Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 4, April -2017

Twitter Stream Analysis for traffic detection in real time

Sayali Dhanawade[1], Rucha Kulkarni[2], Shraddha Raut[3], Prof.D.S.Gogawale[4]

[1]*Department of Computer Engineering, NESGOI, Pune*
[2] *Department of Computer Engineering, NESGOI, Pune*
[3] *Department of Computer Engineering, NESGOI, Pune*
[4] *Department of Computer Engineering, NESGOI, Pune*

Abstract — *Social networking sites have spread in recent year and becoming real time information channel. In this paper we use the Twitter as social networking site for collecting the real time information. In this paper, we use some machine learning technique and text mining algorithms for traffic detection. Firstly, we fetch the tweets then apply the sentiment analysis techniques on that fetched tweets. We use the NLP for extracting the meaningful information. NLP include tokenization, stop word removing stemming. These NLP method apply on that tweets and after that the classify tweets into two types i.e. Traffic or Non Traffic event. If there is traffic then show the alternate path to user.*

Keywords- *Traffic event detection, tweet classification, social sensing, text mining, NLP.*

I. INTRODUCTION

Now a day, many people use the social networking sites and exchange the information. Therefore, social networking has become the large source of information. Twitter is most useful site in daily life for sharing the information. It has more than 600 million active users and they share more than 400 million tweets per day. Therefore, twitter is used for real time event detection. The events are earthquake detection, natural disaster, Traffic detection etc. There are some challenges for event detection using twitter; the most important challenge is size of tweets. The tweets are large in size and it contain large of amount of unwanted data. So that tweets needs some processing to covert small and meaningful form using the text mining.

Traffic is big problem that people faces in daily life.so the twitter is used for traffic detection in real time. In this system, Perform a continuous monitoring of frequently busy roads and highways in order to detect possible traffic events in real-time by analyzing of the Twitter stream coming from those areas.

1.1. PURPOSE

The main purpose of the system is get Public traffic tweets from twitter. for real-time detection of traffic-related events from Twitter stream analysis. 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. Web portal gets array of latitude and longitude and sends return to traffic between that array with causes. Alternate path displayed with traffic.

1.2. SCOPE

This system is generally based on get Public traffic tweets from twitter and Apply tokenization, remove stop words and apply stemming to a particular tweet. Our traffic detection system based on Twitter streams analysis is presented. And it detects the traffic events in real-time. Haversine method is used to calculate the distance between two latitude-longitude pairs, Triangulation for getting GPS Location. After comparing the longitude and latitude having traffic, it is displayed on the maps of Android device. The system is use to Twitter as data source for fetching the all post regarding the road traffic and Accidents.

II. OBJECTIVE

The objective of the system is to detect the small scale event. The events like road. This system is for real-time detection of traffic-related events from Twitter stream analysis. The system is also able to discriminate if a traffic event is due to an external cause. This system is based on text mining and machine learning algorithms.

III. RELATED WORK

Social media is used as source of information. The information extraction is used for event detection. There are two types of events that are small scale events and large scale events. The smalls scale events are traffic, car crashes, fires, etc. And the large scale events are earthquakes, tornados, etc. Smaller numbers of SUMs are used for small scale vent detection.

Large scale events are detected using the Twitter stream by monitoring special keywords and apply the SVM as binary classifier to classify the positive and negative events. For small scale event the TEDAS system is used to retrieve events from tweets.

In this paper, we detect the small scale event i.e. road traffic. And we detect the traffic event by analyzing the SUM using certain criteria.

IV. PROPOSED SYSTEM ARCHITECTURE

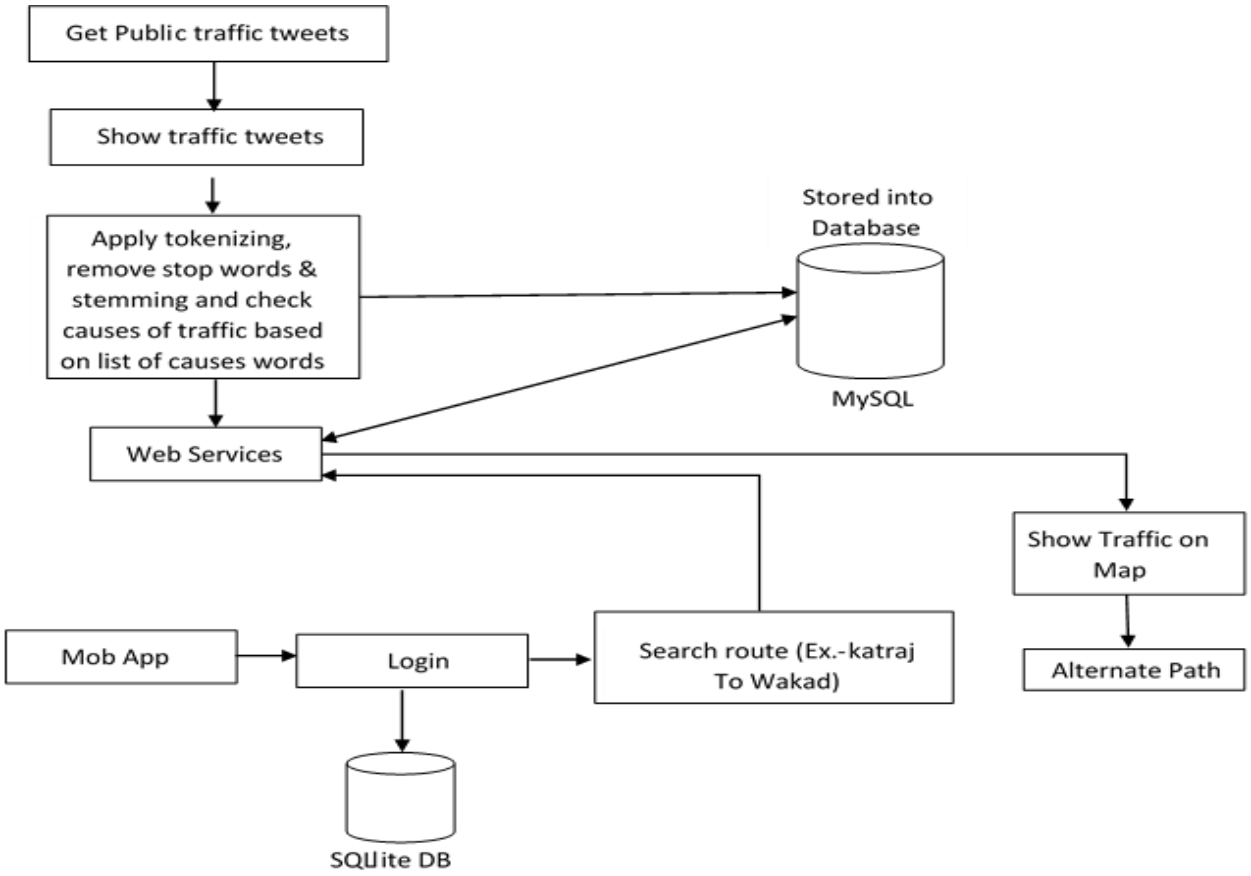

Figure 1: The proposed System Architecture.

Above figure shows the proposed system architecture for real time traffic detection system. Numbers of twitter users are tweets regarding to occurrence of traffic and also gives the cause of traffic. The system gets the public traffic tweets according to several searching criteria. and shows the fetched public traffic tweets on web portal. On that tweets the NLP techniques are apply .Three types of methods are used in NLP for event detection.

4.1.Tokenization

Tokenization is the first step of Text mining. In that the stream of characters are transforms into the stream of processing units called as tokens. The tokenizer removes all punctuation marks and then splits the each Status Update Massage into tokens.

4.2. Stop words removing

In stop word removing, the stop words are eliminate. The stop words includes the words which provide the little or no information. Articles, preposition, conjunctions, pronouns, etc. are the common stop words.

4.3. Stemming

The process of reducing the word i.e. token is called as stemming. In that, words are reduces in the stem or root form. It removes the suffix. Each Status Update message is represented as sequence of stems. That stems are extracted from tokens.

After applying the above methods, classify the tweets into two classes that are Traffic and Non-Traffic. And if there is traffic then it checks the cause of traffic and stores into the database.

There is one mobile application on client side. User can login from mobile application. Through the mobile App user search for the route using source and destination location. Then this query go to web service. web services gets the array of latitude and longitude of searched path and checks into database for traffic. then it compare the latitude and longitude of searched path with latitude and longitude of traffic .If there is traffic then it gives the alternate path to the user.

V.ALGORITHM

Step 1: user search for route.
Step 2: Get location of user.
Step 3: decide route to destination
    Step 3.1: Get user location co-ordinates
    Step 3.2 : Get destination location and determine route
    Step3.3 : Get traffic status alongside route
    Step 3.4 : check conflicts if any occurs
    Step 3.5 : decide alternate route if needed
Step 4: Show alternate route
    Step 4.1 : Get congested locations
    Step 4.2 : Decide alternate paths
    Step 4.3: Check traffic on alternate route
Step5: Display Result.
Step 6: Stop.

VI. MATHEMATICAL MODEL

S= { s, X, Y, T, f_{main}, DD, NDD, f_{friend}, memory shared, CPU_{count}}
Where,
S (system):- Is our proposed system.
s (initial state at time T) :-GUI of Traffic Detection From Twitter Analysis.
X (input to system) :- Input Query.
Y (output of system) :- List of URLs with Snippets..
T (No. of steps to be performed) :- **6**. These are the total number of steps required to process a query and generates results.
f_{main}(main algorithm) :- It contains Process P. Process P contains Input ,Output and subordinates functions.
DD (deterministic data):- It contains Database data.
NDD (non-deterministic data):- No. of input queries.
f_{friend} :- WC And IE.
Memory shared: - Database.
CPU_{count}: - 2.

Subordinate functions:
Identify the processes as P.
S= {I, O, P....}
P= {WC, IE}
Where, WC= Web Crawler for Tweeter service.
IE= Information Extraction that filter tweets
P = processes.
WC= {U, MAX, CP}
Where, U=input Query
MAX = {1, 2, 3, … , n}
CP= output i.e. traffic information on a route
IE= {CP, NLP Techniques, Info}
Where,
CP= input which is tweets
NLP= process for filtering tweets related to traffic, tokenization and other processing of data.

VII. ADVANTAGES

- Easily traffic will be detected, Alternative path will be displayed
- It completes a multi-class classification
- It senses the traffic events in real-time
- It is developed as an event-driven structure, built on an SOA architecture

VIII. APPLICATION

- Accident detection system
- Traffic jam analysis
- Vehicle breakdown
- Intelligent Transportation System

IX. IMPLEMENTATION AND RESULT SET

In this system, we are performing some experiments. This experiments are performed using two dataset that are 2 Class dataset and 3 Class dataset. We fetches 100 number of at each request and perform operation on that tweets. The operations are preprocessing and classification. Preprocessing includes the text mining techniques. After that classify the SUM according to several criteria into 2 Class dataset and 3 Class dataset related to traffic events. The 2 Class dataset includes traffic and non-traffic class and 2 Class dataset includes traffic ,non-traffic and traffic due to external cause. Below figure show some example of classification of tweets.

id	TId	latitude	longitude	tweet	tweetword	username
644	851651476300718080	31.75961	-106.4737	Closed in #ElPaso on Montana Ave EB between ...	traffic	TTN El Paso
645	851651397938323457	18.2682262	73.8842501	#traffic heavy traffic in Balaji nagar due to fire	traffic	sayali dhanawade
646	851650427007455232	31.75961	-106.4737	Closed in #ElPaso on Montana Ave EB between ...	traffic	TTN El Paso
647	851649059236589569	53.2167	6.55	Allmost at home, time for breakfast! Returntrip ...	traffic	Lambert Kamps
648	851648582306455553	44.95764	-93.26926	Accident in #Hennepin on 35W NB at 26th St, s...	accident	TTN Minneapolis
649	851648321164906497	35.7958	-78.6409	Ramp closed for construction: short detour is in ...	construction	Raleigh Traffic

Figure2:Result Set

After that we login into mobile application and enter the path that we have to search e.g. Katraj to wakad . And then this query goes to the web service and web service checks into the database. If there is traffic then alternate path will be shown to user.

The aim is to execute continuous monitoring busy roads and highways. The system is developed and tested for real time monitoring the road network.

X. CONCLUSION

In this system, the real time traffic is detected using the twitter stream analysis. This system is based on SOA. And it fetch and classify the tweets according to several criteria. If there is traffic then the system show alternate path to the user and the system also discriminate the causes of traffic like football match.

REFERENCES

[1] Eleonora DAndrea, Pietro Ducange, Beatrice Lazzerini, Member, IEEE,and Francesco Marcelloni, Member, IEEE, Real-Time Detection of Traffic From Twitter Stream Analysis, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS,VOL. 16, NO. 4, AUGUST 2015.

[2] F. Atefeh andW. Khreich, A survey of techniques for event detection in Twitter,Comput. Intell., vol. 31, no. 1, pp. 132164, 2015.

[3] 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.228238,Feb. 2014.

[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. 919931, Apr. 2013.

[5] M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, Getting there first: Realtime detection of real-world incidents on Twitter in Proc. 2 nd IEEE Work Interactive Vis. Text Anal.Task-Driven Anal. Soc. Media IEEE VisWeek, Seattle,WA, USA, 2012.

[6] R. Li, K. H. Lei, R. Khadiwala, and K. C.-C. Chang, TEDAS: A Twitterbased event detection and analysis system, in Proc. 28Th IEEE ICDE, Washington,DC, USA, 2012, pp. 12731276.

[7] A. Hotho, A. Nrnberger, and G. Paa, A brief survey of text mining, LDV Forum-GLDV J. Comput. Linguistics Lang. Technol.,vol. 20, no. 1, pp. 1962, May 2005.

