "Earthquake [earthquake warning systems]," , 2005. In this paper, the author explains the earthquake warning system. [3]

L. Backstrom, J. Kleinberg, R. Kumar, and J. Novak, "Spatial Variation in Search Engine Queries,", 2008. This model provides not only an estimate of a query's geographic center, but also a measure of its spatial dispersion, indicating whether it has highly local interest or broader regional or national appeal. he also show how variations on our model can track geographically shifting topics over time, annotate a map with each location's "distinctive queries," and delineate the "spheres of influence" for competing qeries in the same general domain. [4]

M. Cataldi, L. Di Caro, and C. Schifanella, "Emerging Topic Detection on Twitter Based on Temporal and Social Terms Evaluation," 2010, In this paper he recognize this primary role of Twitter and he propose a novel topic detection technique that permits to retrieve in real-time the most emergent topics expressed by the community. First, he extract the contents (set of terms) of the tweets and model the term life cycle according to a novel aging theory intended to mine the emerging ones. A term can be defined as emerging if it frequently occurs in the specified time interval and it was relatively rare in the past. Moreover, considering that the importance of a content also depends on its source, he analyze the social relationships in the network with the well-known Page Rank algorithm in order to determine the authority of the users. [5]

H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, A Social Network or A News Media?" 2010, he analyzed the tweets of top trending topics and reported on their temporal behavior and user participation. He has classified the trending topics based on the active period and the tweets and show that the majority (over 85%) of topics are headline news or persistent news in nature. A closer look at retweets reveals that any retweeted tweet is to reach an average of 1, 000 users no matter what the number of followers is of the original tweet. Once retweeted, a tweet gets retweeted almost instantly on next hops, signifying fast diffusion of information after the 1st retweet. [6]

B. Huberman, D. Romero, and F. Wu, "Social Networks that Matter: Twitter Under the Microscope," 2008. Scholars, advertisers and political activists see massive online social networks as a representation of social interactions that can be used to study the propagation of ideas, social bond dynamics and viral marketing, among others. But the linked structures of social networks do not reveal actual interactions among people. Scarcity of attention and the daily rythms of life and work makes people default to interacting with those few that matter and that reciprocate their attention. A study of social interactions within Twitter reveals that the driver of usage is a sparse and hidden network of connections underlying the "declared" set of friends and followers. [7]

T. Rattenbury, N. Good, and M. Naaman, "Towards Automatic Extraction of Event and Place Semantics from Flickr Tags," 2007, He analyze two methods inspired by well-known burst-analysis techniques and one novel method: Scale-structure Identification. He evaluate the methods on a subset of Flickr data, and show that our

Scale-structure Identification method outperforms the existing techniques. The approach and methods described in this work can be used in other domains such as geo-annotated web pages, where text terms can be extracted and associated with usage patterns. [8]

III. PROBLEM STATEMENT

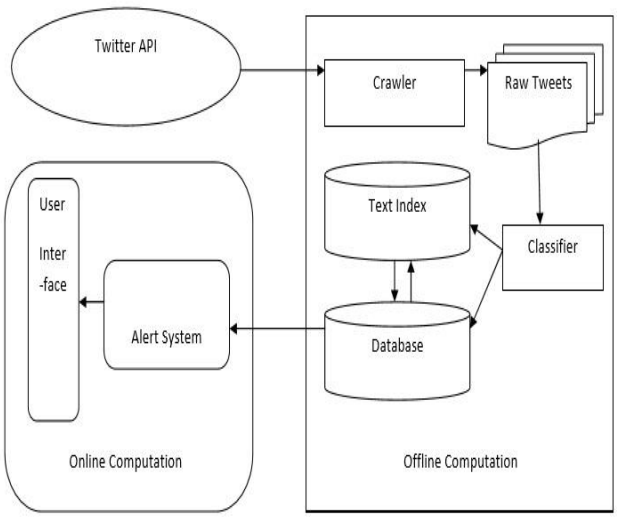
To investigate real-time interaction of events such as earthquakes in Twitter and propose an algorithm to monitor tweets and to detect a target event with semantic analyses applied to tweets to classify them into positive and negative classes based on sensory observations.

IV. EXISTING SYSTEM

In existing system Some issues involved are data privacy, private aggregation communities, and building personalised views and aggregation services for public updates. The linked structures of social networks do not reveal actual interactions among people. And It is of very high complexity. Its takes more time to deliver the messages.

JMA broadcast announcements consume more time to distribute earthquake early warning information's among the advanced users world.

V. PROPOSED SYSTEM



1) Tweet search API window collects tweets regarding events I large scale.

2) We crawl no of tweets using tweeter crawler to find out useful Tweets and scripted to processing.

3) Processed twitter distinguished between "+ class and - class" by using algorithm.

4) From positive class we find out event detection and location using Mining algorithm.

5)Lastly we improve an actual time tweeter operator's method to report real time event detection and analysis of earthquake reporting.

VI. MATHEMATICAL MODEL

System Description:
Input:
Fuction Twits ()
T : Text input.
S : Symbols.
R : Remove unwanted symbols.
C : Classification
D : Store data base.
F : Final Result
Output:
Stored user opinion feedback to the Database.

Input
Function Preprocessing (id, request, feedback)
ID : unique id for each twits.
Request : User request to the server.
Feedback : user opinion feedback.
Output:
User Feedback will preprocess.

Input
Function Classification (id, feedback)
ID : unique id for each feedback.
Feedback : user opinion feedback.
Output:
Classification of final result will display.

VII. CONCLUSION

We investigate the real-time nature of Twitter for event detection. Twitter has real time nature which can be useful in solving social problems such as natural disasters. Twitter user is considered as a sensor and detection of events based on the sensory observations.

Semantic analysis is applied to classify tweets. We are providing the alert messages to users through SMS and e-

mail. As an application we are developing an earthquake reporting system.

REFERENCES

[1] M. Ebner and M. Schiefner, "Microblogging - More than Fun?"Proc. IADIS Mobile Learning Conf.,pp. 155-159, 2008

[2] 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

[3] T. Bleier and F. Freund, "Earthquake [earthquake warning systems],"IEEE Spectrum,vol. 42, no. 12, pp. 22-27, Dec. 2005.

[4] L. Backstrom, J. Kleinberg, R. Kumar, and J. Novak, "Spatial Variation in Search Engine Queries,"Proc. 17th Int'l Conf. World Wide Web (WWW '08),pp. 357-366, 2008.

[5] M. Cataldi, L. Di Caro, and C. Schifanella, "Emerging Topic Detection on Twitter Based on Temporal and Social Terms Evaluation," Proc. 10th Int'l Workshop Multimedia Data Mining(MDMKDD '10), pp. 1-10, 2010.

[6] 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.

[7] 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.

[8] T. Rattenbury, N. Good, and M. Naaman, "Towards Automatic Extraction of Event and Place Semantics from Flickr Tags," Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '07), pp. 103-110, 2007.

