www.ijemr.net ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Volume-7, Issue-3, May-June 2017

International Journal of Engineering and Management Research

Page Number: 823-827

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, attitudes, programming

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.

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.

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.

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.

II. METHODOLOGY

Java Technology

It is both a programming language and a platform**.** Java programming language is a high-level language that can be illustrated as Simple, Architecture neutral, Object oriented, Portable, Distributed, High performance, Interpreted, Multithreaded, Robust, Dynamic, Secure.

With most programming languages, we either compile or interpret a program so that we can run it on our computer. This language is unusual and in that language, a program will be both compiled and interpreted. With the compiler, first, you translate a program into an intermediate language called Java byte codes —the platform-independent codes interpreted by the interpreter on the Java platform. The interpreter will run each Java byte code instruction on the computer. Compilation happens only once but interpretation occurs each time the program is executed. The following figure 1, illustrates how this works.

1. General J2ME architecture.

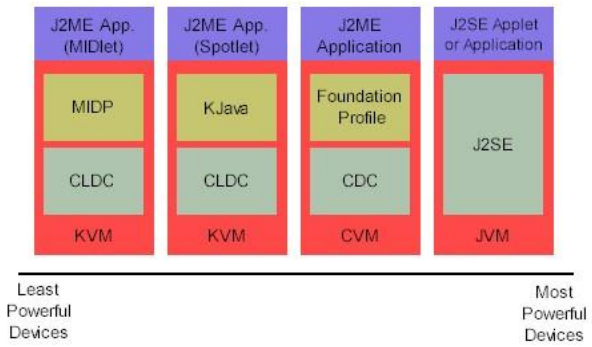

Figure 1: - J2ME Architecture.

J2ME use configurations and profile to modify the Java Runtime Environment (JRE). As a complete Java Runtime Environment, J2ME is comprised of a configuration, which determines the Java Virtual Machine used, and a profile, which defines the application by adding domain-specific classes. The configuration defines the essential run-time environment as a set of core classes and a precise Java Virtual Machine that run on specific types of devices. We'll discuss configurations in detail. The profile defines the application specifically. It adds domain-specific classes to the J2ME pattern to describe the firm use for devices. We'll cover profile in depth. The subsequent graphic depict the relationship between the different virtual machines, configurations, and profiles. It also draws a parallel with the J2SE API and its Java virtual machine. However J2SE virtual machine is usually referred to as a Java Virtual Machine. The J2ME virtual machines, KVM and CVM, are subsets of JVM. Both KVM and CVM can be thought of as a kind of Java virtual machine. It's just that they are shrunken versions of the J2SE JVM and are specific to J2ME.

SYSTEM ANALYSIS EXISTING SYSTEM:

In recent times social networks and media platforms have been commonly used as a source of information for the detection of events such as traffic jamming incident, the natural disaster like earthquakes, storms, fires etc or other events.

Sakaki et al provide work for Twitter streams to spot earthquakes and typhoons by monitoring unique trigger-keywords and by applying an SVM as a binary classifier of positive events like earthquakes and typhoons and negative events are non-events.

Agarwal et al fosses on finding the fires in a factory from Twitter stream analysis by using standard NLP technique and a Naive Bayes (NB) classifier.

Li et al. offer a system called TEDAS, to recover 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 retweets, hashtags, links, and mentions.

Disadvantages of existing system

Incident exposure from social networks analysis is an additional challenging problem than event detection from traditional media like blogs, emails etc where texts are well formatted. Status Update Messages are shapeless and asymmetrical texts, they contain informal or abbreviated words, misspellings or grammatical errors.SUMs contain a huge amount of not useful or meaningless information.

PROPOSED SYSTEM:

In this paper we plan an smart system based on text mining and machine learning algorithms for real-time detection of travel 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 incorporated to organise 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.

We focused on a particular little-scale incident that is road traffic and we planned to notice and analyse traffic events by processing users SUMs belong to a certain area and written 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.

To the most excellent of our information, 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.

ADVANTAGES OF PROPOSED SYSTEM:

Tweets are up to 140 characters, enhancing the real-time and news-oriented nature of the platform. 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. 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. Furthermore the proposed system 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 and recognises non-traffic events that are traffic due to jamming or collide, and traffic due to external events. It detects the traffic events in real-time & it is developed as an event-driven infrastructure, built on an SOA architecture.

III. PRIOR APPROACH

SYSTEM ARCHITECTURE

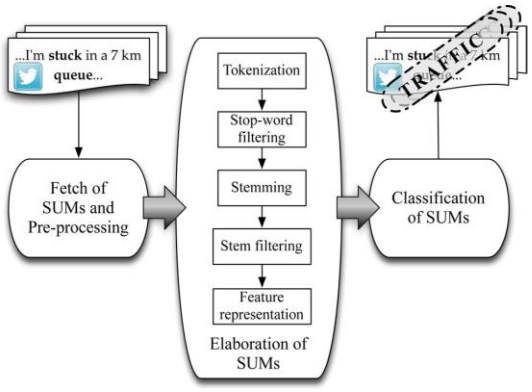

Figure 2: - System Architecture

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 which achieves the most accurate outcome was finally engaged 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.

IV. OUR APPROACH

DATA FLOW DIAGRAM:

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to signify a system in conditions of input data to the system, various processing carried out on this data, and the output data is generated by this system.

The data flow diagram (DFD) is one of the most important modelling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformation that is used as data moves from input to output.

DFD is also known as bubble chart. A Data Flow Diagram may be used to characterise a system at any level of generalisation. DFD may be partition into the level that represents increasing information flow and purposeful feature.

UML DIAGRAMS

UML stands for Unified Modeling Language. It is standardized modeling language in object-oriented software engineering. The standard is managed and was created by the Object Management Group.

The goal is for Unified Modeling Language to become a common language for creating models of object oriented computer software. In present category UML is comprise two major components that is a Meta-model and a notation. In the future a little form of method or process may also be added or associated with UML. The UML uses mostly graphical notations to express the design of software projects.

GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide extensibility and specialisation mechanisms to extend the core concepts.
2. Be free of exacting programming languages and development method.
3. Make available a formal origin for understanding the modelling language.
4. Encourage the growth of Object Oriented tools in market.
5. Integrate best practices.

USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioural diagram defined and created by a Use-case study. Its reason is current graphical overview of the functionality provided by a system in terms of actors their goals represented as use cases and any dependencies between those use cases. The major reason of a use case diagram is to demonstrate what system functions are performed by which actor. Roles of the actors in the system can be depicted.

CLASS DIAGRAM:

A class diagram that describes the structure of a system by viewing the system's classes, their attributes,

operations or methods and the associations among the classes. It explains which class contains information.

SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is the interaction of diagram that shows how processes operate and in what order. It is a build of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

ACTIVITY DIAGRAM:

Activity diagrams are graphical overview of workflows activities and actions with support for option, iteration and concurrency. Within Unified Modeling Language the activity diagrams can clarify the business and operational bit by bit workflows of components in a system. An activity diagram taken as a whole flow of control.

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.