Real-Time Monitoring System For Traffic Event Detection From Twitter Stream Analysis

Pritesh Patil[1], Kavita Pandey[2], Suraj Pawar[3]
[1, 2, 3, 4] Department of Information Technology
[1, 2, 3] Savitribai Phule Pune University, Pune, Maharashtra

Abstract-*People get to know about some road congestion from news and radio. But this scenario sometimes takes more time than required. Due to this many emergency cases are also made to wait in traffic for a longer period of time. Since social networking sites, such as twitter have spread in recent years becoming a new kind of real time information channels .The twitter data i.e. tweets have a benefit over other social networking sites' data that it is generally associated with some meta data and the 140 character constraint , thus allowing news oriented data. So, we present a real time monitoring system of these tweets. The system fetches tweet according to several search criteria, process them by using text mining techniques and finally classifies them. The aim is to assign suitable class label to every tweet as related with an activity of traffic events or not. We use support vector machine as a classified model. Finally we can notify the user of presence of traffic analysis and provide an alternate way.*

Keywords-character recognition, character segmentation, Number plate detection, Toll collection, Vehicle number recognition.

I. INTRODUCTION

The social networking sites like the Twitter, Facebook, Google+ have spread recently thus becoming a new type of real-time information channel. The smartphones and tablets, led to the easiness of use, and real-time nature these sites. To express one self and to tell about some real time event people intensely make use of these social networks.

The message in social networking sites are called as Status Update Messages (SUM). The SUM may contain, apart from the text, some meta-information like the name of the user, timestamp, hashtags, and mentions. These SUMs if analysed properly can be a good source of valuable information .About an event or a topic. In fact, we may regard social network the SUMs are unstructured and irregular texts, they mostly consists of informal or abbreviated words, and misspellings . Mostly, SUMs contain a huge amount of meaningless information, that must filtered. It has been analysed that over 40% of all tweets are pointless. So to get proper information from the tweets we need to apply the text mining techniques on them. Text mining techniques are based on the idea that a document can be represented by the group of words contained in it [4]. At the time of text mining process, many operations can be performed, depending on the goal, like the text filtering by means of specific keywords, and feature selection, i.e., reduction of the number of features in order to take only the relevant ones.

Among social networks platforms, we took into account Twitter as nowadays it is the most popular micro-blogging service; it has more than 600 million active users. Twitter has more advantages over the similar micro-blogging services .First of all like the tweets are having a limit up to 140 characters, enhancing the real-time and news-oriented nature of the platform and the life-time of tweets is usually very short thus related to real-time events. Second, tweet can have meta-information that works as additional information. Third, Twitter messages are easily available .Due to these reasons; Twitter is a good source of information for detection and analysis of the real-time event.

In this paper, we put up an intelligent system, based on text mining and machine learning algorithms, for real-time detection of traffic events from Twitter stream analysis. The system is built on service oriented architecture.

II. RELATED WORK

ET: Events from Tweets

Social media sites such as Twitter and Facebook have emerged as popular tools for people to express their opinions on various topics. The large amount of data provided by these media is extremely valuable for mining trending topics and events. In this paper, there is an efficient, scalable system to detect events from tweets (ET) [1]. It detects events by exploring their textual and temporal components. ET does not require any target entity or domain knowledge to be specified; it automatically detects events from a set of tweets.

Measurement and Analysis of Online Social Networks:

This paper presents a large-scale measurement study and analysis of the structure of multiple online social

networks. It examines data gathered from four popular online social networks: Flickr, YouTube, LiveJournal, and Orkut. It traversed the publicly accessible user links on each site, obtaining a large portion of each social network's graph [2]. The results confirm the power-law, small-world, and scale free properties of online social networks. The data shows that social networks are structurally different from previously studied networks. Social networks have a much higher fraction of symmetric links and also exhibit much higher levels of local clustering.

Real-time Event Detection by Social Sensors:

When an earthquake occurs, people make many tweet related to the earthquake, which enables detection of earthquake occurrence, simply by observing the tweets. As described in this paper, 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 [3]. We consider each Twitter user as a sensor and filtering. System detects earthquakes promptly and sends e-mails to registered users. Notification is delivered much faster than the announcements that are broadcast other news channels.

III. SYSTEM DESIGN

1. 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 re-tweet flag, and the text of the tweet. The text may contain additional information, such as hashtags, links, mentions, and special characters. 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 is applied.

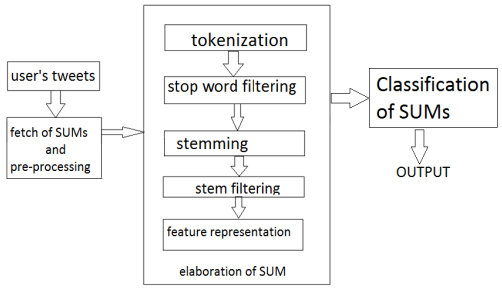

Fig1. System design

2. Elaboration Of Sums:

As we can see in the system design (Fig 1), the second processing module, Elaboration of SUMs, is devoted to transforming the set of pre-processed SUMs [5], i.e., a set of strings, in a set of numeric vectors to be elaborated by the Classification of SUMs module. For this text mining techniques are applied in sequence to the pre-processed SUMs. the text mining steps performed in this module are :

a) 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. The tokenizer removes all punctuation marks and splits each SUM into tokens corresponding to words (bag-of-words representation). At the end of this step, each SUM is represented as the sequence of words contained in it.
b) 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.
c) Stemming is the process of reducing each word (i.e., token) to its stem or root form, by removing its suffix. The purpose of this step is to group words with the same theme having closely related semantics.
d) Stem filtering consists in reducing the number of stems of each SUM. In particular, each SUM is filtered by removing from the set of stems the ones not belonging to the set of relevant stems.
e) Feature representation consists in building, for each SUM, the corresponding vector of numeric features. Indeed, in order to classify the SUMs, we have to represent them in the same feature space.

3. Classification of SUMs:

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

Now a days the social networking sites like Twitter is a medium used by people to express their views and opinions about some event around them .The emergency cases like ambulance and firebrigade need to know about the traffic on the way they are about to go. In case of rallies, people who want to reach faster can lookout for the alternate way. People's tweet can be used as a medium of social sensor. These tweet can help in locating traffic in a given region .we use data mining techniques and the classification technique to identify the tweets. These tweets are related to the traffic and the location as requested by the user.

REFERENCES

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

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

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

[4] M.W. Berry and M. Castellanos, "Survey of Text Mining" NewYork, NY, USA: Springer-Verlag, 2004

[5] Eleonora D'Andrea, Pietro Ducange, Beatrice Lazzerini, "Real Time Detection Of Traffic From Twitter Stream Analysis ", IEEE.vol 11 , no. 5 , Jan 2016