IJMTARC – VOLUME – V – ISSUE - 17 – MAR 2017 ISSN: 2320-1363

ANALYZING THE TWEETS AND DETECT TRAFFIC FROM TWITTER ANALYSIS

V. SRIKANTH(MCA, MTECH, MBA)

Abstract

Social networks have already been recently hired as a source of information for event uncovering, with specialized connection with road traffic activity bottleneck and accidents or earth quack summarizing system. In our paper, we show a problem-solving time exposure of traffic deriving out of Twitter stream evaluation. The system fetches tweets from Twitter as per a number of search criteria; deal with tweets by applying text tapping methods; finally performs the designation of tweets. The aim is to assign suitable class label to every tweet, as relevant with an job of traffic event or not. The traffic detection system or scheme was utilized for real-time monitoring of many areas of the Italian street network, considering detection of traffic events just most in real time, frequently before online traffic news sites. We employed the support vector machine as a distribution model, furthermore, we adept an efficiency value of 90.75% by tackling a binar classification issue (traffic as opposed to non traffic tweets). We were too able to specify if traffic is caused by an external event or rather, by solving a multiclass regulation issue and obtaining efficiency value of 80.89%.

Keywords - Twitter, twitter stream analysis, traffic event detection, tweet classification, text mining, social sensing.

I. INTRODUCTION

Twitter is virulent tweets containing URLs for unsolicited mail, phishing, and malware distribution. Conventional Twitter unsolicited mail find schemes utilize report of features similar to the rate of tweets containing URLs and the account creation date, or relationship features in the Twitter graph. These detection schemes are in effective against feature fabrications or consume much time and resources. Conventional suspicious URL detection schemes apply a variety of features inclusive of lexical features of URLs, URL redirection, HTML content, and progressive action. However, evading techniques similar to time-based circumvention and crawler evasion exist.

In our study, we suggest an original system, based on text mining and natural language processing algorithms, for real-time exposure of traffic events from Twitter stream analysis. The system, after a usefulness study, has been designed and refined from the floor as an event-driven framework, built on a Service Oriented Architecture (SOA). The system exploits accessible technologies according to state of the art techniques for text evaluation and pattern classification. These technologies and

IJMTARC – VOLUME – V – ISSUE - 17 – MAR 2017 ISSN: 2320-1363

strategies happen to be analyzed, tuned, becoming, and mixed in order to build the intelligent system. In unique, we show an experiential study, which has been performed for determining the most effective in association with the different state of the art methodes for text classification. The chosen method was combined into the ultimate structure and used for the on-the-field real-time detection of traffic events.

In our study, we propose an imaginative system, according to text mining and natural language processing algorithms, for problem-solving time detection of network events from Twitter stream analysis. The process, after a feasibility study, out-of-date designed and advanced from the ground as an event-driven framework, build on a Service Oriented Architecture (SOA). The system exploits accessible technologies according to state of the art techniques for text evaluation and pattern classification. These technologies and methods have been analyzed, tuned, becoming, and combined so as to build the intelligent system. In particular, we present an developmental study, that has been performed for figuring out 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.

With connection with current approaches for the use of social media to extract useful message for event detection, we need to correlate small events and huge events. Small-scale acts (e.g., traffic, car crashes, fires, or local manifestations) usually have a small number of SUMs related to them, belong to a precise geographic location, and are fixed in a worthless interval. On any other hand, large scale actions (e.g., earthquakes, tornados, or the selection of a head of state) are in reference to a huge number of SUMs, and by a much wider temporal and geological broadcasting. Consequently, because of the smaller number of SUMs associated with small events, small event detection is a non-trivial task. Several fill in the literature deal with action detection from social networks. Many works manage huge event detection, and only a few works focus on small-scale action. Regarding small event detection, the uncovering of fires in a factory from Twitter stream evaluation, by using usual NLP techniques and also a Naive Bayes (NB) classifier.

In that project, we center around a well known small event, i.e., road traffic, and we aim to detect and analyze traffic facts by processing users' SUMs associated with a particular zone and signed inside the Italian terminology. To this aim, we propose a structure able to produce, produce, and classify SUMs as associated with a road traffic action or not.

II. PROPOSED SYSTEM

We focus on a particular small event, i.e. Road traffic, and we aim to detect and resolve traffic actions by processing users' SUMs belonging to a particular area and written in the Italian language. So

IJMTARC – VOLUME – V – ISSUE - 17 – MAR 2017 ISSN: 2320-1363

that we suggest a system that will ready to fetch, produce, and analyze SUMs as associated with a road traffic event or not. The suggested system may approach the two binate and multi-class regulation problems. As regards paired regulation, we consider traffic-related tweets, and tweets not related with traffic. We use Multi-class classification, to split the traffic-related class into two classes, namely traffic congestion or crash, and traffic because of external event. For this regulation we use hash tag principal

in our system.

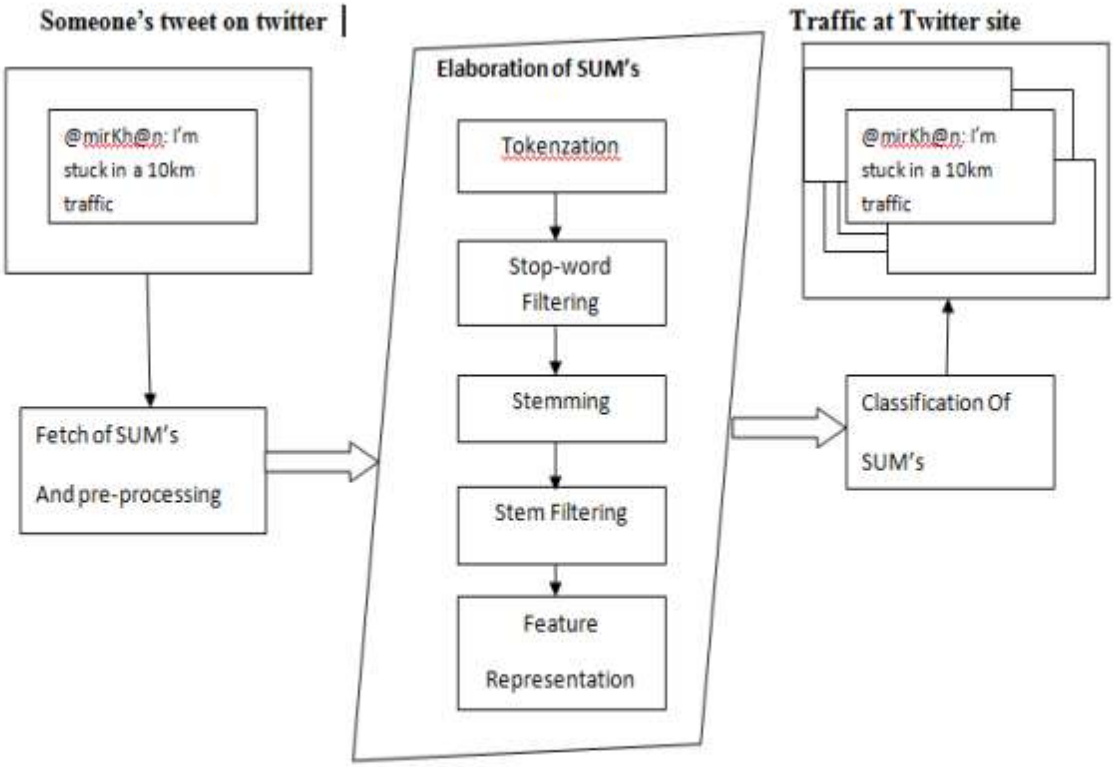

Fig. 1. System architecture for traffic detection from Twitter stream analysis.

III. ALGORITHM

1) Fetch of SUMs and Pre-Processing

The first module, "Fetch of SUMs and Pre-processing", withdraws raw twitters from the Twitter stream, based on a number of search criteria (e.g., geological coordinates, keyword found in the text of

IJMTARC – VOLUME – V – ISSUE - 17 – MAR 2017 ISSN: 2320-1363

your twee t). Each fetched raw twitter contains: the user id, the time stamp, the geological coordinates, a retweet flag, and the text of your tweet. The text may contain additional message, such as hashtags, links, mentions, and unique characters.

After the SUMs have already been fetched consistent with the particular search criteria, SUMs are pre-processed. In request to elicit only the text of every raw tweet and remove all meta instruction linked to it; a Regular Expression filter out is applied.

2) Elaboration of SUMs

The double processing segment, "Elaboration of SUMs", adopt transforming the set of pre-processed SUMs, i.e., a set of strings, within a set of numerical vectors to be elaborated individually "Classification of SUMs" segment. To this aim, a few paragraph mining techniques are utilized sequential to the preprocessed SUMs. In the ensuing, the text mining steps performed in this unit are described in detail:

a) tokenization is usually the first step of the text mining process, and is composed in transforming a stream of characters into a stream of processing units called tokens e.g., syllables, words, or phrases. The tokenizer removes all punctuation marks and splits each SUM into tokens corresponding to words (bag-of-quarrel representation). At the top of this step, each SUMj is described because the sequence of words contained in it

b) stop-word filtering consists in removing stop-words, i.e., words which supply very little info to the text analysis. Common stop-words are articles, conjunctions, prepositions, pronouns, etc. Other stop-words are those having no analytical relevance, that is, those that usually materialize quite often in sentences of the considered language (languagespecific stop-words), or in the set of texts being analyzed (domain-specific stop-words), and can then be thought to be as noise.

c) stemming is the treat of decreasing each word (i.e., token) to its stem or root form, by removing its postfix. The scheme of this step is to group words using the same theme having strictly associated semantics.

d) stem filtering is composed in decreasing method of stems of each SUM. In separate, each SUM is filtered by removing from the set of stems the ones not associated with the set of suitable stems.

e) feature representation consists in a building, for each SUM, the corresponding bearing of numerical features. Indeed, so as to analyze the SUMs, we have to represent them in the same innovation space.

IJMTARC – VOLUME – V – ISSUE - 17 – MAR 2017 ISSN: 2320-1363

3) Classification of SUMs

The third segment, "Classification of SUMs", assigns separately elaborated SUM a class label associated with traffic acts. Thus, the output of this segment is a selection of N labeled SUMs. To the purpose of labeling every single SUM, a classification model engage in. The parameters of your classification model have been identified in the course of the supervised learning stage. The classifier that achieved the most accurate results was ultimately employed for the real time monitoring with the recommended traffic detection system. The system repeatedly monitors a special place and notifies the presence of a traffic event on the basis of a set of rules that can be defined by the system administrator.

For example, when the first tweet determine as a traffic-related tweet, the system may send a warning signal. Then, the particular notification of the traffic event might be sent after the recognition of a certain variety of tweets with the same label.

IV. CONCLUSION

In this paper, we have proposed 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.

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] 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] 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.

IJMTARC – VOLUME – V – ISSUE - 17 – MAR 2017 ISSN: 2320-1363

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

[7] 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

Author's Details:

V. Srikanth(M.Tech, Mca,Mba), Received His Mca From Bharat Institute Of Engineering And Technology Affiliated To Jawaharlal Nehru University Hyderabad And Received The M.Tech Computer Science And Technology With Specialization In Computer Science From Auroras Research And Technological Institute Affiliated To Jawaharlal Nehru Technological University, Hyderabad . Mba(Hrm) And Pgdbm(Hr) Received From Jaipur National Institutes And Mits School Of Distance Education, Now He Is Working As Java Academic Project Trainer In SS Infotech, Hyderabad,Telangana. His Area Of Interest Incuding C, C++ And Java Are Artificial Intelligence, Ai Techniques, Web Technologies, Computer networks And Phython