IN REAL-TIME ANALYSIS TO FIND THE TWITTER STREAM

NATHA PRASANTHI[#1], CH. LAVANYA SUSANA [*2] and SAYEED YASIN[*3]

[#] *Student, M.Tech (C.S.E), Nimra College of Engineering & Technology, A.P., India.*

[*2]*Assistant professor , Dept. of Computer Science & Engineering, Nimra College of Engineering & Technology, A.P., India.*

[*3]*Associate professor & Head, Dept. of Computer Science & Engineering, Nimra College of Engineering & Technology, A.P., India.*

Abstract— **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. 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.**

Index Terms— Sensor Network, Tweet Classification, Text Mining, Social Sensing.

I. INTRODUCTION

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. 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.

II. 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.

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.

A. 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.

III. LITERATURE SURVEY

A. *What's Happening: A Survey of Tweets Event Detection*

Author: Amina Madani, Omar Boussaid, Djamel Eddine Zegour and Algiers, Algeria

Twitter is now one of the main means for spread of ideas and information throughout the Web. Tweets discuss different trends, ideas, events, and so on. This gave rise to an increasing interest in analyzing tweets by the data mining community. Twitter is, in nature, a good resource for detecting events in real-time. In this survey paper, authors have presented four challenges of tweets event detection: health epidemics identification, natural events detection, trending topics detection, and sentiment analysis. These challenges are based mainly on clustering and classification. We review these approaches by providing a description of each one. These last years have been marked by the emergence of microblogs. Their rates of activity reached some levels without precedent. Hundreds of millions of users are registered in these microblogs as Twitter. They exchange and tell their last thoughts, moods or activities by tweets in some words.

B. *ET: Events from Tweets*

Author: Ruchi Parikh, Kamalakar Karlapalem.

Social media sites such as Twitter and Facebook have emerged as popular tools for people to express their opinions on various topics. The large amount of data provided by these media is extremely valuable for mining trending topics and events. In this paper, we build an efficient, scalable system to detect events from tweets (ET). Our approach detects events by exploring their textual and temporal components. ET does not require any target entity or domain knowledge to be specified; it automatically detects events from a set of tweets. The key components of ET are: (1) an extraction scheme for event representative keywords (2) an efficient storage mechanism to store their appearance pat- terns, and (3) a hierarchical clustering technique based on the common co-occurring features of keywords. Authors presented a scalable and efficient system, called ET, to detect real world events from a set of microblogs/tweets. The key feature of this system is the efficient use of con- tent similarity and appearance similarity among keywords, to cluster the related keywords. We demonstrate the effectiveness of this combination in our experiments. ET does not need any human expertise or knowledge from other sources like Wikipedia, and still provides very accurate results. ET is evaluated on two different datasets from two different domains and it yields great results for both of them in terms of the precision.

C. *Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors*

Authors: Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo

Twitter, a popular micro blogging service, has received much attention recently. An important characteristic of Twitter is its real-time nature. For example, when an earthquake occurs, people make many Twitter posts (tweets) related to the earthquake, which enables detection of earthquake occurrence promptly, simply by observing the tweets. As described in this paper, we investigate the real-time interaction of events such as earthquakes, in Twitter, and propose an algorithm to monitor tweets and to detect a target event. To detect a target event, we devise a classifier of tweets based on features such as the keywords in a tweet, the number of words, and their context. Subsequently, we produce a probabilistic spatiotemporal model for the target event that can find the center and the trajectory of the event location. We consider each Twitter user as a sensor and apply Kalman filtering and particle filtering, which are widely used for location estimation in ubiquitous/pervasive computing. The particle filter works better than other compared methods in estimating the centers of earthquakes and the trajectories of typhoons. As an application, we construct an earthquake reporting system in Japan. Because of the numerous earthquakes and the large number of Twitter users throughout the country, we can detect an earthquake by monitoring tweets with high probability (96% of earthquakes of Japan Meteorological Agency (JMA) seismic intensity scale 3 or more are detected). Our system detects earthquakes promptly and sends e-mails to registered users. Notification is delivered

much faster than the announcements that are broadcast by the JMA.

D. Text Detection and Recognition on Traffic Panels From Street-Level Imagery Using Visual Appearance

Authors: Álvaro González, Luis M. Bergasa.

Traffic sign detection and recognition has been thoroughly studied for a long time. However, traffic panel detection and recognition still remains a challenge in computer vision due to its different types and the huge variability of the information depicted in them. This paper presents a method to detect traffic panels in streetlevel images and to recognize the information contained on them, as an application to intelligent transportation systems (ITS). The main purpose can be to make an automatic inventory of the traffic panels located in a road to support road maintenance and to assist drivers. Our proposal extracts local descriptors at some interest key points after applying blue and white color segmentation. Then, images are represented as a "bag of visual words" and classified using Naïve Bayes or support vector machines. This visual appearance categorization method is a new approach for traffic panel detection in the state of the art. Finally, our own text detection and recognition method is applied on those images where a traffic panel has been detected, in order to automatically read and save the information depicted in the panels. We propose a language model partly based on a dynamic dictionary for a limited geographical area using a reverse geo coding service. Experimental results on real images from Google Street View prove the efficiency of the proposed method and give way to using street-level images for different applications on ITS.

IV. CONCLUSION AND FUTURE WORK

In this paper, The existing system detects the traffic related events by analyzing tweets in real-time. This system is basically based on italian tweets, i.e., it classifies only italian tweets. Tweet analysis in italian language is much more easier than english tweets. System classifies the tweets using various machine language techniques like multi class SVM, decision tree and naive bayes classifier. Multi class decision tree provides higher accuracy, precision and recall over all other classifiers; classification based on SVM gave 55.26% accuracy. By using multi class SVM, it is impossible to achieve such higher accuracy levels in english tweets. The proposed system will improve the accuracy of traffic event detection by combining machine language technique multi-class SVM and decision tree classifier. Minimum misclassification techniques used in proposed system. This methods will work well with english tweets. By using these method, we can achieve upto 60.71% accuracy over all other machine language techniques.

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] G. Anastasi *et al.*, "Urban and social sensing for sustainable mobility in smart cities," in *Proc. IFIP/IEEE Int. Conf. Sustainable Internet ICT Sustainability*, 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.

Ms NathaPrasanthi is a student of NIMRA College of Engineering and Technology, IBRAHIMPATNAM,VIJAYAWADA. She is presently pursuing her M.Tech degree from JNTU,Kakinada.

Ms.CH.Lavanya Susana is presently working as Assistant professor in CSE department, NIMRA College of Engineering and Technology, IBRAHIMPATNAM, Vijayawada**.**

SAYEED YASIN received his MTECH in Computer Science & Engg from JNTU Hyderabad. He is pursuing Ph.D., in Rayalaseema University, Kurnool. He is currently working as an Associate Professor & Head in Nimra College of Science & Technology the Department of Computers Science and Engineering & Technology, Jupudi, Ibrahimpatnam,Vijayawada-521456. He has more than Eight years of experience in teaching field. His area of interests are wireless networks & programming, & Mobile Computing.