

Live Detection of Traffic from Twitter

Dr.S. S. Lomte¹, Ms. Sonal N. Gamey²

Principal, Everest College of Engineering & Technology, Aurangabad, India.¹

Student, Computer Science & Engineering, Everest Education Society, Aurangabad, India²

Abstract— online social network is the platforms that users can make the relationships and share interests with others person. Popular social networking sites include MySpace, Facebook, Twitter and Google+, etc. In current era, social networking sites play vital roles in people's life. In 2005, MySpace attracts more page views than Google. In 2009, Facebook overtook MySpace and became the largest social network site[1] Twitter is among the spreading up micro blogging and online social networking services. Twitter was created in March 2006 by Jack Dorsey. Dorsey published the first Twitter message at 9:50 PM Pacific Standard Time (PST) Messages posted on Twitter (tweets) having everything from day to day life stories to the latest local and global news and events. In June 2012.1 over 400 million tweets per day with more than 140 million users Twitter enables users to post status updates, or *tweets*, no longer than 140 characters message to *followers* using various communication services (e.g., cell phones-mails, Web interfaces, or other third-party applications. We also able to focus if traffic is caused by an external event or not, Proposed system is an intelligent system based on text mining and Natural language processing algorithm, for real-time detection of traffic events from Twitter stream and traffic & gives that real time information on Television & FM as the news bulletin[4].we also focus on exact location of user, not the region In the previous paper, we took only Italian language tweets. However, the proposed system can be developed for UK US languages

Keywords:- Traffic event detection, tweet classification, text mining, PST, FM

I INTRODUCTION

Social media sites such as Twitter and Facebook have most popular sites for people to share their opinions on various topics. Twitter is currently the most popular fastest-growing micro blogging service .The large amount of data provided by these media like twitter, is highly important for both large or small scale events. In this paper, we build an efficient system to detect events from tweets (ET). There are many applications, that use twitter for ex. earthquakes In Japan 2011 earthquake, that struck was identified Within

seconds because of the spikes in tweets From affected area, compared to the typical 2 to20 minutes taken by scientific alerts. Another category of events detected by Twitter is sports. During the FIFA World Cup in 2010, Twitter observed huge traffic when a big goal was scored. Instead of 750 tweets per second an average day, there were 2; 940 tweets per second. It is the eighth most popular site in the world.

The main function of Twitter is to allows the users to post and read short messages. And various kind of information, thoughts, opinions, and ideas to update and inform people current affair and what happening around the world. However, Message posted on the Twitter is the status update message via tweets the message length is not longer than 140 characters, Twitter is also one of the unique online social networking service that allows people to create profiles, communicate, and connect with other people on the service. A user can follow any other user without requiring an approval or a reciprocal connection from the followed users. Twitter does not restrict any limits on the number of followers to a user account In this, one user account can typically follow up to 2000 users [2].The user message shared in social networks is called Status Update Message (SUM), and it may contain, apart from the text, meta-information such as timestamp, geographic coordinates (latitude and longitude), name of the user, links to other resources, hash tags, and mentions. SUM contains huge amount of valuable information about an event or a topic .SUMs as sensor information, as it happens with traditional sensors. Now a days social networks and media platforms has widely used as a source of information for the detection of events, around them such as traffic congestion, incidents, natural disasters (earthquakes, storms, fires, etc.), or other events.[3] 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 [2]. In fact, In the Twitter SUMs are unstructured and irregular texts, that contains informal or abbreviated words, misspellings or grammatical errors According to Pear Analytics,[1] with no useful information for the audience.[3]

That information not to be filter out , For this reason, to analyze the information coming from social networks like Twitter, we apply the text mining techniques, It is the field of data mining, machine learning, statistics, and Natural Language Processing (NLP) to extract meaningful information from the SUM. But the most challenging problem is storing and

processing of large amount of unstructured data, handling high velocity data streams, cleaning noise and abnormal data, and finding expected data or meaning full results. Current tools and technologies are not capable to store, process and analyze large amount of data. In this paper, we proposed an architecture that capable and effective storage and analysis of unstructured data, and perform the semantic analysis on that data. By using the Natural language processing (NLP).The NLP evaluation is conducted through streaming Stem filtering, Feature representation, System Features on Twitter data, and performing sentiment analysis. [7].

Text mining process involves structuring the input text and the removal of others unusable text, and simultaneous insertion into a database) similarly recognized patterns for the structured data, and finally gives the validation to output..During the text mining process, various operations can be performed, depending on the particular goal, such as: i) semantic analysis through use of NLP technique, ii) text filtering by means of specific keywords, iii) feature extraction, i.e., conversion of textual features (e.g., words) in numeric features (e.g., weights), that a machine learning algorithm is able to process, and iv) feature selection, i.e., reduction of the number of features in order to take into account only the most relevant ones. The support vector machines (SVMs), decision trees, neural networks, etc. are applied to the documents in the vector space representation for classification, clustering or regression models, and for exact location, by calculating the latitude and longitude, to obtained the final model. Twitter is now a days the most popular micro-blogging service; In 2016 the twitter tweet counts is more than 600 million active users, sharing more than 400 million SUMs per day [3] In the proposed system is based on text mining and machine learning algorithms, for real-time detection of traffic events from Twitter stream and traffic & gives that real time information on Television ,FM as a news bulletin, because of this traffic is minimizes and control the crowd on the road.

II LITERATURE SURVEY

This paper describe the design and development of system by using Stanford Core NLP is Stanford Core NLP provides a set of natural language analysis tools. It can give the base forms of words, their parts of speech, whether they are names of companies, people, etc., normalize dates, times, and numeric quantities, and mark up the structure of sentences in terms of phrases and word dependencies, indicate which noun phrases refer to the same entities, indicate sentiment, extract open-class relations between mentions, etc. one of the most used, & identify the attributes that support to its outcomes[5].In the previous system likes Sakaki and others [8] that uses social sensor to sensing real

time messages from Twitter & the sensors gave earthquake related events. Another system was Asur and uberman[9] use to predict real-world outcomes, from twitter, as focuses on the prediction of box-office revenues for out-coming movies. [10] Tumasjan focusing on the predictions of Elections. They focus on number of tweets reflect voter preferences and comes close to election polls, and the tweets are not only about spreading political opinions, but also to discuss these opinions with other users. Agarwalet al. [7] focus on the detection of fires in a factory from Twitter stream analysis, by using standard NLP techniques and a Naive Bayes (NB) classifier, to detect and analyze small-scale incidents, such as fires. Li et al. [3] propose a system, called TEDAS, to retrieve incident-related tweets. The system focuses on Crime and Disaster-related Events (CDE) such as shootings, thunder storms, 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, hash tags, links, and mentions. Sakakiet al. [7] fetch, elaborate, and classify SUMs as related to a road traffic event or not .

III SYSTEM ARCHITECTURE

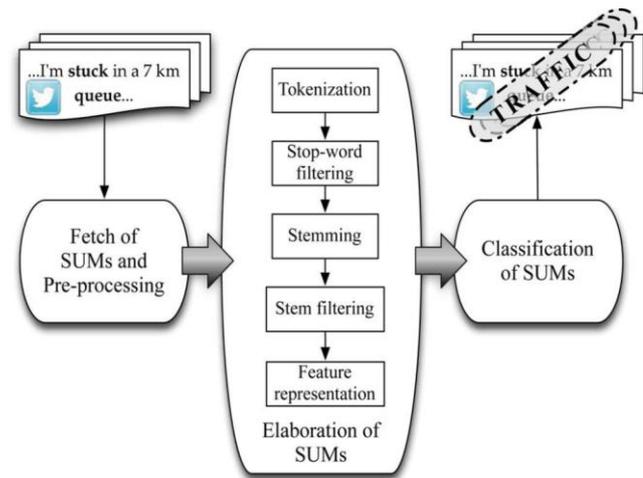


Figure 1 System architecture for traffic detection from Twitter stream analysis.

The System architecture for traffic detection from Twitter stream analysis is composed of three main modules, namely: i) “Fetch of SUMs and Pre-processing”, ii) “Elaboration of SUMs”, iii) “Classification of SUMs”& this is service-oriented and event-driven, The aim of the purpose system is to fetch SUMs from Twitter, process that SUMs by applying text mining steps, and apply the appropriate class label to each SUM. Finally, as shown in Fig. 1.The first module

1. *Fetch of SUMs and Pre-processing*: Extracts raw tweet from the Twitter stream row tweet means , unstructured text, user id, timestamp, the geographic coordinates. Raw tweet remove all meta-information attached with it. In the

previous paper, we took only Italian language tweets. However, the proposed system can be developed for UK US languages.

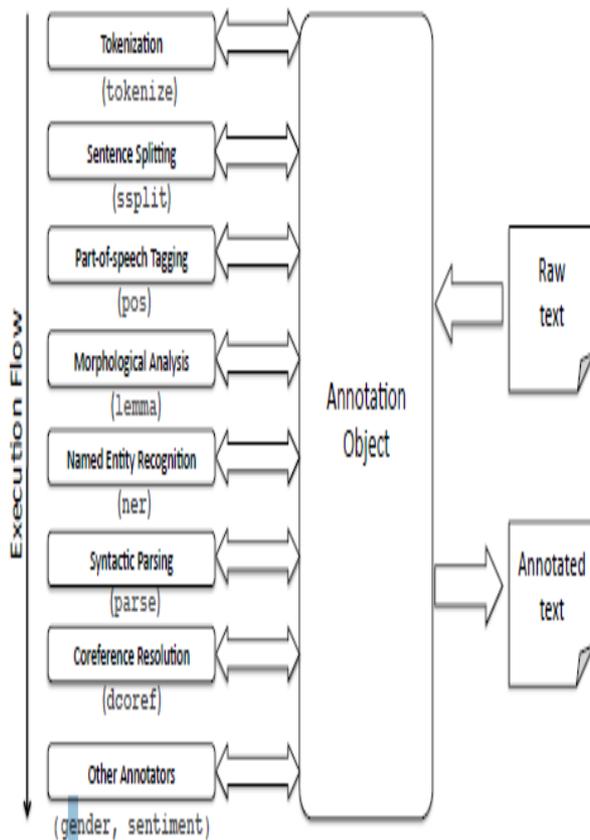
2. *Elaboration of Sums:*

The elaboration of sums contains various blocks that are as:

- a) *Stemming:* It removing suffix from tokens and converting them into their root form.
- b) *Stem filtering:* In this, stems which do not belong to the relevant stems that are removed in the stem filtering.
- c) *Feature representation:* In this, after processing NLP technique, identifying the traffic related words from the relevant set of stems.
- d) *System Features:* It provides notification to users about the location and cause of the traffic. The system analyzes traffic dynamically and suggests the route to avoid traffic congestion. It provides reliable information in a very short time often before online news websites

3. *Classification of SUMs:* The output of this module is a labelling to each SUM, model have been recognized during the supervised learning stage [6].

Figure 2: Overall system architecture: Raw text is put into an Annotation object and then a sequence of Annotators add information in an analysis pipeline.



IV RESULT AND DISCUSSION

Sentiment Score = Sum of Sentiment Scores of words in a tweet Score 1 for positive words is considered be score - 1 negative words is considered and score of 0 is considered to be neutral words. If Score is greater than 0, this means that the sentence has an overall 'positive opinion'. If Score less than 0, this means that the sentence has an overall 'negative opinion'. And if Score is equal to 0, then the sentence is considered to be a 'neutral opinion' [8].

Advantages:

The system solves the problem of traffic in profitable way It saves time by suggesting alternative routes. The system not only detects the traffic but also provides cause related to the traffic also suggested the alternative route . It is flexible and user friendly as it can be used anywhere or any devices that may be cell phones or PCs.

Application:

This system can be used for tourism, Transportation, marketing and Advertisement .In the Transportation system its suggesting the suitable routes mean avoid the crowds, traffic jam or other problems that are related to the traffic and government agencies can take decision of transportation.

Limitations:

The main limitation of the system is the internet connection that should be required throughout the process.

V CONCLUSION AND FUTURE SCOPE

In this paper, we have proposed a system for dynamically analyzes the traffic by using twitter. The system is able to fetch and classify the tweets and notify users about the presence of traffic events and recommend alternative routes. The system can also provide the cause of the traffic efficiently. This system "is for mobile application is limited to android devices but in future it can be developed for iOS , Blackberry, Windows phone OS. The results of this work can be applied by enterprises in sentiment analysis to understand how their customers feel about a particular product or service and to track how those opinions change over time, and also to get information regarding the relative performances of their competitors.[7].This provide Java with ,Stanford Core



NLP attempt to do everything, Stanford Core NLP is used NLP toolkits This system uses the pipeline architecture. Stanford Core NLP [5]

REFERENCES

- [1] Detection of event from the twitter from twitter by Siqu Zhao.
- [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] Real-Time Detection of Traffic From Twitter Stream Analysis EleonoraD'Andrea, PietroDucange, Beatrice Lazzarini, Member, IEEE, and Francesco Marcelloni, Member,.
- [4] Sonal Gamey and Prof. SaadSiddiqui Everest educatio society Aurangabad.
- [5] The Stanford CoreNLP Natural Language Processing Toolkit
- [6] Christopher D. Manning Linguistics & Computer Science Stanford University manning@stanford.edu Mihai Surdeanu SISTA University of Arizona .
- [7] The Movement in Real-Time Analysis Assistant Professor, ChristuJyothi Institute of Technology. Ch. Prudvini, Assistant Professor, ISSN No: 2348-4845 International Journal & Magazine of Engineering, Technology, Management and Research.
- [8] American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) ISSN (Print) 2313-4410, ISSN (Online) 2313-4402 © Global Society of Scientific Research and Researchers.
- [9] Sanchez, G. Mining Twitter with R. Basic Sentiment Analysis in R 2012 [cited 2015; Available from: <https://sites.google.com/site/miningtwitter/questions/sentiment/analysis>]