Real Time Traffic Detection using Twitter Tweets Analysis

Sandeep G Panchal[#1] , Prof. R. S. Apare[*2]

[#]*Dept. Information Technology&Smt. Kashibai Navale College of
Engineering Pune&Savitribai Phule Pune University*

Abstract:*The Social sites have a huge amount of information. The social sites are as Twitter, Facebook, google + and WhatsApp. The social sites used for communication purpose. Using it the user can share an idea, thoughts, emotion, feeling, suggestions, and personal events. Twitter is best because of minimum word it expresses knowledgeable information. It follows news reporter, political leader, movies stars, and businessman. Traffic is a major issue in many cities. Social media is an active site which has many followers, using the traffic related tweets try to control traffic. To implement the real-time traffic detection and analysis of the Twitter tweets coming from those areas in the city. Android application to show and suggest graphical route format of the traffic area. Using text mining technique and natural language programming. Classify traffic related tweets, apply tokenization, stop word filtering, steaming and steam filtering. And also calculate traffic relates tweets coming from which area in latitude and longitude format. The system is real-time because the user travels from one place to another place finding a route on graphical user interface map and select route.*
If the system detects traffic in the route then show the traffic and suggest another alternate route for reach the destination.

Keywords: *NLP, Text Mining, Traffic, Twitter and Tweets.*

I.   INTRODUCTION

The continuous global urbanization raises several challenges to city authority regarding of traffic. As per recent survey 66% of world population is living in cities [2]. All citiesare becoming the smart city. In smart city every people have private vehicle with smartphone [8].That's why traffic on the roads are growing high. Social sites are the new type of real-time information channel. Smartphone is good medium of social sites which as Google+,WhatsApp, Facebook and Twitter. Using that of site user update and share his own thoughts, emotion, feeling and event. Advantage is that have some knowledgeable information. Twitter is most popular in word because of all the news reporter, movie stars, business man and political leader are connected.

Twitter message contains 140 character. Status update message also called tweets, user message shares on social site. Every Status update message added metadata which are name of the user, hashtag, mentions, timestamp, geographical location coordinates in form latitude and longitude.

The main goal is finding route between source and destination. The social media is good information channel for detection of event. Such as incident, traffic congestion, natural disasters or other events. An event can define real word occurrence that happen in a specific time and space. The people can share current traffic situation around them. That message is helpful for other one who has select this route, then he uses alternate route.Social media analysis is challenging problem then event detection from a traditional media like email, blogs, etc. where text is well formatted. Status update message are unstructured and irregular text. That contain informal or abbreviated words misspellings or grammatical error. Text mining is refers to the process of extraction of knowledge and meaningful information from unstructured text. Vagueness of natural language causes the main difficulty encountered in dealing problem with text mining.

II.   RELATED WORK

Eleonora D'Andrea et al. The traffic issue is pointed in the city. The system is using tweets from the twitter and classify them into traffic congestion and car accidents. For the classification the apply natural language programming. The main goal is classify the tweets from twitter, this system is real-time traffic detection. The architecture is divided into three module. The first module is fetch of status update message and preprocessing, in that extract, the tweet from the twitter stream based on one or more search criteria example geographic coordinates, keyword related to appearing in the text of tweet[1].

Second model is an elaboration of the status update message. In that, text mining technique uses some stapes first step is tokenization in that transforming stream character to stream processing unit. The second step is Stop word filtering. The third step is stemming, the main goal this step is the group the word with same related sematic. The fourth step is steam filtering. The last step is feature representation[9].

Yuchao Zhou et al. The system is city event detection at real time from twitter data streams. The event as like traffic, culture, sports, air quality, weather and disaster. The twitter user updates his status message related to his personal information opinion. Every time 500 million people or more people are online with twitter [10]. This user updates his status personal message related to the event,

information, and opinion. Sort out this message with a particular city. To design the general solution and avoid the need of creating keyword set of each and every city, unsupervised method is based on twitter –LDA (Twitter Latent Dirichlet Allocation) is proposed[2].

Table 1: Classification of expected real world events [2]

Category	Traffic influence	People involved	Examples
Traffic	High	Many	Fast traffic ,road network
Culture	High	Potentially many	Sports match, race tournament
Air quality			Description of air pollution incidents
Disaster		Many	Event that causes a huge damage
Weather			Any weather description , includes wind, precipitation, temperature, cloud, sun etc.

Fabio C. Albuquerque et al. The system is based on traffic event detection and twitter messages interpretation. This is one of the objects which is moving from one place to another. Monitor the current state of an object. Detect environment changes that may affect the future and adjust the planned behavior of moving object. News agencies and Government agencies are using the Twitter medium to distribute real-time traffic conditions and notify drivers about planned changes on the road and about a future incident that may affect traffic conditions. Hence, such tweets provide real-time information about the road map, which is exactly the kind of information that provides truck fleet monitoring and similar applications require. TEDO is traffic event domain ontology, the model traffic-related event such as interdiction, accident, breakdown, traffic situation, weather condition and other Event[3].

Dwayne Henclewood et al. Every year United State invests one hundred twenty-one billion Dollars on road Traffic congestion. The goal is a study of predefine system and analysis these disadvantages solve them and update it. Means that after updated version good perform or not. If not goes loop on. There are many traffic simulator and road sensor working in the USA. For this USA government was appointed a committee. That can manage the traffic[4].

Sakkachin Wongcharoen et al. The system is a simulator that store traffic related tweet from the twitter and analysis them for future purpose. First of all collect tweets from twitter, actual road traffic severity, and road data. The second step is training data preparation in that extract tweets and label training data comparing with historical data. The third step is decision tree learning then draw the decision tree. The last step changes the value in database and map also[5].

Tomi Juntunen et al. The system is lightweight web tool that performs traffic related problem. The big city has road complex spider network. One destination has many paths but which one is the shortest path and less time require for that purpose, web tool is developed. The web tool is google map. The overall structure of the tool in that one is a client which means the user interface build using common web technology which is HTML, JavaScript (AJAX) and PHP[6]. Server side uses PHP API is the layer between the user interface and MySQL database. It receives request from web client application.

Supon Klaithin et al. The system is based on twitter data with data mining. The data mining technique[7].

- Data Retrieval: In that collection tweets from twitter.
- Data Cleansing: In that natural language programming used. Detected and removed inconsistency data.
- Normalization: In a twitter network, the user uses to write many shortcut. Normalize the word. Example "way" means "Highway".
- Information Extraction: The main aim of information extraction is to find traffic related information road name, location, traffic accident.
- ROA as Road, DIR as Direction, LOC as location, PST as Position and ACC as Accident.
- Classification: Tweet classification based on Naive Bayes Model.

III.   PROPOSED SYSTEM

The proposed system divided in two application first applicationis web service and second application is android mobile application. In web service they have HDFS (Hadoop dynamic file subsystem) database. Hadoop is used for fast processing and storing high amount of data in to HDFS.

Web service: Web service is taking twitter tweets as input and classify the traffic related tweets from the twitter. For the classification NLP algorithm is used for classification. The main aim is detection traffic related event from social network. It act as multi-class classification which is recognize traffic, non-traffic due to the crash or congestion and traffic due to the external events. The system detect traffic event in real-time. The system is developed an event driven infrastructure.

Android App: It is the android application in that user has to search location and get route. That route is send to the web service and check the traffic in that route if yes then give notification and also give alternate route.

A.   Proposed System steps
Module Partitioning
This session include the partitioning of project into different modules. The Five modules of the system are:
Module 1: Taking real time Public tweets from Twitter.
First of all do registration on Twitter website after completion of registration go to Twitter developer option and create application by enter any name or project name.

Then generate Consumer Key and Consumer secret, Access Token key and Access token Secret key. Each key having different meaning with access permission. Consumer key only for access login of user via key directly to twitter.

Consumer Token key for read and write tweets into the application.

Access Token Key and Access Token Secret key for Accessing twitter account and read and write posts.

Using all keys accessing real time tweets from the twitter. As input tweets to the project.

Module 2: Classify traffic related tweets.

In this module classify traffic related tweets from the twitter and identify sentiments of each tweets in positive, negative and natural. Analysis public traffic tweets in to web service. First of all take Tweets, split tweet into tokens as called Tokenize, remove stop words from tweets and after that compare words with traffic related words called Filter and finally get fully traffic tweets define the class of each tweets.

Module 3: Store data into HDFS.

In this project all the traffic related tweets stored into the Hadoop distributed file system (HDFS). HDFS is used in the system because fast storing, retrieving and processing data.

Module 4: Connectivity between Web service and Android application.

In project divided into two application one is web service and Android application. Web service runs on the computer system and android application runs on android mobile. For connectivity between web service and android application using Json parser. Json paser is independent data exchange. Using json object hit the quarry to web service and get the response in to universe format. Using JSONarray, JSONobject, JSONStringer and JSONTokenizer this method get perfect output.

Module 5: Traffic detected in the Route.

In android application user find Enter source and destination finding non route between them in geographical format.

B. System Overview

For the classification of Tweets from the Twitter used Stanford – core NLP.

In the Stanford core NLP working steps as follows
Steps 1: parser 1
In the parser 1 is StanfordcoreNLP POS tagging.
Tweets passes from Stanford core NLP parser 1 for the tagging each tweets divided in to the tokens.
Step 2: POS tags
Each tokens has been tagged by the POS tags.POS tags in different phases such as noun phrase, verb phrase and pronoun etc.
Step 3: Parser 2
In parser 2 works on keywords matching with corpus.
Tagged key words matches to words that's stored in to the traffic class directory
Traffic class directory.
Step 4: parser 3
In that removing noise from the tweets.
Other than the traffic words removed from the tweets and traffic class word present.
Present word length less than one then is sentiments is negative traffic tweets.
Present word length greater than one then is sentiments is positive traffic tweets.
Present word length is equals to one then is sentiments is neutral traffic tweets.
Finally get the traffic related tweets with their sentiments status.
Stanford coreNLP every time compare all the traffic class keyword.Proposed system classify the traffic tweets in three form positive, negative and neutral. Positive tweets is related to traffic directly, negative tweets is not relation between traffic and neutral tweets is or isn't relation between tweets. More detail shows in following table.

Figure 1: Proposed system architecture

Table 2: classification of tweets

Sr. no.	User Name	Tweet text	Status
1	**Aanand** @aan4nd	#Bengaluru #Traffic will be fun to see from the tops of these buses.	Positive
2	**Skynet Technologies** @skynetindia	#Drupal#website drives visitor #traffic, increase #Googleranking& conversion. http://ow.ly/nztv302YW5n #Drupal8#UK	Negative
3	**Sudip** @iam_Sudip	If this is how it #rain, by tomorrow morning half of #Bangalore will be submerged #under water.	Neutral
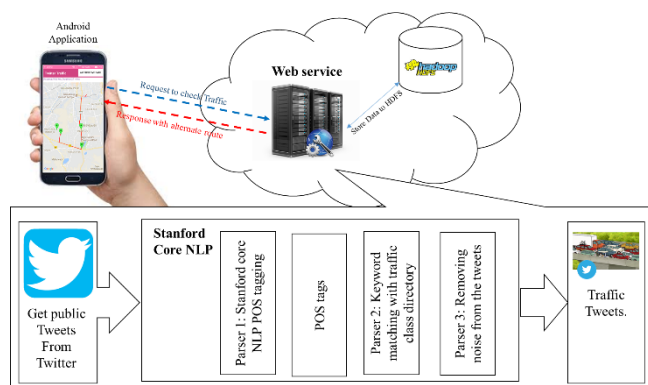
C. Assumptions and Mathematical Formula

The social networking site twitter is now mostly used site for communication so the traffic is also increased for twitter.

Calculate the distance between two pointsto calculate the great-circle distance between two points that is, the shortest distance over the earth's surface. Assuming the value of earth radius as mean radius. The latitude and longitude is also under assumptions.

Formula[11]

$a = \sin^2(\Delta\varphi/2) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2(\Delta\lambda/2)$
$c = 2 \cdot a * \tan2(\ \sqrt{a},\ \sqrt{(1-a)}\ )$
$D = R \cdot c$

Where

φ *is latitude*
λ *is longitude,*
R *is earth's radius (mean radius = 6,371km)*
D *is distance between two points.*

IV. RESULT

The developed system was implemented for the real-time monitoring of several areas of the road network, by means of the analysis of the Twitter stream coming from those areas. The aim is to perform a continuous monitoring of frequently busy roads and highways in order to detect possible traffic events in real-time or even in advance with respect to the traditional news media.

Table 3 Existing vs Proposed system accuracy values

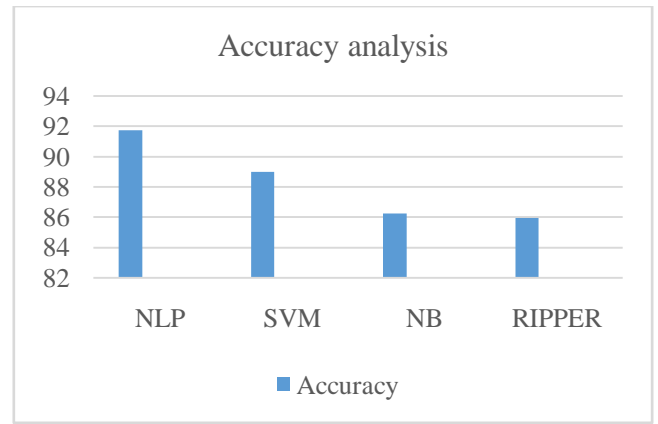| Sr.no.	Classification	Accuracy
1	NLP	91.75
2	SVM	89
3	NB	86.25
4	RIPPER	85.93

Figure 2.1: Accuracy analysis Graphs.

Table 4: Existing vs Proposed system values

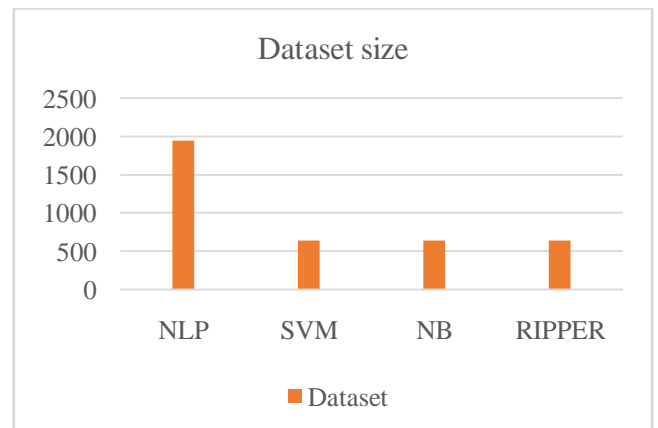| Sr.no.	Classification	Dataset size in tweets
1	NLP	1949
2	SVM	640
3	NB	640
4	RIPPER	640

Figure 2.2: Dataset analysis Graphs

In the above Table 3and 4, row 1 is proposed system output, data set of that is higher than other. And rows 2, 3 and 4 have existing that system output. Figure 2 is the graphical output of incoming tweet from twitter and traffic related tweet. Actual output of the project is Graph.

Following figure shows the graph of total traffic class incoming tweets after application of NLP in them.It only shows traffic related tweets that are to be analyzed.
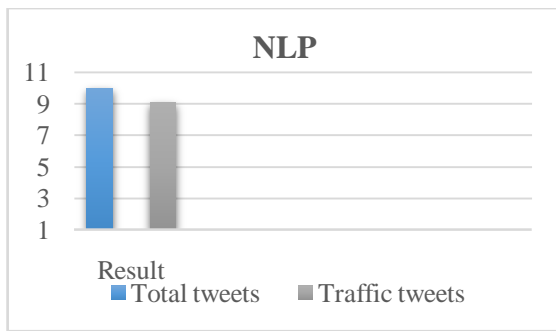
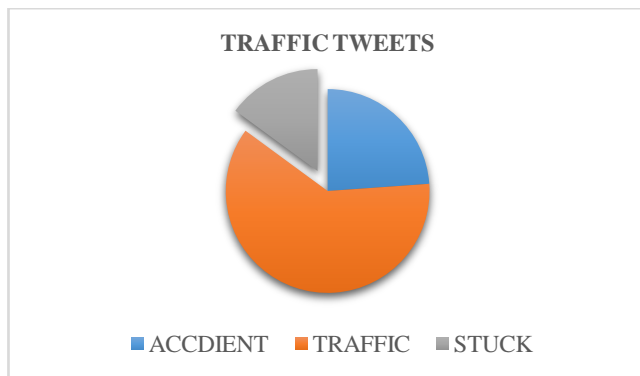Fig 2.3: Bar chart of Input tweets and traffic related tweets.

Fig 2.4: Pie Chart for Traffic classification

Figure 2: Graphical output

Figure 2.2 is shows that traffic tweet with their issues. Detail study of traffic related tweets and which word shows the traffic in the tweets. Classify the words wise traffic tweets.

V. CONCLUSION

I have implementedreal time traffic detection using twitter tweets analysis, and maintained lists like, accident, traffic jams, vehicle breakdown etc. Social network have large amount of information for event detection with particular reference to road traffic congestion and car accident. Using natural language programming and Hadoop framework the system became fast processing and classifies the data with graphical map location.

ACKNOWLEDGMENT

References

[1] Eleonora D'Andrea, Pietro Ducange, Beatrice Lazzerini, and FrancescoMarcelloni ,"*Real-Time Detection of Traffic From Twitter Stream Analysis*" Published By IEEE Journal December 22, 2015

[2] Yuchao Zhou, Suparna De, Klaus Moessner , "*Real world city event extraction from Twitter data streams* " Published By Elsevier Journal 2016

[3] Fabio C. Albuquerque , Marco A. Casanova , Helio Lopes , Luciana R. Redlich , Jose´ Antonio F. de Macedo , Melissa Lemos , Marcelo Tilio M. de Carvalho , Chiara Renso , "*A methodology for traffic-related Twitter messages interpretation* " Published By Elsevier Journal 2015

[4] Dwayne Henclewood, Wonho Suh, Angshuman Guin, Randall Guensler, Richard M. Fujimoto, Michael P. Hunter ," *Real-time data-driven traffic simulation for performance measure estimation* " Published By IEEE Journal 2016

[5] Sakkachin Wongcharoen and Twittie Senivongse, "*Twitter Analysis of Road Traffic Congestion Severity Estimation*" Published By IEEE Journal 2016

[6] Tomi Juntunen, Vassilis Kostakos, Mikko Perttunen, Denzil Ferreira, "*Web Tool for Traffic Engineers: Direct Manipulation and Visualization of Vehicular Traffic Using Google Maps*" Published By ACM Journal 2012

[7] Supon Klaithin and Choochart Haruechaiyasak, "*Traffic Information Extraction and Classification from Thai Twitter*" Published By IEEE Journal 2016

[8] G. Anastasi et al., "*Urban and social sensing for sustainable mobility in smart cities*" published by IEEE 2013.

[9] T. Sakaki, M. Okazaki, and Y.Matsuo, "*Tweet analysis for real-time event detection and earthquake reporting system development*" published by IEEE Apr. 2013.

[10] Twitter Usage Statistics [online] Available: http://www.internetlivestats.com/twitter-statistics, October 21, 2015.

[11] https://www.quora.com/How-does-Google-Maps-calculate-the-distance-from-one-place-to-another