

Traffic Detection Using Tweets on Twitter Social Network

Supriya Bhosale¹, Sucheta Kokate²

^{1,2}Department of Computer Engineering, Savitribai Phule University, Pune, India

Abstract: Social networks can be employed as 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 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. The proposed system uses the system based on semi-supervised approach, which gives training using traffic related dataset. We propose a clustering approach for classification of the tweets in traffic related and non- traffic related tweets. We employ a Euclidean distance to calculate the similarity between the tweets.

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

1. Introduction

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 timestamp, geographic coordinates, name of the user, links to other resources, hashtags, and mentions [1]. 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 [4].

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. Our aim to detect and analyze traffic 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.

2. Motivation

- 1) Tweets are created in real-time.
- 2) Twitter has 140-character-message limit and the popularity of mobile applications of twitter, users tweet and ability to retweet instantly [1].
- 3) Tweets have a broad coverage over events. Every user can report news that is happening around him or her.
- 4) Tweets are not isolated; instead it contains the rich information.
- 5) Tweets cover nearly all aspect of daily life such as breaking news, local events and personal feelings.

3. Related work

Now days the social media is used for the extraction of information related to the real time event detection. Small events have a small number of SUMs related to them [1], belong to a precise geographic location, and are concentrated in a small time interval [9]. Large events such as earthquakes, tornados are detected by a large number of SUMs, and by a wider geographic coverage [9].

Sakaki et al. [6] used twitter streams to detect earthquakes etc. by monitoring special trigger-keywords, and by applying an SVM as a binary classifier of positive events and negative events. In [7], authors presented a method for detecting real-world events, such as natural disasters, tornados by analyzing Tweets by employing both term-frequency-based techniques and NLP [7].

In paper [2] authors proposed a novel system which detect and analyze events from rich information from Twitter. The proposed method supports the following three functionalities (1) detecting new events, (2) ranking events according to their importance, and (3) generating temporal and spatial patterns for events [2]. Initially authors focused on Crime and Disaster related Events (CDE), such as shooting, car accidents etc [2].

Authors in [3] presented a system to dynamically analyze traffic and its causes. Twitter is a social networking site which allows people to share and read tweets related to

public event and personal emotions [3]. Presented system reads the tweets from twitter and applies natural language processing technique on them. The system categorizes the tweets related to traffic; notifies the registered users about it. The NLP focuses on developing efficient algorithms to process text to convert it into machine understandable language [3].

Vikram Singh et.al. proposed an effective tokenization technique which is based on training. In the results, it is shown that tokenization along with pre-processing generates better tokens [4], If less number of token generated then less storage space is required and it facilitates more accuracy in results retrieval [4]. Algorithm also responsible for reducing the time of information retrieval model [4].

Maximilian Walther et. al. in [5] detect Geo-spatial Event using twitter. The paper describes a new scenario and approach to tackle it. In this approach they gathered tweets for target events that can be defined by a user via keywords [5]. Classification and particle filtering methods are used for detecting events. Authors used common theme as if people tweeting from the same place use the same words which means that they talking about that thing only [5].

4. Existing System

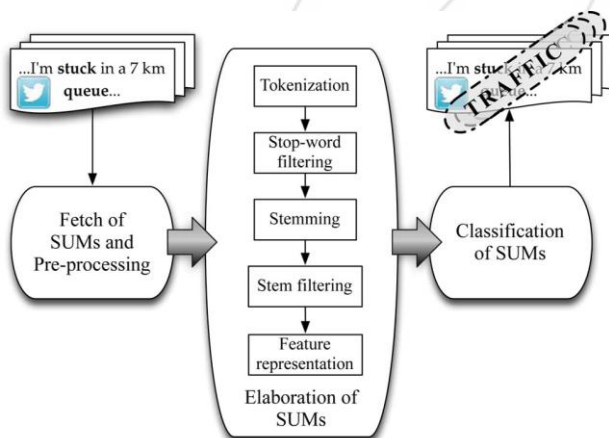


Figure 1: System architecture for traffic detection from Twitter stream analysis.

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]

5. Proposed System

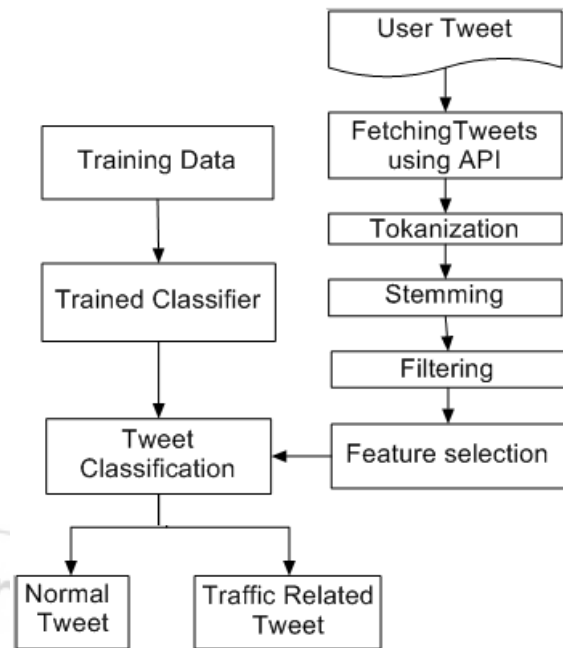


Figure 2: Architecture

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.

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

5.2 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.

5.3 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

6. Conclusion

In this work, we have proposed a system for real-time detection of traffic-related events from Twitter stream analysis. System is able to fetch and classify streams of tweets and to notify the users of the presence of traffic events.

References

- [1] Eleonora D'Andrea, Pietro Ducange, Beatrice Lazzerini, Member, IEEE, and Francesco Marcelloni, Member, IEEE, "Real-Time Detection of Traffic
- [2] Twitter Stream Analysis", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 16, NO. 4, AUGUST 2015
- [3] Rui LI, Kin Hou Lei, Ravi Khadiwala, Kevin Chen-Chuan Chang "FEDAS: a Twitter Based Event Detection and Analysis System", IJCSIT 2014.
- [4] Harshita Rajwani, Srushti Somvanshi, Anuja Upadhye, "Dynamic Traffic Analyzer Using Twitter —, International Journal of Science and Research (IJSR) 2014.
- [5] Vikram Singh and Balwinder Saini —"An Effective Tokenization Algorithm for Information Retrieval System" CS & IT-CSCP 2014
- [6] Maximilian Walther and Michael Kaiser, "Geo-spatial Event Detection in the Twitter Stream", P. Serdyukov et al. (Eds.): ECIR 2013, LNCS 7814, pp. 356367, 2013. cS pringer VerlagBerlinHeidelberg 2013.
- [7] 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.
- [8] M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, "Getting there first: Real-time detection of real-world incidents on Twitter" in Proc. 2nd IEEE Work Interactive Vis. Text Anal.—Task-Driven Anal. Soc. Media IEEE VisWeek," Seattle, WA, USA, 2012.
- [9] A. Schulz, P. Ristoski, and H. Paulheim, "See a car crash: Real-time detection of small scale incidents in microblogs," in The Semantic Web: ESWC 2013 Satellite Events, vol. 7955. Berlin, Germany: Springer-Verlag, 2013, pp. 22–33.
- [10] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," IEEE Intell. Syst., vol. 27, no. 6, pp. 52–59, Nov./Dec. 2012.

Author Profile



Supriya Bhosale, received the B.E. Degree in Computer Science and Engineering from Shivaji University Kolhapure, India, in 2011. She is currently pursuing M.E Degree in Computer Engineering from Savitribai Phule University, Pune, India.



Sucheta Kokate, received the B.E. Degree in Information Technology, and M.Tech Degree in Computer Science and Engineering from Shivaji University, Kolhapure, India. She is currently working as Assistant Professor at TSSMS BSCOER Savitribai Phule University, Pune, India.