ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Novel Method to Study and Analyze the Traffic from Twitter Stream

Thiragabattina Sai Swapna
M.Tech Student,
Vignana Bharathi Institute of Technology.

Rajasekhar Jelli
Assistant Professor,
Vignana Bharathi Institute of Technology.

Abstract

Internet sites are source of info for event detection, with specific mention of the road traffic activity blockage and accidents or earth-quack sensing system. In this paper, we present a real-time monitoring system intended for traffic occasion detection coming from Twitter stream analysis. The system fetches tweets coming from Twitter as per a several search criteria; methods tweets, by applying textual content mining methods; last but not least works the classification of twitter posts. The goal is to assign suitable class packaging to every tweet, because related with an activity of traffic event or perhaps not. The traffic recognition system or framework was utilized for real- time monitoring of various areas of the street network, taking into account detection of traffic occasions just almost in actual time, regularly before on-line traffic news sites. All of us employed the support vector machine like a classification unit, furthermore, we accomplished a great accuracy value of ninety five. 75% by attempting a binary classification issue. All of us were also capable to discriminate if traffic is triggered by an external celebration or not, by resolving a multiclass classification issue and obtaining accuracy worth of 88. 89%.

Key Words: ***Social media; Traffic detection; Text mining; Privacy; Service Oriented Architecture (SOA), machine learning, Twitter stream analysis***

Introduction

Social media platforms are widely used for distributed information about the detection of events, such as traffic blocking, incidents, natural disasters (earthquakes, storms, fires, etc.), or other events. An event is defined as a real- world existence that happens in adefinite time and space [1], [7].

Generally traffic related events, people frequently share by means of an SUM information about the current traffic situation around them while driving. For this purpose, event detection from social networksis also often employed with Intelligent Transportation Systems(ITSs). ITSs afford, e.g., real-time information about weather, traffic congestion or regulation, or plan efficient (e.g., shortest, fast driving, least polluting) routes[4], [6], [8]. Event detection from social networks investigation is a more stimulating problem than event detection from traditional broadcasting like blogs, emails, etc. In fact, SUMs are unstructured and unequal texts, it holds informal or shortened words, mistakes or grammatical errors [1]. SUMs contain a huge amount of not useful or wordless information, which has to be clarified. According to Pear Analytics, it has been estimated that over 40% of all Twitter2 SUMs (i.e., tweets) is senseless with no useful data for the audience.

For all of these reasons, in order to analyze the data coming from social networks or text mining techniques, We use to extract important data [18],of

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

data mining, device learning, numbers, and Natural Language Processing (NLP).In this paper, we propose a brilliant system, based upon text mining and equipment learning algorithms, for current detection of traffic occasions from Twitter stream evaluation.

The system, after having a feasibility study, offers been designed and created from the ground because an event-driven infrastructure, constructed on a Service Focused Architecture (SOA). The program exploits available technologies on state-of- the-art processes for analysis and pattern.These technologies have been analysed, adapted, and integrated to be able to build the intelligent system. In particular, we present a great experimental study, which been performed for the most effective different state-of- the-art approaches text classification.

EXISTING SYSTEM:

- 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 retweets, hashtags, links, and mentions.

DISADVANTAGES OF EXISTING SYSTEM:

- Event detection from social networks analysis is a more challenging problem than event detection from traditional media like blogs, emails, etc., where texts are well formatted.
- SUMs are unstructured and irregular 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 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).
- The system exploits available technologies based on state-of-the-art techniques for text analysis and pattern classification. These technologies and techniques have been analyzed, 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 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, we propose a system able to fetch, elaborate, and classify SUMs as related to a road traffic event or not.

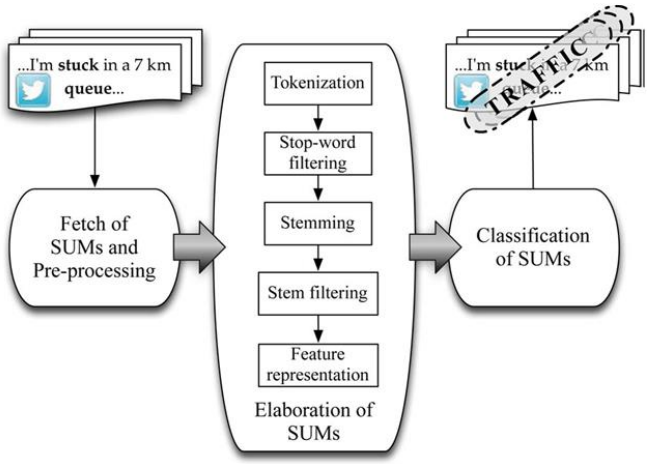
ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

- 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.

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 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, which recognizes non-traffic, traffic due to congestion or crash, and traffic due to external events
- It detects the traffic events in real-time; and iii) it is developed as an event-driven infrastructure, built on an SOA architecture.

SYSTEM ARCHITECTURE:



Mathematical Model

Let S is the Whole System Consists: S= {I, P, O}

I = Input.

P= Procedure

O= Output.

I = {U, T, TS, url, Tk}.

1. Let U is set of number of twitter users in the system.

U = {u1, u2,……un}.

2. T is set of number Twitt or status update of twitter user.

T= {t1, t2, t3…tn}.

3. TS is twitter streamer who analyzes the twits.

4. url is the URL of twitter user who have updated status.

5. Tk is the tokenization of SUM where, SUM is te Status Update Message of twitter user.

P = Procedure.

Step 1: The twitter streamer will collect all the urls from the SUM by users.

$$SUM_j^T = \{t_{j1}^T, \ldots, t_{jh}^T, \ldots, t_{jH_j}^T\}.$$

Where t_{jh}^T is the hth token and H_j is the total number of tokens in SUM_j^T.

Step 2: Filtering: In this step we perform tokenization of SUM and filtered the tokens and ignoring small meaning that is the word which don't have any information which is known as stop-word filtering.

Each SUM is reduced to a sequence of relevant tokens. We denote the jth stop-word filtered SUM as,

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

$SUM_j^{SW} = \{t_{j1}^{SW}, \ldots t_{jk}^{SW}, \ldots t_{jK}^{SW}\}$.

Where t_{jk}^{SW} the kth relevant token and Kj, is with Kj ≤ Hj, is the total number of relevant tokens in SUM_j^{SW}

Step 3: Assigning labels to filtered Tokens:
In this step, system assigns a class label to each SUM related to traffic events. So at last there is collection of N labelled SUMs.

Step 4: In this the classifier that achieved the most accurate results by filtered tokens with labels was finally employed for the real time monitoring with the proposed traffic detec- tion system.

O = Output:
When the first tweet is recognized as a traffic related tweet, the system may send a warning signal. Then, the actual notification of the traffic event may be sent after the identification of a certain number of tweets with the same label.

CONCLUSIONS

We advancing explored the authentication as well as trust and reputation calculation and management of CSPs and SNPs, which are two very critical and hardly explored issues with respect to CC and WSNs integration. Further, we proposed a novel ATRCM system for CC-WSN integration. Discussion and analysis about the authentication of CSP and SNP as well as the trust and reputation with respect to the service provided by CSP and SNP have been presented, followed with detailed design and functionality evaluation about the proposed ATRCM system. All these demonstrated that the proposed ATRCM system achieves the following three functions for CC-WSN integration:

1) Verifying CSP and SNP to avoid spiteful impersonation attacks.
2) Calculating and handling trust and reputation regarding the service of CSP and SNP.
3) Helping CSU choose required CSP and assisting CSP in selecting appropriate SNP,

Based on
(i)The authenticity of CSP and SNP;
(ii)The attribute requirement of CSU and CSP;
(iii)The cost, trust and reputation of the service of CSP and SNP.

REFERENCES:

[1] Eleonora D'Andrea, 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] Alan Mislove, MassimilianoMarcon, Krishna P. Gummadi, Peter Druschel, Bobby Bhattacharjee., "Measurement and Analysis of Online Social Networks," Proc. IEEE, vol. 99, no. 1, pp. 149–167, Jan. 2011.

[3] 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.

[4] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving Programming, Bangalore, India, Jan 2010, pp. 147-158.

[5] M. Yuriyama and T. Kushida, "Sensor-cloud infrastructure Physical sensor management with virtualized sensors on cloud computing," in Proc. 13th Int. Conf. Netw.,-Based Inf. Syst., Sep. 2010, pp. 1–8

[6] 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. 19–36, Jan. 2014.

[7] Álvaro González, Luis M. Bergasa and J. Javier Yebes, "Chen and Harry H.," ACM Trans. Sensor Netw., vol. 5, no. 2, Mar. 2009, Art. ID 10.