Vandana Publications
IJEMR

Volume-7, Issue-3, May-June 2017
International Journal of Engineering and Management Research
Page Number: 818-822

Implementation of Real Time Detection of Traffic from Twitter Stream Analysis

Vaishali Singh[1], Prof. Priti Subramanium[2]
[1] Student of Computer Science & Engineering Department, INDIA
[2]Department of Computer Science & Engineering, INDIA

ABSTRACT

The social network has been newly engaged as a source of information for event discovery, with particular mention to road traffic jamming and the car accident. In this thesis, We are going to present a real-time monitor system for traffic event detection from Twitter stream study. The system fetches tweets from Twitter according to some search criterion like processes tweets, by apply text mining techniques and finally perform the categorization of tweets. The aim is to allocate the appropriate class label to each tweet as related to a traffic incident or not. The traffic detection system was engaged for real-time monitoring of several areas of the Italian road network, allowing for detection of traffic events almost in real time, often before online traffic news web sites. We employed the support vector machine as a classification model and we achieved an accuracy value of 95.75% by solving a binary classification problem (traffic versus non-traffic tweets). We were also able to discriminate if traffic is caused by an external event or not, by solving a multiclass classification problem and obtain accuracy value of 88.89%.

Keywords-- activities, definition, possible, professional

I. INTRODUCTION

What Is A Social Network?

Google defines a social network service as a service which focuses on the online social networks for communities of people who share interests and activities or who are interested in exploring the interests and activities of others and which necessitate the use of the software.

Another definition of social networking sites is Web sites mainly designed to smooth the progress of interaction between users who share interests, attitudes and activities, such as Facebook, Mixi and MySpace.[10]

What Can Social Networks Be Used For?

Social networks can provide a range of benefits to members of an organisation:

Support for learning: Social networks can improve casual learning and support social links within groups of learners and with those involved in the hold of learning.

Support for members of an organisation: Social networks can potentially be used our all members of an organisation, and not just those involved in working with students. Social networks can help the development of communities of practice.

Engaging with others: Unreceptive use of social networks can provide important business intelligence and opinion on institutional services (although this may give rise to ethical concerns).

Ease of access to information and applications: The simplicity use of many social networking services can offer profit to users by simplifying right to use other tools and applications. The Facebook Platform provides an example of how a social networking service can be used as an environment for other tools.

Common interface: A possible benefit of social networks may be the common interface which spans work or social boundaries. Since such services are often used in a personal capacity the interface and the way the service works may be familiar, thus minimising training and support needed to exploit the services in a professional context. This can be a barrier to those who wish to have strict boundaries between work and social activities.

Examples of popular social networking services include:

Facebook: Facebook is a social networking Web site that allows people to communicate with their friends and exchange information. In May 2007 Facebook launched the Facebook Platform which provides a framework for developers to create applications that interact with core Facebook features.

MySpace: It is a social networking Web site present an interactive customer-submitted network of connections, personal profiles, blogs and groups, generally used for distribution of photos, music and videos [3].

Ning: An online stage for creating social Web sites and social networks designed for users who want to create networks about a specific benefit or have restricted technical skills.

Twitter: Twitter is an example of a micro-blogging service. Twitter can be used in a variety of ways including sharing brief information with users and providing support for one's peers.

Note that this brief list of popular social networking services omits popular social sharing services such as Flickr and YouTube [1].

Opportunities and Challenges:

The fame and easiness of social networking services have energised institutions with their potential in various areas. However the effective use of social networking services poses a number of challenges for institutions including long-term sustainability of the services; user concerns over the use of social tools in a work or study context like a variety of technical issues and legal issues such as copyright, privacy, accessibility etc. Institutions would be advised to consider carefully the implications before promoting the significant use of such services.[3]

II. SYSTEM ARCHITECTURE

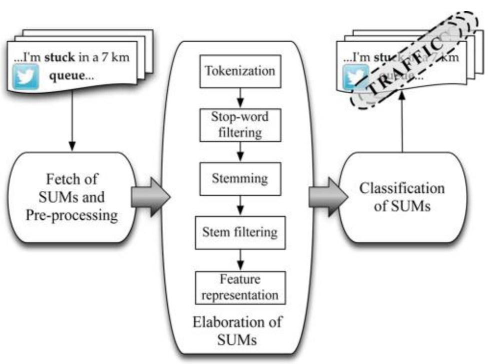

Figure 2: - System Architecture.

MODULES:

- Fetch of SUMs and Pre-Processing
- Elaboration of SUMs
- Classification of SUMs
- Setup Of the System

MODULES DESCRIPTION:

Fetch of SUMs and Pre-Processing

The first module that is Fetch of Status Update Messages and Pre-processing which extracts unprocessed tweets from the Twitter stream based on one or more search criteria, for example, geographic coordinates, keywords appearing in the text of the tweet. Each fetched raw tweet contains the user id, the timestamp, the geographic coordinates, a retweet flag, and the text of the tweet. The text may hold extra information after the SUMs have been fetched according to the definite search criteria and 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. Extra meta-information redundant are user id, timestamp, geographic coordinates, hashtags, links, mentions, and special characters. Lastly, a case-folding operation is applied to the texts, in order to convert all characters to lower case. At the end of this explanation, each fetched SUM appears as a string that is a sequence of characters.

Elaboration of SUMs

The second processing module is Elaboration of SUMs. This module devoted to transforming the set of pre-processed SUMs that is a set of strings in a set of numeric vectors to be elaborated by the Classification of SUMs module. To this plan, some text mining techniques are used in sequence to the pre-processed SUMs. In the following text mining steps performed in information, such as hashtags, links, mentions, and special characters. In this paper, we took only Italian language tweets into account. However, the system can be easily adapted to cope with different languages. This module is described in detail. 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 for example syllables, words, or phrases. Stop-word filtering consists in eliminating stop-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 like language-specific stop-words or in the set of texts being analysed domain-specific stop-words and can, therefore, be considered as noise. Stemming is the procedure of dropping each word that is taken to its branch or root form by removing its suffix. The purpose of this step is to group words with the same theme having closely related semantics.

Classification of SUMs

The third module is Classification of SUMs. This module assigns every complicated SUM as a class label related to traffic events. Thus the output of this module is a collection of *N* labelled SUMs. To the plan of labelling every SUM, a classification model is in use. The parameters of the classification model have been identified during the supervised learning stage. Actually, as it will be discussed different classification models have been considered and compared. The classifier that achieved the most accurate results were finally employed for the real-time monitoring of 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. Such as when the first tweet is recognised as a traffic-related tweet, the system may send a warning signal. Then the actual notice of the traffic event may be sent later than the identification of a convinced number of tweets with the same label.

Setup of the System

As confirmed before a supervised learning stage is required to perform the setup of the system. In particular, we need to identify the set of relevant items, the weights associated with each of them, and the parameters that describe the classification models. We use a collection of Ntr labelled SUMs as the instruction set. During the learning stage, each SUM is elaborated by applying the tokenization, stop-word filtering, and stemming steps. Finally, the tweets were manually labelled with two possible class labels as related to road traffic event (traffic) for example accidents, jams, queues, or non-traffic. The additional feature first we read then interpreted and correctly assigned a traffic class label to each candidate traffic class tweet.

III. INPUT DESIGN

The input design is the connection between the information method and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data into a usable form for processing can be achieved by inspecting the computer to convert data from a written or printed document. It can occur by having people keying the data straight into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in a way so that it provides security and simplicity and use with retain the confidentiality. Input Design considered the following things:

- What data should be given as input?
- How should the data be arranged or coded?
- The dialogue to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when the error occurs.

OBJECTIVES

1. Input Design is the process of converting a user-oriented explanation of input into a computer oriented system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerised system.

2. It is achieved by creating user oriented screens for the data access and handle the large volume of information. The goal of designing input is to make data entry easier and to be free from errors. The data access screen is planned in a way that every data manipulate and well performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as for when needed so that the user will not be in maize of instant. Thus the goal of input plan is to create an input layout that is easy to follow.

OUTPUT DESIGN

A quality output is one which meets the requirements of the end user and presents the information clearly. In any system results of processing and communicating of the users and another system through outputs. In output design, it is determined how the information is to be displaced for urgent need and also the hard copy output. It is the main straight source of information to the client. Efficient and intelligent output design improves the system's relationship to help user decision-making.

Designing computer output should proceed in an organised, well thought out manner the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. Analysis Design should identify the specific output that is needed to meet the requirements.

Select methods for presenting information. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

IV. RESULTS

We focus on a particular small-scale event that is road travel and we plan to spot and analyse traffic events by processing users' SUMs belonging to a definite area and printed in the Italian language. To this aim, we propose a system able to fetch, elaborate, and classify SUMs as related to a road traffic event or not.

Figure 4.1 illustrates how to sign in on twitter account so that we could fetch all tweets.

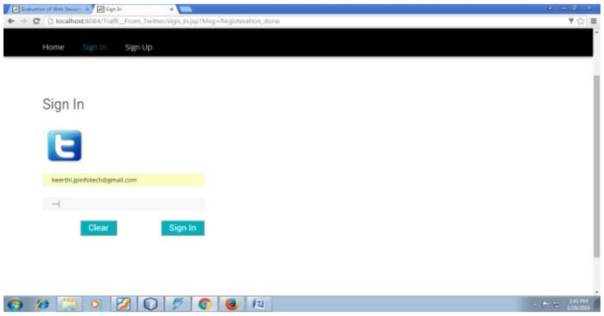

Figure 4.1: - Twitter Account Sign In

ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Tweets are up to 140 characters, enhancing the real-time and news-oriented nature of the platform. In this Figure 4.2 shows all friend lists which are added to our twitter account.

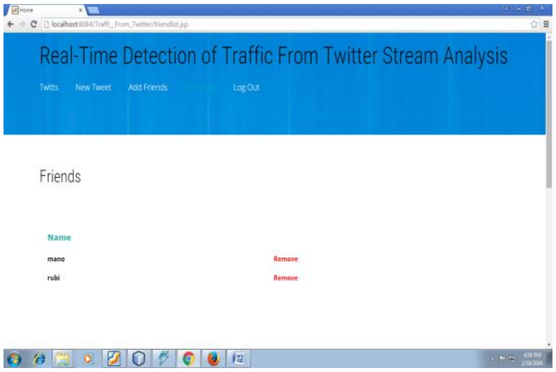

Figure 4.2: - Twitter Friend List

In fact, the lifetime 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. Figure 4.3 follows how many friends are available on our twitter account.

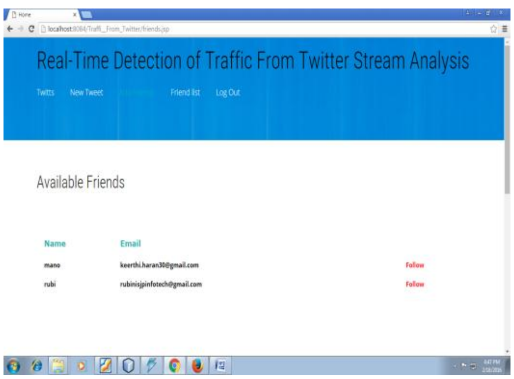

Figure 4.3: - Available Friends On Twitter

Each tweet can be directly associated with meta-information that constitutes additional information. Figure 4.4 send simple hi tweets on our account.

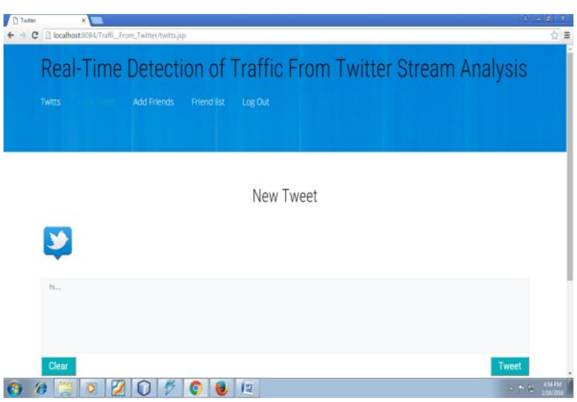

Figure 4.4: - Simple Hi Tweet

Twitter messages are public. They are directly available with no privacy limitations.

Figure 4.5 illustrates stuck on road tweets on our twitter account so that friends will be updated about this jam.

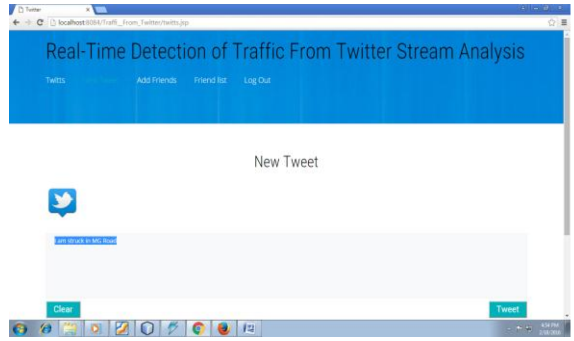

Figure 4.5: - New Tweet Stuck

For all of these reasons, Twitter is a good source of information for real-time event detection and analysis. Figure 4.6 shows traffic alert from tweets.

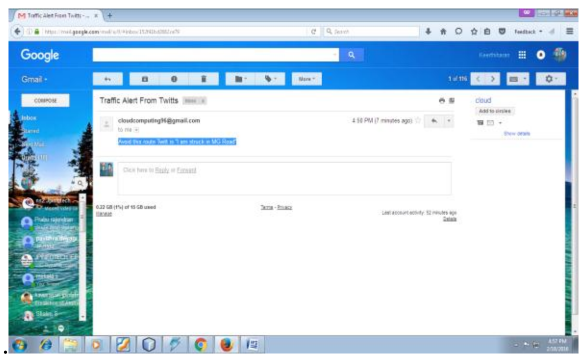

Figure4.6: - Traffic Alert From Tweets

It performs a multiclass classification, which recognises non-traffic, traffic due to congestion or crash, and traffic due to external events. After getting information about traffic congestion on road through tweets in our account then we can sign out our twitter account. Figure 4.7 shows sign out of our twitter account.

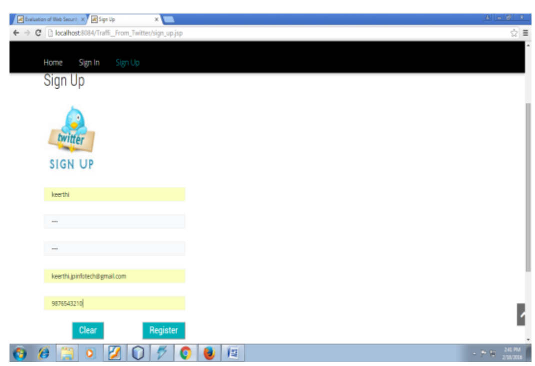

Figure 4.7: - Twitter Account Sign Up

V. CONCLUSION

In this term paper, we have proposed a system for real moment detection of traffic related actions from Twitter stream analysis. The method built on an SOA is able to fetch and classify streams of tweets and to notify the users of the presence of traffic events. We have exploited available software packages and 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 overall system for traffic event detection. Among the analysed classifiers, we have shown the superiority of the SVMs, which have achieved the accuracy of 95.75%, for the 2-class problem, and of 88.89% for the 3-class problem, in which we have also considered the traffic due to the external event class. The best classification model has been engaged for real time monitor of numerous area of the Italian road network. We have shown the results of a monitoring campaign, performed in September and early October 2014. We have discussed the ability of the system to detect traffic events almost in real time, often before online news web sites and local newspapers.

REFERENCES

[1] F. Atefeh and W. Khreich, “A survey of techniques for event detection in Twitter,” Comput. Intell., vol. 31, no. 1, pp. 132–164, 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. 613–620.

[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. 29–42.

[4] 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.

[5] J. Allan, Topic Detection and Tracking: Event-Based Information Organization. Norwell, MA, USA: Kluwer, 2002.

[6] K. Perera and D. Dias, “An intelligent driver guidance tool using location based services,” in Proc. IEEE ICSDM, Fuzhou, China, 2011, pp. 246–251.

[7] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, “Real-time event extraction for driving information from social sensors,” in Proc. IEEE Int. Conf. CYBER, Bangkok, Thailand, 2012, pp. 221–226.

[8] B. Chen and H. H. Cheng, “A review of the applications of agent technology in traffic and transportation systems,” IEEE Trans. Intell. Transp. Syst., vol. 11, no. 2, pp. 485–497, Jun. 2010.

[9] A. Gonzalez, L. M. Bergasa, and J. J. Yebes, “Text detection and recognition on traffic panels from street-level imagery using visual appearance,” IEEE Trans. Intell. Transp. Syst., vol. 15, no. 1, pp. 228–238, Feb. 2014.

[10] 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.

