International Journal of Scientific Research in Computer Science and Engineering
Vol.4, Issue.3, pp.34-37, June (2016)

Review Paper

E-ISSN:2320-7639

Monitoring and Analysis of Real time detection of traffic from twitter stream analysis

A. Jalaparthi[1*], A.S. Kumar[2]

[1]Dept. of CSE, Sanketika Vidya Parishad Engineering College, Visakhapatnam - India
[2]Dept. of CSE, Sanketika Vidya Parishad Engineering College, Visakhapatnam - India

Corresponding Author: anusha.jalaparthi@gmail.com

Available online at www.isroset.org

Received: 12 May 2016, Revised: 25 May 2016, Accepted: 18 Jun 2016, Published: 30 Jun 2016

Abstract— Social networks have been, at a recent time utilized as a source of information for event detection, with particular referral to road traffic congestion and for the car accidents. In this paper, we present a real-time monitoring system for traffic event detection from the twitter stream analysis. The system fetches tweets from the twitter according to several search criteria; processes tweets, by applying the text mining techniques; and finally performs the classification of tweets. The aim is to assign the suitable class label to each tweet, as relevant to a traffic event or not. The traffic detection system was exploited for real-time monitoring of several areas of the Italian road network, allowing for detection of traffic events almost in real time, generally before online traffic news web sites. We employed the support vector machine as a classification model, and we attain an accuracy value of 95.75% by solving a binary classification problem (traffic versus non-traffic tweets). We would also able to discriminate if traffic is caused by an external event or not, by solving a multiclass classification issue and obtaining accuracy value of 88.89%.

Keywords- *Twitter; Traffic event detection; tweet classification; text mining, social sensing.*

I. INTRODUCTION

Twitter stream to detect earthquakes and hurricanes, and positive events (earthquakes and typhoons) and adverse events (non-events or other occasions) as a binary SVM applying seed. Agarwal et al. NLP and naive Bayes (NB) using standard techniques workbook, Twitter stream analysis to focus on the detection of a fire in the factory. Lee et al. TEDAS the proposed system, to restore tweets about the event. In this system the fire, thunderstorms, car accidents and crime, as well as events related crashes (CDE- of mind) focuses on, and remember the events of CDE- Keywords spatial and temporal information, and the user's followers and the restoration of a number of hash , links, and the rating of the United States of exploiting the nomination aims to tweets.

Social network analysis is where events such as formatted text, blogs, e-mails and other issues, such as the traditional media are much more difficult to detect the event. Unstructured text and the occasional sound of the non-formal or short, contain spelling or grammatical errors. With the huge amount of data is useful or useless. In this paper, we analyze the Twitter streams of text mining algorithms and machine learning to detect the traffic in real-time, based on an intelligent system have been proposed. System, and the feasibility study has been designed and developed from the ground infrastructure, SOA architecture (SOA), based on the event-driven. Systems for the analysis of text and pattern-site state-of-the-art techniques based on the exploitation of the technology available. These technologies and techniques to analyze, tune and adaptive, and integrated to create an intelligent system. In particular, we classify the various state-of-the-art approach to the text of the present experimental study, which was conducted to determine the most effective. Once the system is integrated into the system and real-time traffic incident has been used to identify fields. In this paper, we have a specific event on a smaller scale, no traffic on the streets, we exaggerated the users belonging to a specific area to detect and analyze the traffic incident and aim to focus on writing in Italian language processing. In order to achieve this goal, we have the system, not on the streets or in the event that relate to the mode of traffic, and the site is able to bring that recommendation.

To our knowledge, for the detection of traffic using twitter stream analysis has suggested that some of the papers. However, with respect to our work, all of them, focusing on the language of the input feature a variety of Italian and / or feature selection algorithm employed, and only considers bilateral classifications. Tweets to 140 characters, and the real-time nature of the news media and platforms. In fact, the life-time favorites are usually very small, and therefore, suitable for the study of Twitter is related to events in real-time on a social network platform. Additional information is up to each of tweets that can be connected directly with descriptive information. Twitter messages in public, that is, they are directly without any confidentiality restrictions.

For these reasons, the Twitter real-time analysis to detect the event is a good source of information. To provide coverage of a wide range of low-cost road network, with the addition of traffic sensors, the system can offer a job (for example, rings, cameras, infrared cameras to detect) and the monitoring of the traffic problem is exhibited, especially in those areas where traditional motion sensors (eg, city and suburbs), Because it recognizes the event of non-commercial, in which the multi-layer, and due to traffic congestion or disaster sites, and traffic will take place. It shows real-time traffic incident. And iii) and SOA framework, which was built on an infrastructure driven event, as it developed.

II. LITERATURE SURVEY

A. *What's Happening: A Survey of Tweets Event Detection*

Twitter is now one of the main modes for spreading of ideas and information throughout the Web. Tweets discuss different style, ideas, events, and so on. This gave rise to an increasing interest in examining tweets by the data mining community. Twitter is, in nature, a good resource for identifying 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 micro blogs. Their rates of activity reached some levels without precedent. Hundreds of millions of users are registered in these micro blogs as Twitter. They exchange and tell their last thoughts, moods or activities by tweets in some words [1].

B. *ET: Events from Tweets*

Social media sites some of which are Twitter and Facebook have emerged as popular tools for people to express their ideas on various topics. The huge amount of data provided by these media is greatly valuable for mining trending topics and events. In this paper, we build an adequate, scalable system to detect events from tweets (ET). Our approach detects events by analyzing their textual and temporal components. ET does not require any target entity or domain knowledge to be stated; it automatically detects events from a set of tweets.

The key components of ET are:

- An extraction scheme for event representative keywords
- An adequate storage mechanism to store their appearance patterns, and
- A hierarchical clustering technique based on the common co-occurring features of keywords.

Authors presented a scalable and adequate system, called ET, to detect real world events from a set of micro blogs/tweets. The key feature of this system is the adequate use of content similarity and appearance similarity among keywords, to cluster the related keywords. We demonstrate the adequateness of this combination in our experiments. ET does not need any human expertise or knowledge from other sources like Wikipedia, but still provides very accurate results. ET is evaluated on two different data sets from two different domains and it produces great results for both of them in terms of the precision [2].

C. *Measurement and Analysis of Online Social Networks:*

The online social networking sites like Orkut, YouTube, and Flickr are out of the most popular sites on the Internet. Users from these sites form a social network, which provides a powerful means of sharing, organizing, and finding content and the contacts. The vogue of these sites provides an opportunity to study the characteristics of online social network graphs at an immense scale. Knowing these graphs is vital, both to improve the current systems and to design the new applications of online social networks. This paper shows a large scale measurement study and scrutiny of the structure of multiple online social networks. We scrutinize data gathered from four vogue online social networks: Flickr, YouTube, Live Journal, and Orkut. We crawled the publicly accessible user links on each of the site, obtaining a huge portion of each social network's graph [3]. Our data set contains over 11.3 million users and had 328 million links. We suppose that this is the first study to examine multiple online social networks at scale. Our results explain the power law, small world, and scale free properties of online social networks. We find that the in degree of user nodes tends to match the out degree; that the networks have a densely connected core of high-degree nodes; and that this core links small groups of strongly clustered, low-degree nodes at the fringes of the network. Lastly, the implications of these structural properties for the design of social network based systems. Presented an analysis of the structural properties of online social networks using data sets collected from four vogue sites. Our data shows that social networks are structurally different from previously studied networks, specifically the Web. Social networks have a much higher fraction of symmetric links and also display much higher levels of local clustering. We have outlined how these properties may affect the algorithms and applications designed for the social networks [4].

D. *Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors*

Twitter, a vogue micro blogging service, has received much attention recently. A significant characteristic of Twitter is its real-time nature. For instance, when an earthquake occurs, people make many Twitter posts (tweets) related to the earthquake, which facilitates detection of earthquake occurrence promptly, simply by observing the tweets. As described in this paper, we scrutinize the real-time interaction of events such as earthquakes, in Twitter, and suggest an algorithm to monitor tweets and to find a target event. To find a target event, we arrange a classifier of tweets

based on features such as the keywords in a tweet, the number of words, and their context. Later, we produce a probabilistic spatiotemporal model for the target event that can find the center and the trajectory of the event location.

We then consider each Twitter user as a sensor and apply Kalman filtering and particle filtering, which are generally used for location estimation in ubiquitous/pervasive computing [5]. The particle filter works better than other compared methods in judging 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 huge number of Twitter users throughout the country, we can investigate 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 finds earthquakes promptly and sends e-mails to registered users. Notification is delivered much faster than the announcements that are broadcast by the JMA [6].

E. Text Detection and Recognition on Traffic Panels From Street-Level Imagery Using Visual Appearance

Traffic sign detection and recognition has been completely studied for a long time. Yet, traffic panel finding and recognition still remains a challenge in computer vision due to its different types and the immense variability of the information illustrated in them. This paper presents a method to detect traffic panels in street level images and to recognize the information contained on them, as an application to intelligent transportation systems (ITS) [7]. The main purpose can be, to make an automatic inventory of the traffic panels located in a road to support road maintenance and to help 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 methodology for traffic panel detection in the state of the art [8]. Lastly, our own text identification and recognition method is applied on those images where a traffic panel has been identified, so automatically read and save the information illustrated in the panels. We suggest a language model partially based on a dynamic dictionary for a finite geographical area using a reverse geo coding service. Experimental results on real images from Google Street View prove the efficiency of the suggested method and give a way to use street level images for different applications on ITS [9].

III. RELATED WORK

A. FETCH OF SUMS AND PRE-PROCESSING

The first module, "Fetch of SUMs and Pre-processing", extracts raw tweets from the Twitter stream, based on one or more search criteria (e.g., geographic coordinates, keywords appearing in the text of the tweet). Each fetched raw tweet contains: the user id, the timestamp, the geographic coordinates, a retweet flag, and the text of the tweet. The text may contain additional information, such as hash tags, links, mentions, and special characters. In this paper, we took only Italian language tweets into account. However, the system can be easily adapted to cope with different languages. After the SUMs have been fetched according to the specific search criteria, SUMs are pre-processed. In order to extract only the text of each raw tweet and remove all meta-information associated with it; a Regular Expression filter [8] is applied. More in detail, the meta-information discarded are: user id, time stamp, geographic coordinates, hash tags, links, mentions, and special characters.

B. ELABORATION OF SUMS:

"Elaboration of SUMs", is devoted to transforming the set of pre-processed SUMs, i.e., a set of strings, in a set of numeric vectors to be elaborated by the "Classification of SUMs" module. To this aim, some text mining techniques are applied in sequence to the pre-processed SUMs. In the following, the text mining steps performed in this module are described in detail: Tokenization is typically the first step of the text mining process, and consists in transforming a stream of characters into a stream of processing units called tokens (e.g., syllables, words, or phrases). During this step, other operations are usually performed, such as removal of punctuation and other non-text characters [10], and normalization of symbols (e.g., accents, apostrophes, hyphens, tabs and spaces).

In the proposed system, the tokenizer removes all punctuation marks and splits each SUM into tokens corresponding to words (bag-of-words representation). A stop-word filtering consists in eliminating stop-words, i.e., words which provide little or no information to the text analysis. Common stop-words are articles, conjunctions, prepositions, pronouns, etc. Other stop-words are those having no statistical significance, that is, those that typically appear very often in sentences of the considered language (language-specific stop-words), or in the set of texts being analyzed (domain-specific stop-words), and can therefore be considered as noise. The authors in [10] have shown that the 10 most frequent words in texts and documents of the English language are about the 20– 30% of the tokens in a given document. In the proposed system, the stop-word list for the Italian language was freely downloaded from the Snowball Tartarus website6 and extended with other ad hoc

defined stop-words. At the end of this step, each SUM is thus reduced to a sequence of relevant token [11].

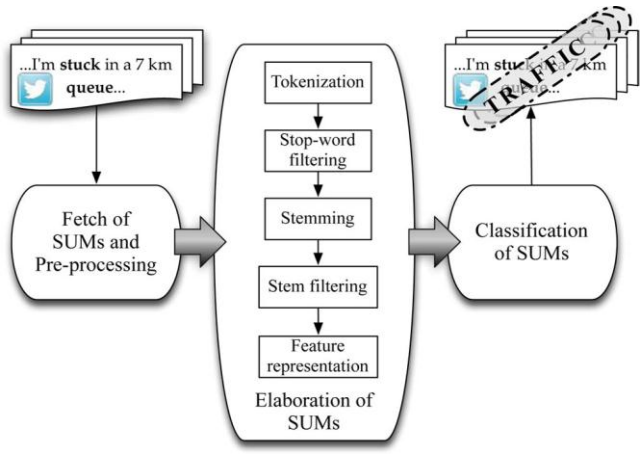

Figure 1. System Architecture

C. CLASSIFICATION OF SUMS:

The third module, "Classification of SUMs", assigns to each elaborated SUM a class label related to traffic events. Thus, the output of this module is a collection of N labeled SUMs. To the aim of labeling each SUM, a classification model is employed. The parameters of the classification model have been identified during the supervised learning stage.

Actually, as it will be discussed in Section V, different classification models have been considered and compared. The classifier that achieved the most accurate results was finally employed for the real time monitoring with the proposed traffic detection system. The system continuously monitors a specific region and notifies the presence of a traffic event on the basis of a set of rules that can be defined by the system administrator.

IV. CONCLUSION

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.

REFERENCES

[1]. F. Atefeh, W. Khreich, "*A survey of techniques for event detection in Twitter*", Computer Intelligence, Vol.31, No.1, pp. 132–164, 2015

[2]. P. Ruchi, K. Kamalakar, "*ET: Events from tweets*," in Proc. 22nd Int. Conf. World Wide Web Computer, Brazil, pp. 613–620, 2013.

[3]. A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, B. Bhattacharjee, "*Measurement and analysis of online social networks*", in Proc. 7th ACM SIGCOMM Conf. Internet Meas., San Diego (USA), pp. 29–42, 2007.

[4]. J. Kothari, T.i Shah, B. Nagaria, A. Choubey, S.D.i Pabba, "*Automated Real Time In-Store Retail Marketing Using Beacon*", International Journal of Computer Sciences and Engineering, Vol.4, Issue.2, pp.110-113, 2016.

[5]. T. Sakaki, M. Okazaki, Y. Matsuo, "*Tweet analysis for real-time event detection and earthquake reporting system development",* IEEE Transaction Knowledage Data Engineering, Vol.25, No. 4, pp. 919–931, 2013.

[6]. M. Krstajic, C. Rohrdantz, M. Hund, A. Weiler, *"Getting there first: Real-time detection of real-world incidents on Twitter*", 2nd IEEE Work Interactive Vis. Text Anal—Task-Driven Anal. Soc. Media IEEE VisWeek, Seattle (USA),pp.128-134, 2012.

[7]. J. Yin, A. Lampert, M. Cameron, B. Robinson, R. Power, "*Using social media to enhance emergency situation awareness",* IEEE Intell. Syst., Vol.27, No. 6, pp. 52–59, 2012.

[8]. T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, K. Nawa, "*Real-time event extraction for driving information from social sensors*", IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), Bangkok, pp. 221-226, 2012.

[9]. H. Ning, H. Liu, "*Cyber-physical-social based security architecture for future internet of things",* Advances in Internet of Things, Vol.2, Issue.1, pp.1-12, 2012.

[10]. A. Schulz, P. Ristoski, H. Paulheim, "*I see a car crash: Real-time detection of small scale incidents in microblogs*", The Semantic Web: ESWC 2013 Satellite Events, Vol.7955, pp. 22–33, 2013.

[11]. P. Agarwal, R. Vaithiyanathan, S. Sharma, G. Shro, *"Catching the long-tail: Extracting local news events from Twitter*", in Proc. Sixth International AAAI Conference on Weblogs and Social Media, Ireland, pp. 379–382, 2012.

