ISSN: 2321-7782 (Online)
Impact Factor: 6.047

Volume 4, Issue 7, July 2016

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Traffic Detection Using Tweets on Twitter Social Network

Supriya Bhosale[1]
Department of Computer Engineering
Savitribai Phule University
Pune, India

Sucheta Kokate[2]
Department of Computer Engineering
Savitribai Phule University,
Pune, India

Abstract: ***Social networks is 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 or collect tweets from twitter and then processes tweets using text mining techniques. Lastly performs the classification of tweets. The system aim is to assign the appropriate class label to each tweet, whether it is related to a traffic event or not. System used the support vector machine as a classification model.***

The proposed system uses the 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.

I. INTRODUCTION

Now a days social networking sites are becoming the real time information channel over time. With the help of portable devices many users are able to share the real life events on the internet. This feature of the social networking sites made them more popular and valuable. Basically these social networking sites are used for the maintaining social relationship, finding friends and users with the similar interest .The text message shared by user in social networks that message is called Status Update Message [1]. In that SUM the text, meta-information like timestamp, name of the user, geographic coordinates, links of the other resources are present. The SUM's considered in a specific area may provide the proper information [1].An Intelligent transport system integrates ICTs (Information and Communication Technologies) connect with transport networks. This allows improving safety and management of transport networks in better way [4].

Twitter has many advantages over other micro-blogging services. Twitter can be used to study the real time events using SUM's. Each tweet is made of meta information which provides additional information. Another major advantage of twitter is SUM's are publically and easily available. Our works aims at the detection of road traffic events in a small region. Proposed work is useful to detect and analyse traffic related events by processing users 'SUM's which is from a specific area and written in the English language [2]. Our Proposed system is able to extract, analyse SUMs as traffic event or non-traffic event. The event detection from the social networking sites is more complicated and difficult work as compared to traditional media like blogs and emails, where texts are well formatted [2]. On social sites SUMs are not in structured form and irregular texts also it having informal words along with spelling mistakes and grammatical errors [1]. Due to very brief, small nature SUM's are becoming an incomplete source of information. The SUMs having a large number of non-useful or meaningless information with respect to different events. Such unrelated information must be filtered [2].

II. LITERATURE SURVEY

This approach is used to getting information which is useful[1] for event detection, it need to classified a small-scale and large-scale events.

Authors of[6] used twitter streams to detect earthquakes event which is a large scale event by monitoring special trigger-keywords. For detection of earthquake authors used SVM as a binary classifier of positive events and negative events [6]. In [7] shakki et.al. presented a method is used for detecting real-world event like natural disasters by analyzing Tweets by using both term-frequency-based techniques and NLP. A novel system is presented in [2] for detection and analysis of events from rich information of the twitter stream. The Authors presented the following three functionalities (1) New event detection, (2) Event ranking according to importance, and (3) collecting temporal and spatial patterns for events [2]. The work was initially focused on bad and Disaster related Events(CDE), eg. shooting,accidents etc [2]. Traffic, fires, crashes and local manifestations are small-scale events. They have a small number of SUMs related to that events. Small number of SUMs categories them to the small scale events. Small scale events belong from small time interval. Large scale events like a earthquakes, tornados, elections are characterized by a large number of SUMs. Such event has wider temporal and geographic coverage.

In reference [3]authors presented a system which analyze traffic and its causes. Twitter is a social media which allows people to share and read tweets related to all events[3]. The system is able to read the tweets from twitter and this system uses natural language processing technique on them. Then system categorizes the tweets related to traffic and notifies the registered users about it. The natural language processing focuses on developing efficient algorithms to process text. The NLP also focus to convert text into machine understandable language [3].

In paper [4] Vikram Singh et. al. proposed an effective tokenization technique based on training. The presented method results in the better tokens using tokenization along with preprocessing. If less number of token generated then minimum storage space is required, This facilitates more accuracy in results retrieval [4]. Proposed algorithm takes responsibility for reducing the time of retrieval information [4].

Maximilian Waltheret. al.in [5] detect Geo-spatial event using twitter SUM's. The proposed approach authors gathered tweets for target events that can be defined by a user through keywords [5]. Classification and particle filtering methods are used for finding this geo-spatial events [5]. Authors used common theme as if people are tweeting from the same place or area uses similar words which, means that these users are talking about that thing only[5].

III. PROPOSED SYSTEM

Proposed system architecture is service oriented, event-driven. The systems has following main functionality

1) Extraction of SUMs and Preprocessing.

2) Elaboration of SUMs

3) Classification of SUMs

Initially user tweets are fetched from the twitter stream using twitter API's. At the same time the traffic event related training must be provided to the system using trained classifier. Once the tweets are fetched further processing such as tokenization, stemming, filtering are applied on the tweets. After these task the feature selection is performed. Then using selected features the tweets are classified in traffic related and non-traffic related tweets.

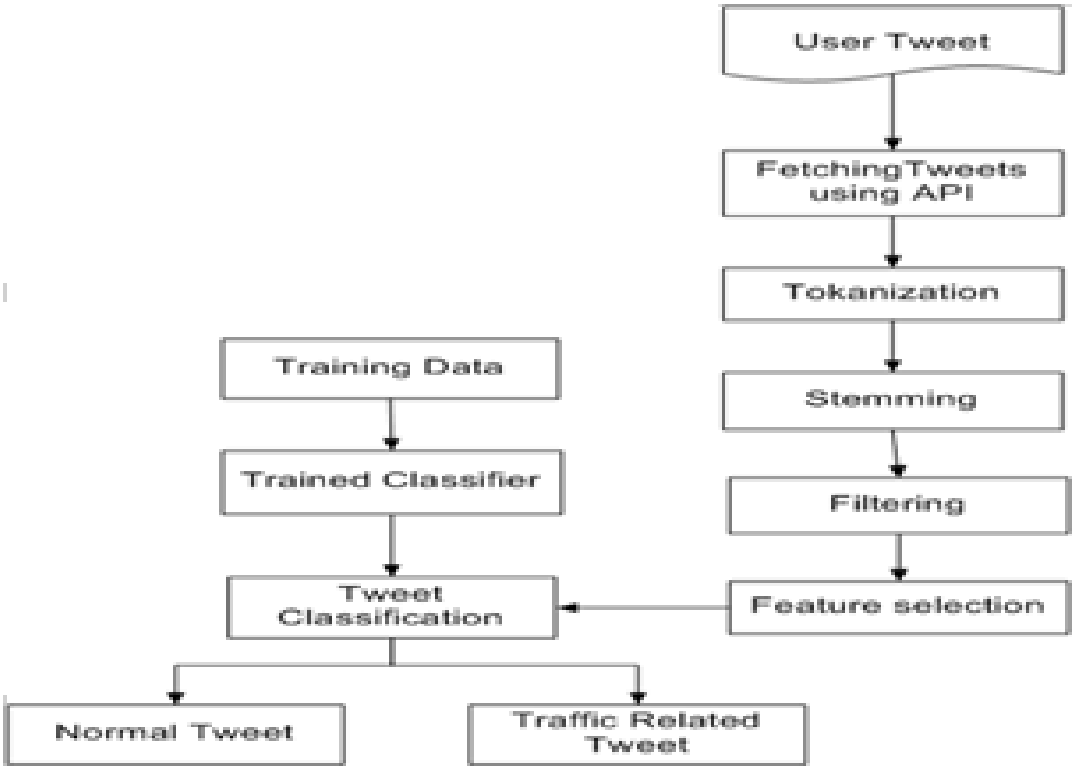

Fig. 1: Architecture

Initially user tweets are fetched from the twitter stream using twitter API's. At the same time the traffic event related training must be provided to the system using trained classifier. Once the tweets are fetched further processing such as tokenization, stemming, filtering are applied on the tweets. After these task the feature selection is performed. Then using selected features the tweets are classified in traffic related and non-traffic related tweets.

1) Fetch of SUMs and Pre-Processing

SUM's from the user profile are fetched and the further processing is applied on it to extract the raw information. Fetched raw tweet contains information like: user id, timestamp, geographic coordinate, re-tweet flag and text of the tweet [1].

2) Elaboration of SUMs

In preprocessing module elaboration of SUMs is done. This module transforms the set of strings into a set of numeric vectors are elaborated by the Classification of Status Update Message module [1]. Different text mining techniques are used in sequence to the pre-processed SUMs to achieve this. The text mining steps done in this module are

a) Tokenization is normally the first step of the text mining process which is used for transforming a stream of words into a stream of processing units called tokens. In this step different operations used like removal of other non-text characters and punctuation, and normalization of symbols [1].

b) In stop-word filtering remove unnecessary words which does not provide any information to the text analysis.

c) Stemming is the process of finding the root word from specific word. It removes its suffix.

d) Stem filtering is used to reducing the number of stems of each SUM. Each SUM is filtered by removing from the set of stems the ones not from the set of relevant stems [1].

3) Classification of SUMs

Classification of SUMs module is used to assigns each SUM a class label related to traffic events. This module outputs a collection of N labeled SUMs [1].

IV. PROPOSED ALGORITHM

Input: Training Dataset T, Test dataset D,

Output: Clustered Tweet set.

1) Initially train the classifier by using semi-supervised traffic related training dataset.

2) Fetch tweets of user from tweeter account.

3) Store it in DB

4) For each tweet in DB finding the similarity using Euclidean distance by trained data.

5) If(similarity > Threshold)

6) Add that tweet to traffic related tweet set

7) Else

8) Add to normal (not related to traffic) tweet set.

9) Return classified tweets

The proposed clustering algorithm takes the tweets from the twitter as input along with the traffic related training dataset. This approach initially trains the classifier using available training data. After that the fetched tweets are taken for classification, the similarity between the traffic related and non-traffic related tweet is calculated using the Euclidean similarity measure. Lastly this algorithm presents a set of traffic related and non-traffic related tweets.

V. RESULT AND GRAPH

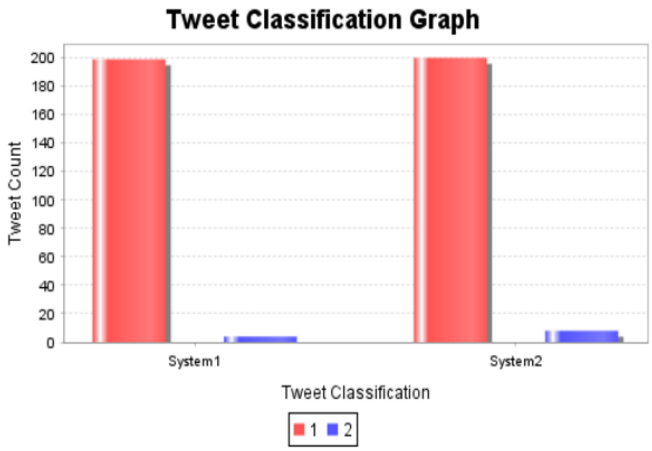

Fig.2: Tweet Classification

The above graph shows the result and comparison of existing system SVM and proposed semi-supervised method. System 1 is SVM and system 2 is semi-supervised method. Recall, precision values are greater for proposed semi-supervised approach than existing system.

VI. CONCLUSION

This paper presents a semi-supervised approach for the detection of traffic events using twitter. The proposed approach uses the social network to detect the traffic events on the go. This approach helps user to detect accident and traffic events to save the time. The proposed approach fetches the tweets at runtime and finds the appropriate class label for that tweet to classify them into traffic related tweet or non-traffic related tweets.

ACKNOWLEDGEMENT

We are thankful to prof. Sucheta Kokate and prof. N.B. Pokale for their support and guidelines to complete this work.

References

1. EleonoraD' Andrea, Pietro Ducange, Beatrice Lazzerini, Member, IEEE, and Francesco Marcelloni, Member, IEEE ,"Real-Time Detection of Traffic FromTwitter Stream Analysis",IEEE transaction on intelligent transportation system, VOL. 16, NO. 4, AUGUST 2015.
2. Rui LI, Kin Hou Lei, Ravi Khadiwala, Kevin Chen-Chuan Chang "TEDAS: a Twitter Based Event Detection and Analysis System", IJCSIT 2014.
3. Harshita Rajwani, Srushti Somvanshi, AnujaUpadhye," Dynamic Traffic Analyzer Using Twitter", International Journal of Science and Research (IJSR) 2014.
4. Vikram Singh and Balwinder Saini "An Effective Tokenization Algorithm for Information Retrieval System" CS and IT-CSCP 2014
5. Maximilian Walther and Michael Kaisser,"Geo-spatial Event Detection in the Twitter Stream", P. Serdyukov et al. (Eds.): ECIR 2013, LNCS 7814, pp. 356367, 2013 .springer Verlag Berlin Heidelberg 2013.
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. 919-931, Apr. 2013.
7. 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.
8. 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.
9. 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.
10. K. Boriboonsomsin, M. Barth, W. Zhu, and A. Vu, "Ecorouting navigation multisource historical and real-time traffic information," IEEE Trans. Intell. Transp. Syst., vol. 13, no. 4, pp. 1694-1704, Dec. 2012.

AUTHOR(S) 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.

