IJRSET MARCH 2017 Volume 4, Issue 3 Pages: 15-19
ISSN 2394-739X
International Journal for Research in Science Engineering and Technology

GEO-TRAFFIC IDENTIFICATION VIA CENTRALIZED WEB STREAM ANALYSIS

[1] Ms. J. Sharmila, [2] Mr. V. Shriram, [3] Mr. R. Vickram, [4] Mr. S. k Raveenraj
[1, 2, 3, 4] Asst Professor
[1, 2, 3, 4] Department of CSE,
[1, 2, 3, 4] Sri Krishna College of Technology,
[1, 2, 3, 4] Coimbatore, India.

ABSTRACT: Social networks can be employed as a source of information for event detection such as road traffic congestion and car accidents. The Existing system present a real-time monitoring system for traffic event detection from twitter. The system fetches tweets from twitter and then; processes tweets using text mining techniques. Lastly performs the classification of tweets. The aim of the system is to assign the appropriate class label to each tweet, whether it is related to a traffic event or not. System employed the support vector machine as a classification model. Road traffic prediction is a critical component in modern smart transportation systems. It provides the basis for traffic management agencies to generate proactive traffic operation strategy for alleviate congestion. Existing work on near-term traffic prediction (forecasting horizons in the range of 5 minutes to 1 hour) relies on the past and current traffic conditions. However, once the forecasting horizon is beyond 1 hour, i.e., in longer-term traffic prediction, these techniques do not work well since additional factors other than the past and current traffic conditions start to play important roles. To address the above problem, examination is done to check whether it is possible to use the rich information in online social media to improve longer-term traffic prediction. Analysis is done to check the correlation between traffic volume and tweet counts with various granularities. Finally, the classification of tweets is done using the k nearest neighbor algorithm. The aim of the system is to assign the appropriate class label to each tweet, whether it is related to a traffic event or not. The traffic detection system can be employed for real-time monitoring of several areas of the road network, allowing for detection of traffic events almost in real time, often before online traffic news web sites.

Keywords: [Social media; Traffic detection; Text mining; machine learning, Twitter stream analysis; K-nearest neighbor algorithm]

1. INTRODUCTION

Social networks are flattering the real information channel over a moment in time. As we can use the portable devices for sharing information over the social network, its use is increased to large extent. Because of the portability and ease of use these networking sites become the channel of valuable and real time information. Now a days the people are using the social media for the reporting the real life events. These social networking sites are used for maintaining the social relationship, finding the users with similar interests. The message shared by user in societal networks is called Status Update Message (SUM). SUM may contain, apart from the text, meta-information such as timestamp, geographic

coordinates, name of the user, links to other resources, hashtags, and mentions. SUM considered in a specific geographic area may provide the accurate information. Social networks and media platforms are being widely used as a source of information for the detection of events, such as traffic, incidents, and natural disasters [1]. An Intelligent transport system is an infrastructure which integrating ICTs (Information and Communication Technologies) with transport networks, allows improving safety and management of transport networks. In telligent Transportation Systems (ITSs). ITSs afford, e.g., real-time information about weather, traffic congestion or regulation, or plan efficient (e.g., shortest, fast driving, least polluting) routes[4], [6], [8]. Twitter has several advantages over the similar micro blogging services. The life-time of tweets is usually small, thus Twitter is the social network platform that is used to study sums related to real-time events. Each tweet is related with the Meta information that constitutes the additional information. Twitter messages are directly available as they are public. In this paper, our focus is on particular small-scale event road traffic. The aim is to detect and analyze traffic related events by processing users' messages belonging to a specific area and written in the English language. A system has been proposed to extract, analyze and classify the SUMs as related to a road traffic event or not in an efficient way. The traffic detection system will be useful in areas such as traffic congestion, incidents, natural disasters (earthquakes, storms, fires, etc.), or other events.

2. RELATED WORK

2.1 Survey of Techniques for Event Detection in Twitter

Micro blogging is a broadcast medium that allows users to exchange small digital content such as short texts, links, images, or videos. Although it is a relatively new communication medium compared with traditional media, micro blogging has gained increased attention among users, organizations, and research scholars in different disciplines. The popularity of micro blogging stems from its distinctive communication services such as portability, immediacy, and ease of use, which allow users to instantly respond and extend information with restricted or no limitations on content. in effect any person witnesses or involved in any event is nowadays able to disseminate real-time information, which can reach the other side of the world as the event unfolds. For occurrence, during recent social cataclysm and crises, millions of people on the ground turned to T witter to report and follow significant events.

2. 2. ET: Events from Tweets

An important characteristic of Twitter is its real time nature. Its availability on web as well as mobile devices (cell phones) enables users to post tweets at any time and place. This ease of publishing messages on Twitter makes it a popular source of data to detect real-world events. There have been many events which were highlighted by Twitter users almost at the same time or before they were identified by the traditional news media. For example, Twitter revealed the 2010 cholera outbreak in Haiti two weeks before the health oficials2. There are many applications that use Twitter to track earthquakes. The disastrous earthquake that struck Japan in 2011 was identified within seconds because of the spikes in tweets from affected area, compared to the typical 2 to 20 minutes taken by scientific alerts3. Another prominent category of events detected by Twitter is sports.

2.3 Measurement and Analysis of Online Social Networks

The Internet has spawned different types of information sharing systems, including the Web. Recently, online social networks have gained significant popularity and are now among the most popular sites on the Web. Unlike the Web, which is largely organized around content, online social networks are organized around users. Participating users join a network, publish their profile and (optionally) any content, and create links to any other users with

whom they associate. The resulting social network provides a basis for maintaining social relationships, for finding users with similar interests, and for locating content and knowledge that has been contributed or endorsed by other users.

3. EXISTING SYSTEM

Recently, social networks and media platforms have been widely used as a source of information for the detection of events, such as traffic congestion, incidents, natural disasters (earthquakes, storms, fires, etc.), or other events. Twitter streams to detect earthquakes and typhoons, by monitoring special trigger keywords, and by applying an SVM as a binary classifier of positive events (earthquakes and typhoons) and negative events (non-events or other events). Focus on the detection of fires in a factory from Twitter stream analysis, by using standard NLP techniques and a Naive Bayes (NB) classifier Event detection from social networks analysis is a more challenging problem than event detection from traditional media like blogs, emails, etc., where texts are well formatted The main difficulty encountered in dealing with problems of text mining is caused by the vagueness of natural language. Feature selection is particularly important, since one of the main problems in text mining is the high dimensionality of the feature space. Text mining is a difference on a field called data mining [2] SUMs are unstructured and irregular texts, they contain informal or abbreviated words, misspellings or grammatical errors. SUMs contain a huge amount of not useful or meaningless information.

4. SYSTEM ARCHITECTURE

An intelligent system based on text mining and supervised learning is used for real-time detection of traffic events from Twitter stream analysis. 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. The chosen approach was integrated into the final system and used for the on-the-field real-time detection of traffic events. As regards binary classification, we consider traffic-related tweets, and tweets not related with traffic. System could work together with other traffic sensors ITS monitoring systems for the detection of traffic difficulties, providing a low-cost wide coverage of the road. It performs a multi-class classification, which recognizes non-traffic, traffic due to congestion or crash, and traffic due to external events. The architecture of the proposed system is clearly depicted in the Figure 1.

4.1 Dataset Collection and Fetching of tweets

The tweets can be crawled from the complete large Twitter datasets using crawlers using Twitter API. The tweets were fetched in a time span of about four hours from the same geographic area.

4.2 User Content Posting

This module is used to design the system based on twitter to handle on tweets. It has a login page for twitter users and for administrator to handle the event reporting system and extract the tweets from the twitter. Users can post their traffic related information based on location and time in online websites, which can also be used as an input for the project.

4.3 Data Preprocessing

Data preprocessing can be used to extract only the text of each raw tweet and remove all meta-information associated with it. The tokenizer removes all punctuation marks and splits each SUM into tokens corresponding to words (bag-of-words representation). Stop-word filtering consists in eliminating stop-words, i.e., words which provide little or no information to the text analysis. The stop words used are insignificant words like the, a, etc. Stemming is also performed to convert the particular raw text to the stem word. It helps in decreasing the number of each of the SUM. E.g.,' Driving' and 'Drived' can be converted to the root word 'drive'.

4.4 Data Classification

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. K nearest neighbor algorithm used to classify the traffic related information and normal posts. Identify the traffic related post and retrieve the location and within 1hr posts. Actual notification of the traffic event may be sent after the identification of a certain number of tweets with the same label.

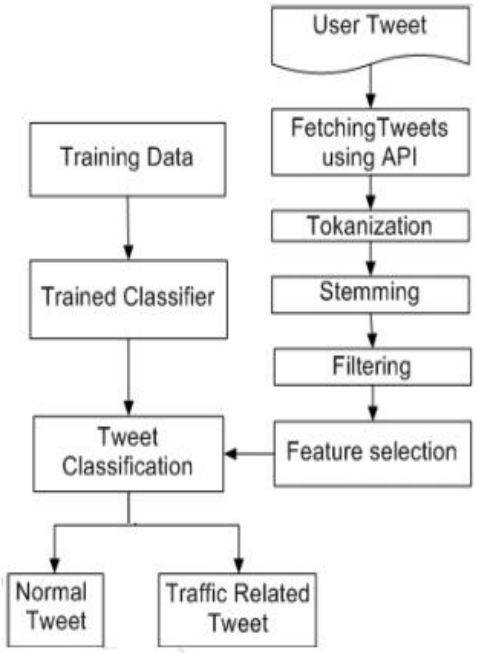

Figure - 1 Traffic Analysis System

5. KNN CLASSIFICATION

KNN algorithm is one of the simplest algorithms which give high competitive results. The algorithm uses the Euclidean distance metric to predict the class label for the tweet as “traffic related” or “non-traffic”. The classification may help people to get an alert and they may be directed to take a different route to reach their destination.

6. EXPERIMENTAL RESULTS

The dataset used for the project comprises of the attributes date, day, timestamp, user id and user tweets.

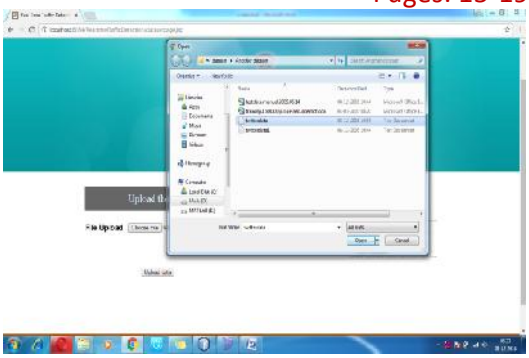

Figure - 2 Dataset Collection through Twitter API

The above Figure 2 illustrates the fetching of tweets from twitter API.

Figure - 3 Notification for successful upload of the dataset

The above figure 3 illustrates the successful upload of the dataset.

Figure - 4 User Content Posting

The Figure 4 illustrates that the users can post the traffic related information in online website.

Table 1

Text of tweet	Class
To all friends of Catania, crash on the bypass! Queue from Talegaon to Akurdi	Traffic
Crash on Mumbai highway! Traffic slowed	Traffic
I happy to see that this great player is well after crash he had	Non Traffic

Figure - 5 Data Classification

The Figure 5 classifies whether the tweets posted by the user are traffic related tweets or non-traffic related tweets.

CONCLUSION AND FUTURE WORK

In the current paper, a system for real-time detection of traffic-related events from Twitter stream analysis and users can also post traffic related information has been proposed. System is able to fetch and classify streams of tweets and to notify the users of the presence of traffic events. The most important shortcoming of Twitter data for use in geographical analysis in general is that only a very limited number of tweets can be geo-located by a coordinate pair that is available in the metadata of the tweets. As a consequence it is difficult to detect small-scale events in the large amounts of data that are generated through Twitter every day. Real-world events that are successfully identified in this thesis using spatio-temporal analysis are detected from massive and temporary crowds at festivals, fairs, concerts or sport events. Without sophisticated noise filtering, machine learning or language processing, it seems challenging to detect small-scale events with analysis that is solely based on the spatio-temporal characteristics of tweets. The results of a monitoring campaign, performed. 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.

For future work, along with the traffic related information notification of the shortest route to the preferred destination can be given by incorporating Google Maps.

REFERENCES

[1]. AminaMadani, Omar Boussaid, Djamel EddineZegour and Algiers, Algeria, "What's Happening: A Survey of Tweets Event Detection," J. Internet Services Appl., vol. 1, no. 1, pp. 7–18, 2010.

Parikh, Kamalakar Karlapalem., "ET: Events from Tweets," Future Generat. Comput. Syst., vol. 25, no. 6, pp. 599–616, Jun. 2009.

[2]. Alan Mislove, MassimilianoMarcon, Krishna P. Gummadi, Peter Druschel, Bobby Bhattacharjee., "Measurement and Analysis of Online Social Networks," Proc. IEEE, vol. 99, no. 1, pp. 149–167, Jan. 2011.

[3]. 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.

[4]. T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving Programming, Bangalore, India, Jan 2010, pp. 147-158.

[5]. M. Yuriyama and T. Kushida, "Sensor-cloud infrastructure Physical sensor management with virtualized sensors on cloud computing," in Proc. 13th Int. Conf. Netw.,-Based Inf. Syst., Sep. 2010, pp. 1–8

[6]. Chen and Harry H., "A Review of the Applications of Agent Technology in Traffic and Transportation Systems," Wireless Commun. Mobile Comput, vol. 14, no. 1, pp. 19–36, Jan. 2014.

[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. Intel. Transp. Syst., vol. 11, no. 2, pp. 485–497, Jun. 2010.