ISSN 2395-1621

Traffic Detection from Twitter Stream Analysis

[#1]Prof. Ms.Rupali Shinganjude, [#2]Pratiksha Jambhulkar, [#3]Isha Shende, [#4]Divya Thembekar [#5]Bhagyashree Sawarkar

[1]rupalishinganjude@gmaill.com
[2]pratikshajambhulkar660@gmail.com
[3]isha.shende26@gmail.com
[4]divyatembhekar@gmail.com
[4]sawarkarbhagyashri@gmail.com

[#1]Professor, Department of Information Technology
[#2345]Student, Department of Information Technology
Priyadarshini Bhagwati College of Engineering
Nagpur

ABSTRACT

Social networks have been recently employed as a source of information for event detection with particular reference to road traffic congestion and car accidents. Netizens these days are getting more and more active regarding their everyday activity. Somehow their addiction can result in more productive application. One such is regarding Sedan drivers posting every minute detail of traffic and route. In this project, user gets public traffic tweets from Twitter. Apply tokenize, remove stop words and apply stemming to tweet. Detect the causes on that tweet based on Words that are stored into database. User logs in via android application, User can search path (Ex-Katraj to Wakad). Web portal gets array of latitude and longitude and sends return to traffic between that array with causes. Alternate path displayed with traffic.

ARTICLE INFO

Article History

Received: 14th March 2017

Received in revised form :

14th March 2017

Accepted: 16th March 2017

Published online :

17th March 2017

Keywords: web portal, tokenize, social networks, stemming, tweets.

I. INTRODUCTION

A few years ago information about events of any kind came almost exclusively from specialized information. With the increase of social networks' usage, information has become more decentralized; as each user become a consumer/provider of information. This paradigm becomes relevant when it is crucial to disseminate information in real-time about an event. A good example where social networks had a great impact in disseminating events in real-time and namely for disaster relief, was by the time of the Haiti earthquake in January 2010, when posts to Facebook, Flickr, YouTube and Twitter were numerous [1]. Research revealed that those messages sent during the Haiti earthquake that were considered "actionable" and more useful than the others were generally the ones that included locations [2]. Moreover, Twitterers themselves are more likely to pass along, or re-tweet, messages that include geo-location and situational updates [3], indicating that Twitterers find such messages important. Location estimation using the content in social networks has its own challenges, namely the uncontrolled and the limited content, are two important drawbacks. In fact, one of the most important limitations in extracting useful information from a tweet is its 140-character limit. Additionally, users can make spelling mistakes, and use varying conventions and abbreviations; the quality of content is not as good as in news articles. In addition to the credibility of the content, the credibility of user profiles and their geo-references are questionable as well. Not all users have to provide their GPS locations, or their city of residence in their profiles. As a result, the most obvious problem is uncertainty and the lack of rich and reliable data. Moreover, tweet density depends heavily on the population and Twitter usage in a region. That means, if an event occurs in a small town, the number of Twitterers posting tweets about that event would be considerably lower than the number of users if the event would have occurred in a huge capital. Especially, once an event is broadcasted on the mediaor shared among people via phone calls or Internet, people from anywhere can post a tweet about it, making the location center more obscure to detect. Therefore, a practical problem to solve is how to normalize these tweets according to the user density of the area, and obtain a fair result to compare and evaluate tweets from different locations. In the scope of this work, we focus on the analysis and detection of traffic events, where information disseminated by social networks can have an important role. Social networks can be seen as a mechanism which allows the detection of very small events, such as a damaged car in a side street. In addition to that, the interval between the occurrence of a traffic incident and the publication of a tweet about such incident usually tends to be much less, when compared to the time required for a news agency to broadcast about such particular event. Extracting useful information from tweets presents some challenges, as described by the authors in [4] and [5]. Taking into account that the information is completely unstructured, tweets can contain grammatical errors and abbreviations. Each user has its own style of writing, the information can be incomplete (e.g. a street name with no more information about city, etc.), false or not credible. More specifically, and to what this work is concerned, determining geographic interpretations for place names, or toponyms, involves resolving multiple types of ambiguity. From our point of view, and considering traffic related tweets, the relevance of a tweet message is directly proportional with the number of similar tweets posted by others in a short period of time. Moreover, in order to automatically recognize the relevance of tweet message for the purpose of detecting traffic events is a challenging task (e.g. the tweet “Excessive speed is the main cause of car accidents in Liverpool” is not relevant for the purpose of the work to be addressed here, but it can be considered as a traffic related tweet). social networks have been recently employed as a source of information for event detection with particular reference to road traffic congestion and car accidents. Netizens these days are getting more and more active regarding their everyday activity. Somehow their addiction can result in more productive application. One such is regarding Sedan drivers posting every minute detail of traffic and route. In this project, user gets public traffic tweets from Twitter. Apply tokenize, remove stop words and apply stemming to tweet. Detect the causes on that tweet based on Words that are stored into database. User logs in via android application, User can search path .Web portal gets array of latitude and longitude and sends return to traffic between that array with causes. Alternate path displayed with traffic. The traffic detection system was employed for real-time monitoring of several areas of the Italian road network, allowing for detection of traffic events almost in real time, often before online traffic news web sites. Twitter users to external attack servers. To cope with malicious tweets, several Twitter spam detection schemes have been proposed. These schemes can be classified into account feature-based, relation feature-based, and message feature based schemes.

II. Literature Survey

This system is generally based on get Public traffic tweets from twitter and Apply tokenization, remove stop words and apply stemming to a particular tweet. We have maintained lists of causes (eg. Accidents, Traffic Jams, Vehicle breakdowns, etc.) Twitter is, in nature, a good resource for detecting events in real-time. The real-time detection capability allows obtaining reliable information about traffic events in a very short time, often before online news web sites and local newspapers.

III. Problem statement

The developed system was installed and tested for the real-time monitoring of several areas of the road network. It means of the analysis of the Twitter stream coming from those areas. The goal is to perform a continuous monitoring of frequently busy roads and highways in order to detect possible traffic events in real-time or even in advance with respect to the traditional news media

IV. Proposed system

We have proposed a system for real-time detection of traffic-related events from Twitter stream analysis. The system is also able to discriminate if a traffic event is due to an external cause, such as football match, procession and manifestation, or not. Web portal gets array of latitude and longitude and sends return to traffic between that array with causes. Alternate path displayed with traffic. Our traffic detection system based on Twitter streams analysis is presented .and it detects the traffic events in real-time. Also Web part gets array of latitude and longitude of searched path and then the latitude and longitude of the traffic is compared with searched path with their causes. After comparing the longitude and latitude having traffic, it is displayed on the maps of Android device. Haversine method used to calculate distance between two latitude-longitude pairs, Triangulation for getting GPS Location.

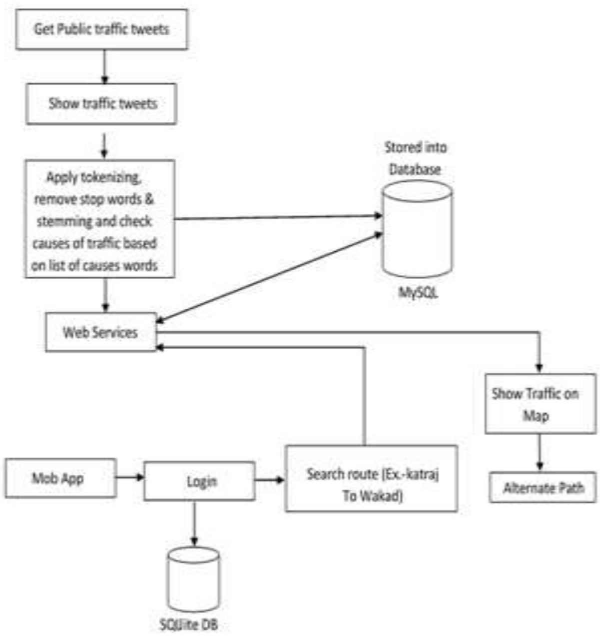


Fig 1. proposed system architecture

V. Implementation

Social networks have been recently employed as a source of information for event detection with particular reference to road traffic congestion and car accidents. Netizens these days are getting more and more active regarding their everyday activity. Somehow their addiction can result in more productive application. One such is regarding Sedan drivers posting every minute detail of traffic and route. In this project, user gets public traffic tweets from Twitter. Apply tokenize, remove stop words and apply stemming to tweet. Detect the causes on that tweet based on Words that are stored into database. User logs in via android application, User can search path .Web portal gets array of latitude and longitude and sends return to traffic between that array with causes. Alternate path displayed with traffic. The traffic detection system was employed for real-time monitoring of several areas of the Italian road network, allowing for detection of traffic events almost in real time, often before online traffic news web sites. Twitter users to external attack servers. To cope with malicious tweets, several Twitter spam detection schemes have been proposed. These schemes can be classified into account feature-based, relation feature-based, and message feature based schemes.

Functionality :

Web Part-

1)Get Public traffic tweets from twitter.

2)Apply tokenization, remove stop words and apply stemming to a particular tweet.

3)We have maintained lists of causes (eg. Accidents, Traffic, Jams, Vehicle breakdowns, etc.) and we check these causes in that particular tweet.

4)Tweets and their causes are stored into database.

5)Web part gets array of latitude and longitude of searched path and then the latitude and longitude of the traffic is compared with searched path with their causes.

6) After comparing the longitude and latitude having traffic, it is displayed on the maps of Android device.

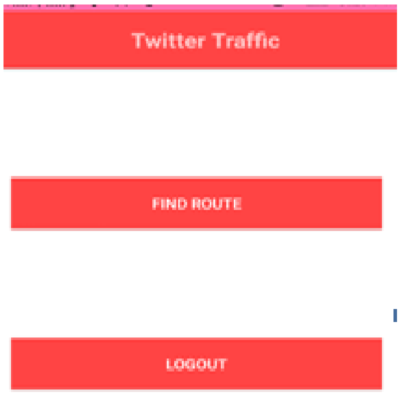
Android Part-

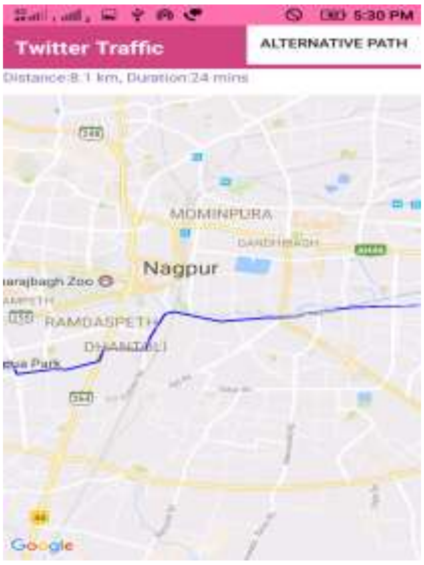
1)Login and registration of user.

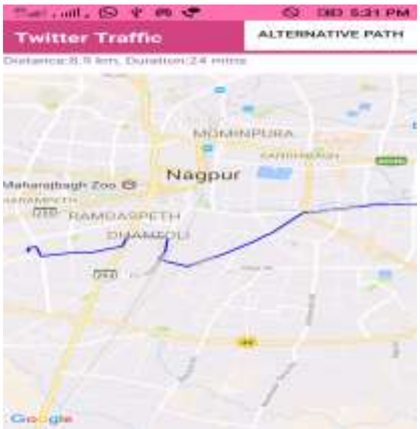
2)User can search path (Ex- Katraj to Wakad) and they can see traffic on that way.

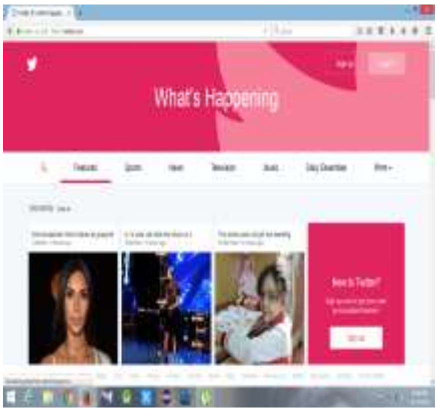
3)And user can select the alternate path.

VI. Experimental Results









VII. CONCLUSION.

In this system we have proposed a system for real-time detection of traffic-related events from Twitter stream analysis and we have also maintained lists of causes (eg. Accidents, Traffic, Jams, Vehicle breakdowns, etc.). We check these causes in that particular tweet: Showing traffic tweet with causes and Showing traffic between two points

REFRENCES

[1] H. Gao, G. Barbier and R. Goolsby, "Harnessing the Crowdsourcing Power of Social Media for Disaster Relief," IEEE Intelligent Systems,vol. 26, no. 3, pp. 10-14, 2011.

[2] R. Munro, "Subword and spatiotemporal models for identifying actionable information in Haitian Kreyol," in Proceedings of the Fifteenth Conference on Computational Natural Language Learning, Portland, 2011.

[3] S. Vieweg, A. Hughes, K. Starbird and L. Palen, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, 2010.

[4] A. Bifet and E. Frank, "Sentiment knowledge discovery in twitter streaming data," Discovery Science, pp. 1-15.

[5] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan and N. A. Smith, "Part-of-speech tagging for Twitter: annotation, features, and experiments," in Proceedings of the 49th Annual Meeting of the Association forComputational Linguistics: Human Language Technologies: short papers, 2011.

[6] R. Costa, P. Figueiras, P. Maló, M. Jermol and K. Kalaboukas, "MobiS -Personalized Mobility Services for energy efficiency and security through advanced Artificial Intelligence techniques," in 5th KES International Conference on Intelligent Decision Technologies,Sesimbra, 2013.

[7] N. Wanichayapong, W. Pruthipunyaskul, W. Pattara-Atikom and P. Chaovalit, "Social-based traffic information extraction and classification," in International Conference on Telecommunications,2011.

[8] A. Schulz, P. Ristoski and H. Paulheim, "I See a Car Crash: Real-time Detection of Small Scale Incidents in Microblogs," in ESWC 2013 Satellite Events: Revised Selected Papers, 2013.

[9] [Online]. Available: http://twitraffic.co.uk/.

[10] S. A., C. J., H. D., L. D. and P. F., "AT&T at TREC-7," in Proceedings of the Seventh Text REtrieval Conference, 1999.

[11] H. Paulheim and J. Fümkranz, "Unsupervised generation of data mining features from linked open data," in Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics, 2012.

[12] M. Walther and M. Kaisser, "Geo-spatial event detection in the twitter stream," in Proceedings of the 35th European conference on Advances in Information Retrieval, Moscow, 2013.

[13] R. Li, K. Lei, R. Khadiwala and K. Chang, "TEDAS: A Twitter-based Event Detection and Analysis System," in IEEE 28th International Conference on Data Engineering, Washington, 2012.

[14] F. Abel, C. Hauff, G. J. Houben, R. Stronkman and K. Tao, "Semantics+ filtering+ search= twitcident. exploring information in social web streams," in Proceedings of the 23rd ACM conference on Hypertext and social media, 2012.

[15] J. Rogstadius, M. Vukovic, C. A. Teixeira, V. Kostakos, E. Karapanos and J. A. Laredo, "CrisisTracker: Crowdsourced social media curation for disaster awareness," IBM Journal of Research and Development, vol. 57(5), pp. 4-1, 2013.

[16] O. Okolloh, "Ushahidi, or'testimony': Web 2.0 tools for crowdsourcing crisis information," Participatory learning and action, vol. 59(1), pp. 65-70, 2009.

[17] J. Yin, A. Lampert, M. Cameron, B. Robinson and R. Power, "Using Social Media to Enhance Emergency Situation Awareness," IEEE Intelligent Systems, pp. 52-59, 2012.

[18] G. P. C. Fung, J. X. Yu, P. S. Yu and H. Lu, "Parameter free bursty events detection in text streams," in Proceedings of the 31st international conference on Very large data bases, 2005.

[19] R. Guillén, "GeoParsing Web Queries," in Advances in Multilingual and Multimodal Information Retrieval, Berlin, Springer-Verlag, 2008, pp. 781-785.

[20] Rapid-I GmBH, "RapidMiner," 2012. [Online]. Available: http://rapidi.com. [Accessed 3 September 2012].

[21] G. Salton, A. Wong and C. Yang, “A vector space model for automatic indexing,” Communications of the ACM, pp. 613-620, 1975.

