International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 23 Issue 7 –OCTOBER 2016.

Detection of Traffic on Twitter in Real-Time Social Network

D.SRILAKSHMI BAI[#1] and T.KARUNALATHA[*2]

[#] *M.TECH (CSE), NIMRA INSTITUTE OF SCIENCE AND TECHNOLOGY, A.P., INDIA.*

[*] *ASSISTANT PROFESSOR, DEPT. OF COMPUTER SCIENCE & ENGINEERING, NIMRA INSTITUTE OF SCIENCE AND TECHNOLOGY, A.P., INDIA.*

***Abstract*— TWITTER, is a popular micro blogging 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. Social networks is a source of information for event detection such as road traffic congestion and car accidents. Existing system present a real-time monitoring system for traffic event detection from twitter. The system fetches or collect tweets from twitter and then processes tweets using text mining techniques. Lastly performs the classification of tweets. The system aim is to assign the appropriate class label to each tweet, whether it is related to a traffic event or not. System used the support vector machine as a classification model. In this paper, we present a real-time monitoring system intended for traffic occasion detection coming from Twitter stream analysis. The system fetches tweets coming from Twitter as per a several search criteria; methods tweets, by applying textual content mining methods; last but not least works the classification of twitter posts. The goal is to assign suitable class packaging to every tweet, because related with an activity of traffic event or perhaps not. The traffic recognition system or framework was utilized for real-time monitoring of various areas of the street network, taking into account detection of traffic occasions just almost in actual time, regularly before on-line traffic news sites.**

Index Terms*— *Traffic Event Detection, Tweet Classification, Text Mining, Social Networks.

I. INTRODUCTION

Twitter is categorized as a micro blogging service. Micro blogging is a form of blogging that enables users to send brief text updates or micro media such as photographs or audio clips. Micro blogging services other than Twitter include Tumblr, Plurk, Jaiku, identi.ca, and others.3 Our study, which is based on the real-time nature of one social networking service, is applicable to other microblogging services, but we specifically examine Twitter in this study because of its popularity and data volume. An important characteristic that is common among microblogging 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 tumblr. In such a manner, numerous update results in numerous reports related to events. They include social events such as parties, baseball games, and presidential campaigns. They also include disastrous events such as storms, fires, traffic jams, riots, heavy rainfall, and earthquakes. Actually, Twitter is used for various real-time notification such as that necessary for help during a large-scale fire emergency or live traffic updates. Twitter is prone to malicious tweets containing URLs for spam, phishing, and malware distribution. Conventional Twitter spam detection schemes utilize account of features such as the ratio of tweets containing URLs and the account creation date, or relation features in the Twitter graph[1][2]. These detection schemes are ineffective against feature fabrications or consume much time and resources. Conventional suspicious URL detection schemes utilize several features including lexical features of URLs, URL redirection, HTML content, and dynamic behavior. However, evading techniques such as time-based evasion and crawler evasion exist[3]. In this paper, we propose an intelligent system, based on text mining and machine learning algorithms, for realtime detection of traffic events from Twitter stream analysis. The system, after a feasibility study, has been designed and developed from the ground as an event-driven infrastructure, built on a Service Oriented Architecture (SOA)[4]. The system exploits available technologies based on state-of-the-art techniques for text analysis and pattern classification. These technologies and techniques have been analyzed, tuned, adapted, and integrated in order to build the intelligent system. In particular, we present an experimental study, which has been performed for determining the most effective among different state-of-the-art approaches for text classification. The chosen approach was integrated into the final system and used for the on-the-field real-time detection of traffic events. In the existing system attackers use shortened malicious URLs that redirect 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. Account feature-based schemes use

the distinguishing features of spam accounts such as the ratio of tweets containing URLs, the account creation date, and the number of followers and friends. However, malicious users can easily fabricate these account features. The relation feature-based schemes rely on more robust features that malicious users cannot easily fabricate such as the distance and connectivity apparent in the Twitter graph. Extracting these relation features from a Twitter graph, however, requires a significant amount of time and resources as a Twitter graph is tremendous in size. The message feature-based scheme focused on the lexical features of messages. However, spammers can easily change the shape of their messages. A number of suspicious URL detection schemes have also been introduced.

II. PROPOSED SYSTEM

In our paper, we present a real-time monitoring system for traffic occasion detection from Twitter stream analysis. The system fetches tweets from Twitter as per a several search criteria; procedures tweets, by applying text mining methods; lastly performs the classification of tweets. The aim is to assign suitable class label to every tweet, as related with an activity of traffic event or not. The traffic detection system or framework was utilized for real-time monitoring ofseveral areas of the Italian stree network, taking into consideration detection of traffic events just almost in real time, regularly before online traffic news sites. We employed the support vector machine as a classificatio model, furthermore, we accomplished an accuracy value of 95.75% by tackling a binar classification issue (traffic versus nontraffic tweets). We were also able to discriminate if traffic is caused by an external event or not, by solving a multiclass classification problem and obtaining accuracy value of 88.89%.

A. Advantages of Proposed System:

- Tweets are up to 140 characters, enhancing the real-time and news-oriented nature of the platform. In fact, the lifetime of tweets is usually very short, thus Twitter is the social network platform that is best suited to study SUMs related to real-time events.
- Each tweet can be directly associated with meta information that constitutes additional information.
- Twitter messages are public, i.e., they are directly available with no privacy limitations. For all of these reasons, Twitter is a good source of information for real time event detection and analysis.
- Moreover, the proposed system could work together with other traffic sensors (e.g., loop detectors, cameras, infrared cameras) and ITS monitoring systems for the detection of traffic difficulties, providing a low-cost wide coverage of the road network, especially in those areas (e.g., urban and suburban) where traditional traffic sensors are missing.
- It performs a multi-class classification, which recognizes non-traffic, traffic due to congestion or crash, and traffic due to external events.

III. LITERATURE SURVEY

A. A.Text Detection and Recognition on Traffic Panels From Street-Level Imagery Using Visual Appearance Authors: Álvaro González, Luis M. Bergasa.

Traffic sign detection and recognition has been thoroughly studied for a long time. However, traffic panel detection and recognition still remains a challenge in computer vision due to its different types and the huge variability of the information depicted in them. This paper presents a method to detect traffic panels in streetlevel images and to recognize the information contained on them, as an application to intelligent transportation systems (ITS). The main purpose can be to make an automatic inventory of the traffic panels located in a road to support road maintenance and to assist drivers. Our proposal extracts local descriptors at some interest key points after applying blue and white color segmentation. Then, images are represented as a “bag of visual words” and classified using Naïve Bayes or support vector machines. This visual appearance categorization method is a new approach for traffic panel detection in the state of the art. Finally, our own text detection and recognition method is applied on those images where a traffic panel has been detected, in order to automatically read and save the information depicted in the panels. We propose a language model partly based on a dynamic dictionary for a limited geographical area using a reverse geo coding service. Experimental results on real images from Google Street View prove the efficiency of the proposed method and give way to using street-level images for different applications on ITS.

B. B.What’s Happening: A Survey of Tweets Event Detection Author: Amina Madani, Omar Boussaid, Djamel Eddine Zegour and Algiers, Algeria

Twitter is now one of the main means for spread of ideas and information throughout the Web. Tweets discuss different trends, ideas, events, and so on. This gave rise to an increasing interest in analyzing tweets by the data mining community. Twitter is, in nature, a good resource for detecting events in real-time. In this survey paper, authors have presented four challenges of tweets event detection: health epidemics identification, natural events detection, trending topics detection, and sentiment analysis. These challenges are based mainly on clustering and classification. We review these approaches by providing a description of each one. These last years have been marked by the emergence of microblogs. Their rates of activity reached some levels without precedent. Hundreds of millions of users are registered in these microblogs as Twitter. They exchange and tell their last thoughts, moods or activities by tweets in some words.

C. C.Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors Authors: Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo

Twitter, a popular micro blogging service, has received much attention recently. An important characteristic of Twitter is its real-time nature. For example, when an

earthquake occurs, people make many Twitter posts (tweets) related to the earthquake, which enables detection of earthquake occurrence promptly, simply by observing the tweets. As described in this paper, 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. Subsequently, we produce a probabilistic spatiotemporal model for the target event that can find the center and the trajectory of the event location. We consider each Twitter user as a sensor and apply Kalman filtering and particle filtering, which are widely used for location estimation in ubiquitous/pervasive computing. The particle filter works better than other compared methods in estimating the centers of earthquakes and the trajectories of typhoons. As an application, we construct an earthquake reporting system in Japan. Because of the numerous earthquakes and the large number of Twitter users throughout the country, we can detect an earthquake by monitoring tweets with high probability (96% of earthquakes of Japan Meteorological Agency (JMA) seismic intensity scale 3 or more are detected). Our system detects earthquakes promptly and sends e-mails to registered users. Notification is delivered much faster than the announcements that are broadcast by the JMA.

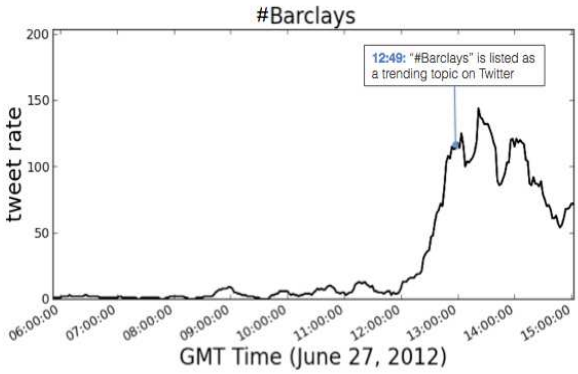
IV. REALATED WORK

Twitter is an interesting example of the most recent type of social media. Numerous researchers have examined Twitter. Regarding similar research to that presented in this paper, some researchers have attempted topic detection using Twitter.

First, some researchers specifically examine the network structure of Twitter and investigate Twitter network features of various kinds. Java et al. analyzed Twitter as early as 2007. They described the social network of Twitter users and investigated the motivations of Twitter users [2]. Haewoon et al. crawled a vast amount of Twitter data, analyzed the Twitter follower-following topology and ranked users by Pagerank [4]. Huberman et al. analyzed more than 300 thousand users. They discovered that the relation between friends (defined as a person to whom a user has directed posts using an "@" symbol) is the key to understanding interaction in Twitter [3].

Second, some researchers have examined characteristics of Twitter as social media. Recently, Boyd et al. have continued their investigation of retweet activity, which is the Twitter-equivalent of e-mail forwarding, by which users post messages that were originally posted by others [5]. Tumasjan et al. crawled many tweets referring to the election in Germany and attempted to predict the results of the election: which political parties would win the election [6]. Oconnor extracts public opinion from Twitter using sentiment analysis and reports the possibility of using a proposed method instead of polls.

Third, some studies elucidate the benefits of novel applications of Twitter: Ebner and Schiefner establish a microblogging community and studies how to use Twitter as a tool for mobile e-learning. The integration of the Semantic Web and microblogging was described in a previous report in which a distributed architecture is proposed and the contents are aggregated. In contrast to the small number of academic studies of Twitter, numerous Twitter applications exist. Some are used for analyses of Twitter data. For example, Tweettronics14 provides an analysis of tweets related to brands and products for marketing purposes. It can classify positive and negative tweets, and it can identify influential users. The classification of tweets might be done similarly to our algorithm. Web2express Digest15 is a website that auto discovers information from Twitter streaming data to find real-time interesting conversations. It also uses natural language processing and sentiment analysis to discover interesting topics, as we do in our study.



V. CONCLUSION & FUTURE SCOPE

In this paper, we have proposed a system for real-time detection of traffic-related events from Twitter stream analysis. The system, built on a SOA, is able to fetch and classify streams of tweets and to notify the users of the presence of traffic events. Furthermore, 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. We have exploited available software packages and state-of the- art techniques for text analysis and pattern classification. These technologies and techniques have been analyzed, tuned, adapted and integrated in order to build the overall system for traffic event detection. Among the analyzed classifiers, we have shown the superiority of the SVMs, which have achieved accuracy of 95.75%, for the 2-class problem, and of 88.89% for the 3-class problem, in which we have also considered the *traffic due to external event* class. The best classification model has been employed for real time monitoring of several areas of the Italian road network. We have shown the results of a monitoring campaign, performed in September and early October 2014. We have discussed the capability of the system of detecting traffic events almost in real time, often before online news web sites and local newspapers.

As future work, we are planning to integrate our system with

an application for analyzing the official traffic news web sites, so as to capture traffic condition notifications in real-time. Thus, our system will be able to signal traffic-related events in the worst case at the same time of the notifications on the web sites. Further, we are investigating the integration of our system into a more complex traffic detection infrastructure. This infrastructure may include both advanced physical sensors and social sensors such as streams of tweets. In particular, social sensors may provide a low-cost wide coverage of the road network, especially in those areas (e.g., urban and suburban) where traditional traffic sensors are missing.

REFERENCES

[1] F. Atefeh and W. Khreich, "A survey of techniques for event detection in Twitter," Comput. Intell., vol. 31, no. 1, pp. 132–164, 2015.
[2] P. Ruchi and K. Kamalakar, "ET: Events from tweets," in Proc. 22nd Int. Conf. World Wide Web Comput., Rio de Janeiro, Brazil, 2013, pp. 613–620.
[3] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in Proc. 7th ACM SIGCOMM Conf. Internet Meas., San Diego, CA, USA, 2007, pp. 29–42.
[4] G. Anastasi *et al.*, "Urban and social sensing for sustainable mobility in smart cities," in *Proc. IFIP/IEEE Int. Conf. Sustainable Internet ICT Sustainability*, Palermo, Italy, 2013, pp. 1–4.
[5] A. Rosi *et al.*, "Social sensors and pervasive services: Approaches and perspectives," in *Proc. IEEE Int. Conf. PERCOM Workshops*, Seattle, WA, USA, 2011, pp. 525–530.
[6] T. Sakaki, M. Okazaki, and Y.Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 4, pp. 919–931, Apr. 2013.
[7] J. Allan, *Topic Detection and Tracking: Event-Based Information Organization*. Norwell, MA, USA: Kluwer, 2002.
[8] K. Perera and D. Dias, "An intelligent driver guidance tool using location based services," in *Proc. IEEE ICSDM*, Fuzhou, China, 2011, pp. 246–251.
[9] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving information from social sensors," in *Proc. IEEE Int. Conf. CYBER*, Bangkok, Thailand, 2012, pp. 221–226.
[10] B. Chen and H. H. Cheng, "A review of the applications of agent technology in traffic and transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 485–497, Jun. 2010.
[11] A. Gonzalez, L. M. Bergasa, and J. J. Yebes, "Text detection and recognition on traffic panels from street-level imagery using visual appearance," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 228–238, Feb. 2014.