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

The Movement in Real-Time Analysis to Find the Twitter Stream

K.Jaya Shree
Assistant Professor,
Christu Jyothi Institute of Technology.

Ch.Prudvini,
Assistant Professor,
Christu Jyothi Institute of Technology.

ABSTRACT:

Social network congestion and vehicle road traffic accident with reference to the source of information for the detection of the event, as has been used in recent years. In this paper, we analyze the traffic situation and to identify the Twitter stream to provide real-time monitoring system. The system according to different search criteria, brings tweets from Twitter. Tweets are text-mining operations through the application of techniques, and ultimately lead Rating tweets. The traffic incident or not, appropriate class tag is set for each Tweet. The traffic on the Internet news sites, often in different areas of the Italian road network traffic in real-time monitoring system for detection work, and to allow for the real-time traffic incident detection. We support vector machine is used as a model site, and we are the site of the binary problem (non-commercial traffic tweets v) solution has achieved 95.75% accuracy standards. We have the distinction of traffic through an external event or not, multiclass classification problems and were able to get the value of precision.

I.INTRODUCTION:

Recently, social networks and media outlets such as traffic jams, accidents and natural disasters (earthquakes, storms, fires, etc.), or other events as well as events, to be used as a source of information for detection. Sakaki and others. Keywords trigger the monitoring, using the Twitter stream to detect earthquakes and hurricanes, and positive events (earthquakes and typhoons) and adverse events (non-events or other occasions) as a binary SVM applying seed. Agarwal et al. NLP and naive Bayes (NB) using standard techniques workbook, Twitter stream analysis to focus on the detection of a fire in the factory. Lee et al. TEDAS the proposed system, to restore tweets about the event. In this system the fire, thunderstorms, car accidents and crime, as well as events related crashes (CDE- of mind) focuses on, and remember the events of CDE- Keywords spatial and temporal information, and the user's followers and the restoration of a number of hash , links, and the rating of the United States of exploiting the nomination aims to tweets.

Social network analysis is where events such as formatted text, blogs, e-mails and other issues, such as the traditional media are much more difficult to detect the event. Unstructured text and the occasional sound of the non-formal or short, contain spelling or grammatical errors. With the huge amount of data is useful or useless. In this paper, we analyze the Twitter streams of text mining algorithms and machine learning to detect the traffic in real-time, based on an intelligent system have been proposed. System, and the feasibility study has been designed and developed from the ground infrastructure, SOA architecture (SOA), based on the event-driven. Systems for the analysis of text and pattern-site state-of-the-art techniques based on the exploitation of the technology available. These technologies and techniques to analyze, tune and adaptive, and integrated to create an intelligent system. In particular, we classify the various state-of-the-art approach to the text of the present experimental study, which was conducted to determine the most effective. Once the system is integrated into the system and real-time traffic incident has been used to identify fields.

In this paper, we have a specific event on a smaller scale, no traffic on the streets, we exaggerated the users belonging to a specific area to detect and analyze the traffic incident and aim to focus on writing in Italian language processing. In order to achieve this goal, we have the system, not on the streets or in the event that relate to the mode of traffic, and the site is able to bring that recommendation. To our knowledge, for the detection of traffic using twitter stream analysis has suggested that some of the papers. However, with respect to our work, all of them, focusing on the language of the input feature a variety of Italian and / or feature selection algorithm employed, and only consider bilateral classifications. Tweets to 140 characters, and the real-time nature of the news media and platforms. In fact, the life-time favorites are usually very small, and therefore, suitable for the study of Twitter is related to events in real-time on a social network platform. Additional information is up to each of tweets that can be connected directly with descriptive information. Twitter messages in public, that is, they are directly without any confidentiality restrictions.

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

For these reasons, the Twitter real-time analysis to detect the event is a good source of information. To provide coverage of a wide range of low-cost road network, with the addition of traffic sensors, the system can offer a job (for example, rings, cameras, infrared cameras to detect) and the monitoring of the traffic problem is exhibited, especially in those areas where traditional motion sensors (eg, city and suburbs), the. Because it recognizes the event of non-commercial, in which the multi-layer, and due to traffic congestion or disaster sites, and traffic will take place. It shows real-time traffic incident. And iii) and SOA framework, which was built on an infrastructure-driven event, as it developed.

II. RELATED WORK:
A. FETCH OF SUMS AND PRE-PROCESSING

The first module, "Fetch of SUMs and Pre-processing", extracts raw tweets from the Twitter stream, based on one or more search criteria (e.g., 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 contain additional 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. After the SUMs have been fetched according to the specific search criteria, 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 [33] is applied. More in detail, the meta-information discarded are: user id, timestamp, geographic coordinates, hashtags, links, mentions, and special characters.

B. ELABORATION OF SUMS:

The second processing module, "Elaboration of SUMs", is devoted to transforming the set of pre-processed SUMs, i.e., a set of strings, in a set of numeric vectors to be elaborated by the "Classification of SUMs" module. To this aim, some text mining techniques are applied in sequence to the pre-processed SUMs. In the following, the text mining steps performed in this module are described in detail:

a. 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 (e.g., syllables, words, or phrases). During this step, other operations are usually performed, such as removal of punctua tion and other non-text characters [18], and normalization of symbols (e.g., accents, apostrophes, hyphens, tabs and spaces). In the proposed system, the tokenizer removes all punctuation marks and splits each SUM into tokens corresponding to words (bag-of-words representation).

a.stop-word filtering consists in eliminating stop-words, i.e., 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 (language-specific stop-words), or in the set of texts being analyzed (domain-specific stop-words), and can therefore be considered as noise [34]. The authors in [35] have shown that the 10 most frequent words in texts and documents of the English language are about the 20–30% of the tokens in a given document. In the proposed system, the stop-word list for the Italian language was freely downloaded from the Snowball Tartarus website6 and extended with other ad hoc defined stop-words. At the end of this step, each SUM is thus reduced to a sequence of relevant tokens.

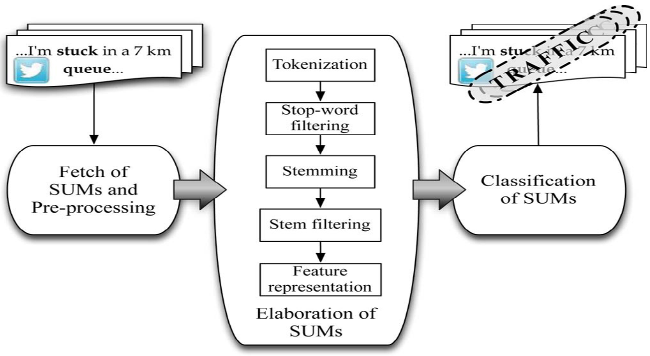

Fig: Basic System Architecture

C.CLASSIFICATION OF SUMS:

The third module, "Classification of SUMs", assigns to each elaborated SUM a class label related to traffic events. Thus, the output of this module is a collection of N labeled SUMs. To the aim of labeling each SUM, a classification model is employed. The parameters of the classification model have been identified during the supervised learning stage.

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

Actually, as it will be discussed in Section V, different classification models have been considered and compared. The classifier that achieved the most accurate results was finally employed for the realtime monitoring with 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.

II.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. We have exploited available software packages and 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 overall system for traffic event detection. Among the analyzed classifiers, we have shown the superiority of the SVMs, which have achieved 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 external event class.

REFERENCE:

[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] G. Anastasi et al., “Urban and social sensing for sustainable mobility in smart cities,” in Proc. IFIP/IEEE Int. Conf. Sustainable Internet ICTSustainability, Palermo, Italy, 2013, pp. 1–4.
[5] A. Rosi et al., “Social sensors and pervasive services: Approaches and perspectives,” in Proc. IEEE Int. Conf. PERCOM Workshops, Seattle, WA, USA, 2011, pp. 525–530.
[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] J. Allan, Topic Detection and Tracking: Event-Based Information Organization. Norwell, MA, USA: Kluwer, 2002.
[8] K. Perera and D. Dias, “An intelligent driver guidance tool using location based services,” in Proc. IEEE ICSDM, Fuzhou, China, 2011, pp. 246–251.
[9] 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.
[10] 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.
[11] 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.
[12] 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.
[13] P. M. d’Orey and M. Ferreira, “ITS for sustainable mobility: A survey on applications and impact assessment tools,” IEEE Trans. Intell. Transp. Syst., vol. 15, no. 2, pp. 477–493, Apr. 2014.
[14] K. Boriboonsomsin, M. Barth, W. Zhu, and A. Vu, “Eco-routing navigation system based on multisource historical and real-time traffic information,” IEEE Trans. Intell. Transp. Syst., vol. 13, no. 4, pp. 1694–1704, Dec. 2012.
[15] J. Hurlock and M. L. Wilson, “Searching twitter: Separating the tweet from the chaff,” in Proc. 5th AAAI ICWSM, Barcelona, Spain, 2011, pp. 161–168.