

To identify or detect traffic by analysing various tweets on Social Networking Media like Twitter

^{#1}Abhishek Pandharinath Khutwad, ^{#2}Mayur Hiranman Gorgal

¹khutwadabhishek@gmail.com

²Mayur.gorgal.11@gmail.com

^{#12}Department of Computer Engineering

Smt. Kashibai Navale College of Engineering



ABSTRACT

Twitter has become a popular micro-blogging service. An important characteristic of Twitter is its real-time nature. For example, when an earthquake occurs people make many tweets related to the earthquake, which enables detection of earthquake occurrence promptly, simply by observing the tweets. In this report 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. We also establish a temporal model to estimate time of occurrence. Finally, we use geo-charts from Google to show our results on maps.

Keywords: Tweet classification, Traffic event detection, Data mining, text mining, and social sensing.

ARTICLE INFO

Article History

Received: 13th April 2017

Received in revised form :
13th April 2017

Accepted: 18th April 2017

Published online :

18th April 2017

I. INTRODUCTION

In Social networks are becoming the real information channel over a 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 the maintaining the social relationship ,finding the users with similar interests. The message shared by user in social networks is called Status Update Message (SUM). SUM may contain, apart from the text, meta-information such as time stamp, geographic coordinates, name of the user, links to other resources, hash-tags, and mentions.

In they came to a point It is important to understand the societal concerns over the increased rate of resource consumption and waste production and therefore the policy makers have encouraged recycling and reuse strategies to reduce the demand for raw

materials and to decrease the quantity of waste going to land.

In 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. 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 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 trace. Our aim to detect and analyze trace related events by processing users messages belonging to a specific area and written in the English language. We propose a system able to extract, analyze SUMs as related to a road traffic event or not. Few papers have been proposed for traffic detection using Twitter stream analysis. 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.

II. EXISTING SYSTEM

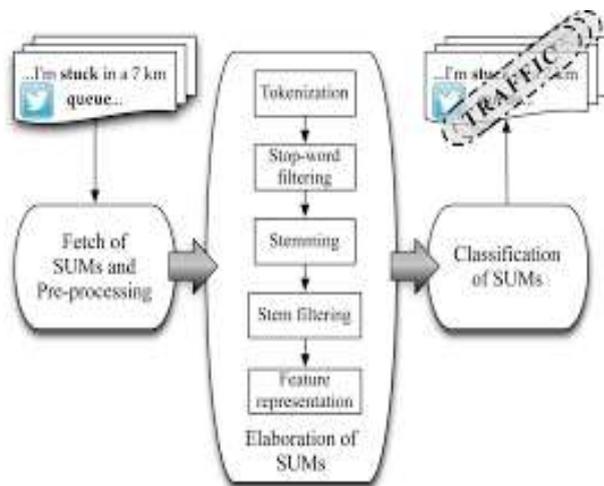


Figure 1: System architecture for traffic detection from Twitter stream analysis

Existing system propose an intelligent system, based on text mining and machine learning algorithms, for real-time 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) [1].

The system exploits available technologies based on state-of-the-art techniques for text analysis and pattern classification [4]. These technologies and techniques have been analyzed, adapted, and added with existing in order to build the intelligent system [1]. In particular, system 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 is added into the final system and then used for the on-the-field real-time detection of traffic events. [1]

III. PROPOSED SYSTEM

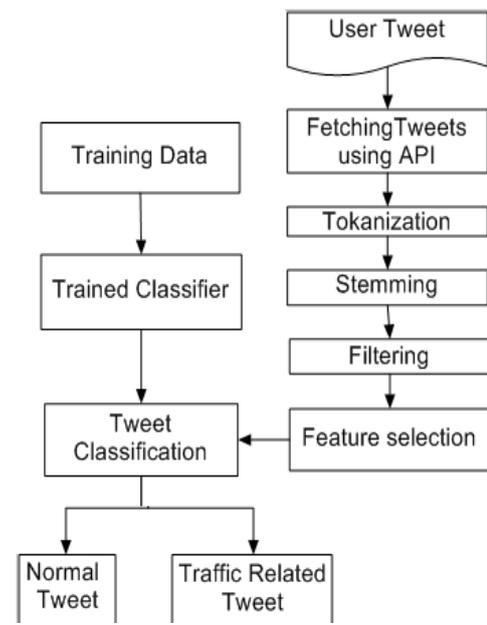


Figure 2: Architecture

In this section, our traffic detection system based on Twitter streams analysis is presented. The system architecture is service-oriented and event-driven, and is composed of three main modules, namely:

- i) Extraction of SUMs and Pre-processing
- ii) Elaboration of SUMs
- iii) Classification of SUMs.

The purpose of the proposed system is to fetch SUMs from Twitter and process SUMs by using text mining steps, and to assign the appropriate class label to each SUM. Finally, as shown in Fig. 1, by analyzing the classified SUMs, the system is able to notify the presence of a traffic event.

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. Each fetched raw tweet contains: the user id, the timestamp, the geographic coordinate, retweet flag along with the text of the tweet.

Elaboration of SUMs

The second processing module is Elaboration of SUMs. This 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:

- a) Tokenization is typically the first step of the text mining process. This process is used for transforming a stream of characters into a stream of processing units

called tokens [1]. During this step, other operations are usually performed, such as removal of punctuation and other non-text characters [8], and normalization of symbols

b) Stop-word filtering eliminates stop-words, the words which provide little or no information to the text analysis.

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.

Classification of SUMs

Last module is, Classification of SUMs. This module assigns each elaborated SUM a class label related to traffic events. Thus, the output of this module is a collection of N labeled SUMs. Proposed Clustering Algorithm: Input: Training Dataset T, Test dataset D, Output: Clustered Tweet set.

Method:

1. Initially train the classifier using semi-supervised traffic related training dataset.
2. Fetch user tweets from tweeter account
3. Store in DB
4. For each tweet in DB
5. Calculate the similarity using Euclidean distance with trained data.
6. If(similarity > Threshold)
7. Add tweet to traffic related tweet set
8. Else
9. Add to normal tweet set.
10. End if
11. End for
12. Return classified tweets

IV. CONCLUSION

In this work we have proposed a system of real time detection of traffic related event from twitter stream analysis. System is able to fetch and classify stream of tweets and to notify the user of the presence of traffic event.

REFERENCES

- 1 F. Atefeh and W. Khreich, A survey of techniques for event detection in Twitter, *Comput. Intell.*, vol. 31, no. 1, pp. 132164, 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. 613620.

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. 2942.

4. The Smarty project. [Online]. Available: <http://www.smarty.toscana.it>

5. T. Sakaki, M. Okazaki, and Y. Matsuo, Tweet analysis for real-time event detection and Traffic reporting system development, *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 4, pp. 919931, Apr. 2013.