ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Tweet Analysis for Real-Time Traffic Detection and Alert Reporting System

Ramiz Shaikh, Sameer Shinde, Vaibhav Pawar, Prof. Prashant Kale

B. E. Student, Dept. of Computer Engineering, G.N Sapkal College of Engg, Anjaneri, Nashik, India

B. E. Student, Dept. of Computer Engineering, G.N Sapkal College of Engg, Anjaneri, Nashik, India

B. E. Student, Dept. of Computer Engineering, G.N Sapkal College of Engg, Anjaneri, Nashik, India

Asst. Professor of Dept. of Computer Engineering, L.G.N Sapkal College of Engg, Anjaneri, Nashik, India

ABSTRACT: Social networking sites are source of information for event detection, with specific reference of the road traffic activity blockage and accidents or earth-quack sensing system. During this paper, we have a tendency to present a time period observation system supposed for traffic occasion detection coming back from social media stream analysis. The system fetches tweets coming back from social media/network as per a many search criteria; ways tweets/posts, by applying matter content mining methods; last however not least works the classification of social networks posts. The goal is to assign appropriate category packaging to each posts, as a result of connected with Associate in Nursing activity of traffic event or maybe not. The traffic recognition system or framework was utilized for time period observation of varied areas of the road network, taking into consideration detection of traffic occasions simply virtually in actual time, frequently before on-line traffic news sites. All people utilized the support vector machine sort of a classification unit; what is more, we tend to accomplish a good accuracy price of 95.76% by making an attempt a binary classification problem. All people were conjointly capable to discriminate if traffic is triggered by Associate in nursing external celebration or not, by partitioning a multi category classification issue and getting accuracy price of 88.89.

KEYWORDS*:* Social media, Traffic detection, Text mining; Privacy, Service oriented architecture (SOA), machine learning, Twitter/social media's stream analysis.

I. INTRODUCTION

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 [1] [2]. These detection schemes are ineffective against feature fabrications or consume much time and resources. Conventional suspicious URL detection schemes utilize several features including lexical features of URLs, URL redirection, HTML content, and dynamic behaviour. However, evading techniques such as time-based evasion and crawler evasion exist [3]. In this paper, we 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) [4]. The system exploits available technologies based on state-of-the-art techniques for text analysis and pattern classification. These technologies and techniques have been analysed, tuned, adapted, and integrated in order to build the intelligent system. In particular, we 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 was integrated into the final system and used for the on-the-field real-time detection of traffic events. In the existing system attackers use shortened malicious URLs that redirect Twitter users to external attack servers. To cope with malicious tweets, several Twitter spam detection schemes have been proposed. These schemes can be classified into account feature-based, relation feature-based, and message feature based schemes. Account feature-based schemes use the distinguishing features of spam accounts such as the ratio of

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

tweets containing URLs, the account creation date, and the number of followers and friends. However, malicious users can easily fabricate these account features. The relation feature-based schemes rely on more robust features that malicious users cannot easily fabricate such as the distance and connectivity apparent in the Twitter graph. Extracting these relation features from a Twitter graph, however, requires a significant amount of time and resources as a Twitter graph is tremendous in size. The message feature-based scheme focused on the lexical features of messages. However, spammers can easily change the shape of their messages. A number of suspicious URL detection schemes have also been introduced.

A. Purpose:

The main purpose of the system is get Public traffic tweets from twitter. For real-time detection of traffic-related events from Twitter stream analysis. 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. Web portal gets array of latitude and longitude and sends return to traffic between that arrays with causes. Alternate path displayed with traffic.

B. Scope: This system is generally based on get Public traffic tweets from twitter and Apply tokenization, remove stop words and apply stemming to a particular tweet. Our traffic detection system based on Twitter streams analysis is presented. And it detects the traffic events in real-time. Haversine method is used to calculate the distance between two latitude-longitude pairs, Triangulation for getting GPS Location. After comparing the longitude and latitude having traffic, it is displayed on the maps of Android device. The system is use to Twitter as data source for fetching the all post regarding the road traffic and Accidents.

C. Objectives: Our traffic detection system based on Twitter streams analysis is presented. And it detects the traffic events in real-time. Also Web part gets array of latitude and longitude of searched path and then the latitude and longitude of the traffic is compared with searched path with their causes. After comparing the longitude and latitude having traffic, it is displayed on the maps of Android device. Haversine method used to calculate distance between two latitude-longitude pairs, Triangulation for getting GPS Location.

II. LITERATURE SURVEY

Twitter is, in nature, a good resource for detecting events in real-time. In this survey paper, we have presented four challenges of tweets event detection: health epidemics identification, natural events detection, trending topics detection, and sentiment analysis. Get Public traffic tweets from twitter and Apply tokenization, remove stop words and apply stemming to a particular tweet, We have maintained lists of causes (e.g. Accidents, Traffic, Jams, Vehicle breakdowns, etc.) and we check these causes in that particular tweet. Web part gets array of latitude and longitude of searched path and then the latitude and longitude of the traffic is compared with searched path with their causes.

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. Sakaki et al. use 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). Agarwal et al. focus on the detection of fires in a factory from Twitter stream analysis, by using standard NLP techniques and a Naive Bayes (NB) classifier. Li et al. propose a system, called TEDAS, to retrieve incident-related tweets. The system focuses on Crime and Disaster-related Events (CDE) such as shootings, thunderstorms, and car accidents, and aims to classify tweets as CDE events by exploiting a filtering based on keywords, spatial and temporal information, number of followers of the user, number of rewets, hash tags, links, and mentions.

III. PROPOSED SYSTEM

A. Proposed System:

In this paper, 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). The system exploits available technologies based on state-of-the-art techniques for text analysis and pattern classification. These technologies and techniques have been analysed, tuned, adapted, and integrated in order to build the intelligent system as shown in Fig.1. In particular, we present an experimental study, which has been performed for determining the most

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

effective among different state-of-theart approaches for text classification. The chosen approach was integrated into the final system and used for the on-the field real-time detection of traffic events. In this paper, we focus on a particular small-scale event, i.e., road traffic, and we aim to detect and analyze traffic events by processing users' SUMs belonging to a certain area and written in the Italian language. To this aim, a system able to fetch, elaborate, and classify SUM's as related to a road traffic event or not. To the best of our knowledge, few papers have been proposed for traffic detection using Twitter stream analysis. However, with respect to our work, all of them focus on languages different from Italian, employ different input features and/or feature selection algorithms, and consider only binary classifications.

B. Proposed System Architecture:

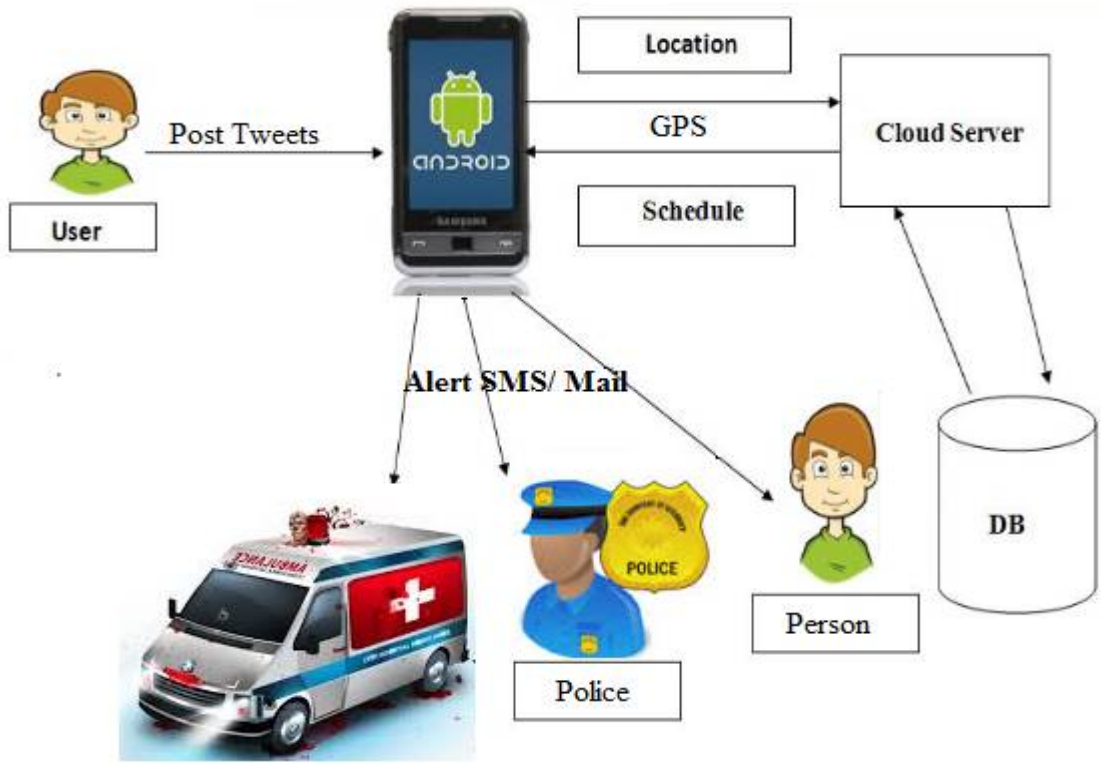


Fig.1: Proposed System Architecture

C. Advantages of Proposed System:

- Tweets are up to 140 characters, enhancing the real-time and news-oriented nature of the platform. In fact, the life-time of tweets is usually very short, thus Twitter is the social network platform that is best suited to study SUMs related to real-time events.
- Each tweet can be directly associated with meta-information that constitutes additional information. Twitter messages are public, i.e., they are directly available with no privacy limitations. For all of these reasons, Twitter is a good source of information for real-time event detection and analysis.
- Moreover, the system architecture could work together with other traffic sensors (e.g., loop detectors, cameras, infrared cameras) and ITS monitoring systems for the detection of traffic difficulties, providing a low-cost wide coverage of the road network, especially in those areas (e.g., urban and suburban) where traditional traffic sensors are missing.
- It performs a multi-class classification, which recognizes non-traffic, traffic due to congestion or crash, and traffic due to external events.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

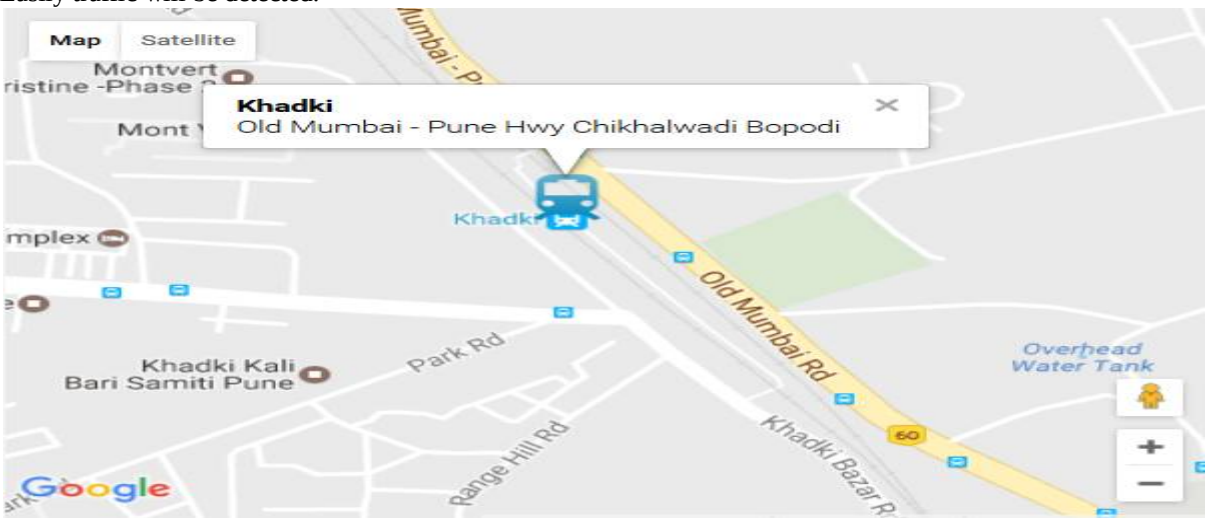
Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

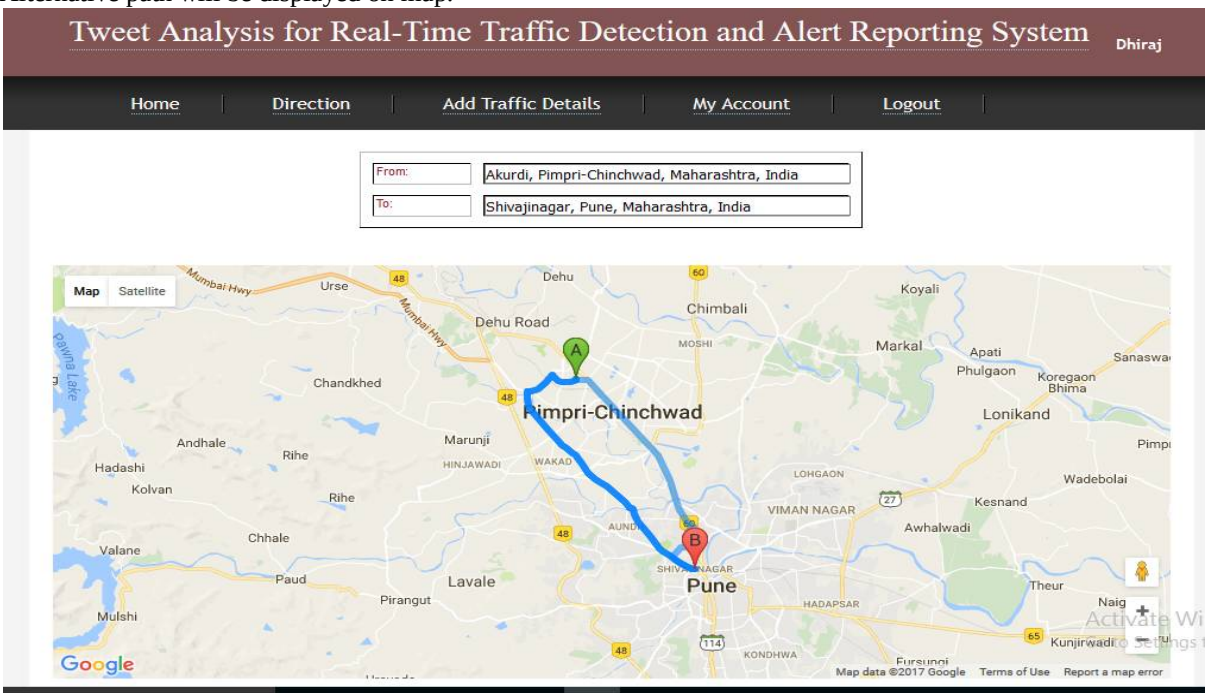
- It detects the traffic events in real-time; and iii) it is developed as an event-driven infrastructure, built on SOA architecture.

IV. RESULTS

- Easily traffic will be detected.



- Alternative path will be displayed on map.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

- Traffic alert will be sending on new user's mail/SMS.

V. CONCLUSION

In this paper, 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.

VI. FUTURE SCOPE

As future work, we are planning to integrate our system with an application for analysing the official traffic news web sites, so as to capture traffic condition notifications in real-time. Thus, our system will be able to signal traffic-related events in the worst case at the same time of the notifications on the web sites. Further, we are investigating the integration of our system into a more complex traffic detection infrastructure. This infrastructure may include both advanced physical sensors and social sensors such as streams of tweets. In particular, social sensors may provide a low-cost wide coverage of the road network, especially in those areas (e.g., urban and suburban) where traditional traffic sensors are missing.

REFERENCES

1. Eleonora DAndrea, Pietro Ducange, Beatrice Lazzerini, Member, IEEE, and Francesco Marcelloni, Member, IEEE, Real-Time Detection of Traffic From Twitter Stream Analysis, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 16, NO. 4, AUGUST 2015.
2. F. Atefeh and W. Khreich, "A survey of techniques for event detection in Twitter," Comput. Intell., vol. 31, no. 1, pp. 132–164, 2015.
3. 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. 1936, Jan. 2014.
4. Eleonora D'Andrea, Pietro Ducange, Beatrice Lazzerini, Member, IEEE, and Francesco Marcelloni, Member, IEEE ,‖Real-Time Detection of Traffic.
5. Vikram Singh and Balwinder Saini ―An Effective Tokenization Algorithm for Information Retrieval System‖ CS & IT-CSCP 2014.
6. 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. 919931, Apr. 2013.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

7. M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, "Getting their 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.
8. N. Wanichayapong, W. Pruthipunyaskul, W. Pattara-Atikom, and P. Chaovalit, "Social-based traffic information extraction and classification," in Proc. 11th Int. Conf. ITST, St. Petersburg, Russia, 2011, pp. 107– 112.

BIOGRAPHY

Ramiz Khalik Shaikh pursuing BE in Department Of Computer Engineering, L.G.N Sapkal college of Engg, Anjneri hill, Trambakeshwer, Nashik.

Sameer Balasaheb Shinde pursuing BE in Department Of Computer Engineering, L.G.N Sapkal college of Engg, Anjneri hill, Trambakeshwer, Nashik.

Vaibhav Natha Pawar pursuing BE in Department Of Computer Engineering, L.G.N Sapkal college of Engg, Anjneri hill, Trambakeshwer, Nashik.

Prof. Prashant Kale Assistant Professor Department Of Computer Engineering, L.G.N Sapkal college of Engg, Anjneri hill, Trambakeshwer, Nashik.