A Proposed System for Real Time Detection of Traffic from twitter using NB Classifier

[1]Akshay Chugh, [2]Siddharth Deshmukh, [3]Amol Gunjal, [4]Madhuri Kamble

Department of Computer Engineering,
MIT College of Engineering, Kothrud, Pune, India

***Abstract*— Social networks have been recently employed as a source of information for event detection its real-time nature. Twitter has received much attention recently as the investigation of real time events such earthquakes, traffic congestion, street rallies is the need of the existing busy life of the users. In this paper, a real-time monitoring system for traffic event detection is proposed. The system fetches tweets from Twitter according to several search criteria; processes tweets, by applying text mining techniques; and finally performs the classification of tweets. The aim is to assign the appropriate class label to each tweet, as related to a traffic event or not. After the traffic event is detected, further classification is performed to decide the class of the traffic tweet. In other words, the status of the traffic in terms of heavy traffic, medium traffic, road jams are analyzed from the traffic tweet.**

***IndexTerms*— Location estimation, Traffic congestion, Tweet classification, Text mining, Twitter, Naïve Bayes Classifier, Real-time event detection.**

I. Introduction

Twitter, a popular microblogging service, has received much attention recently due to its real time nature. This online social network is used by millions of people around the world to remain socially connected to their friends, family members, and co-workers through their computers and mobile phones. People intensely use twitter to report (personal or public) real life events happening around them or simply to express their opinion on a given topic, through a public message. People post Tweets, which may contain photos, videos, links and up to 140 characters of text. These messages are posted to your profile, sent to your followers, and are searchable on twitter search[1],[2].

In most of the microblogging services like Facebook, Gmail the user can send or update text, video, audio. In the proposed system, the twitter database is used to detect time and location of real time event like traffic or other event. The real time nature of twitter is useful to get accurate outcome. In the proposed system the tweets from user are collected, and then tweets are classified using NB classifier to detect the target event. Traffic congestion is a huge problem the world is facing nowadays. People suffer immensely in terms of money and time. In this paper, we present a system to dynamically analyze traffic and its causes, using twitter stream analysis[3],[4].

II. LITERATURE SURVEY

Sakaki *et al.* [4] use Twitter streams to detect earthquakes and typhoons, by monitoring special trigger-keywords, and by applying an SVM as a binary classifier of positive events (earthquakes and typhoons) and negative events (non-events or other events). In [5], the authors present a method for detecting real-world events, such as natural disasters, by analyzing Twitter streams and by employing both NLP and term-frequency-based techniques.

Agarwal *et al.* [6] focus on the detection of fires in a factory from Twitter stream analysis, by using standard NLP techniques and a Naive Bayes (NB) classifier. In [7], information extracted from Twitter streams is merged with information from emergency networks to detect and analyze small-scale incidents, such as fires.

Wanichayapong *et al.* [8] extract, using NLP techniques and syntactic analysis, traffic information from microblogs to detect and classify tweets containing place mentions and traffic information.

Li *et al.* [9] propose a system, called TEDAS, to retrieve incident-related tweets. The system focuses on Crime and Disaster-related Events (CDE) such as shootings, thunderstorms, 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, hashtags, links, and mentions.

Sakaki *et al.*[10] extract, based on keywords, real-time driving information by analyzing Twitter's SUMs, and use an SVM classifier to filter "noisy" tweets not related to road traffic events. Schulz *et al.* [11] detect small-scale car incidents from Twitter stream analysis, by employing semantic web technologies, along with NLP and machine learning techniques. They perform the experiments using SVM, NB, and RIPPER classifiers.

III. PROPOSED SYSTEM

The main purpose of the proposed system is to fetch tweets from twitter according to several search criteria which are- processing the tweets; by applying text mining techniques; and finally performing the classification of tweets. Secondly, the proposed system is used to classify the traffic event into various classes such as heavy traffic, moderate traffic, no traffic etc. 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 as an event-driven infrastructure, built on a Service Oriented Architecture (SOA) [12]. 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.

IV. MATHEMATICAL MODEL

Input: Tweets
Output: Proposes real-time detection of traffic-related events from Twitter stream analysis.

Set Theory Analysis:
a) Let 'S' be the | Dynamic survey system with location tracker as the final set
S = {…………}

b) Identify the inputs as C, P and M
Input= {C, P, M, U, I, B, A, L}
C = {C1, C2, C3, …, Cn | 'C' given current inputs}
P= {P1, P2, P3, …, Pn | 'P' gives previous inputs}
M= {M1, M2, M3 | 'M' gives control sync alerts}
U = {U1, U2, U3, …, Un | 'U' as Username}
I = {I1, I2, I3, … In |'I' password}
B = {B1, B2, B3, …Bn | 'B' is entity related information}
A = {A1, A2, A3, …An | 'A' is IMEI number of mobile}
L = {L1, L2, …, Ln | 'L' is the location of entity}

c) Identify the outputs as O
Output= {O, D}
O = {O1, O2, O3, … On | 'O' is alert message sent for control updation}
D = {D1, D2, D3, … Dn | 'D' is the survey information shown on map}

d) Identify the functions as 'F'
F = {Registration (), Authentication (), delete (), Control Updation(), Map View()}
Registration (U, A) =R' :: takes username(U), Password(I).
R' = {d|'d'contains customer information for registration}
Authentication (U,I) = AU' :: Verifies the customer's username and password AU' = {d|'d' contains valid customer's Username and Password}
Delete (B, A) = D':: delete the customers particular account
D' = {d|'d' account details}
Alarm (L, M) = Q':: Generate alert messages for control updations.
Q = {d|'d' generate the messages}

e) Identify the success as Su.
S = {Input, Output, F, Su, ...} where,
Su =Success is when tweets are fetched successfully and users are notified in real time.

f) Identify the failure as F.
S = {Input, Output, Su, F, ...}
where,
F= Failure is when user does not get notified within specified time.

V. SYSTEM ARCHITECTURE

The diagram in Figure 1 shows the architecture diagram of the proposed system where the tweets shared by registered users of twitter will be stored at twitter server. One more server named Traffic Analyser Server will fetch tweets from twitter server and will process the tweets. The tweets detected related to traffic event will be further classified using NB classifier. The notifications related to traffic status in a particular area will be sent to the registered users on the android application.

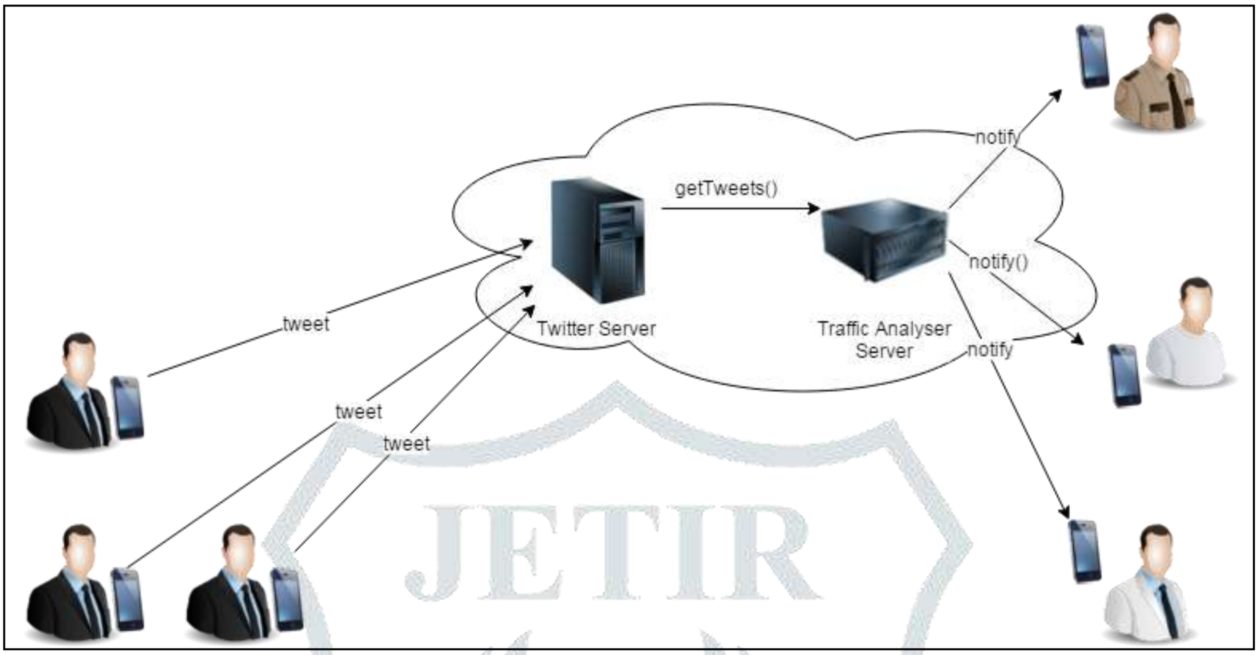

Figure 1. Architecture Diagram of Proposed System

The system architecture is service-oriented and event-driven, and is composed of three main modules, namely: 1) “Fetch of SUMs and Pre-processing”, 2) “Elaboration of SUMs”, 3) “Classification of SUMs”. The proposed system is to fetch SUMs from Twitter, to process SUMs by applying a few text mining steps, and to assign the appropriate class label to each SUM. Figure 2, shows text mining techniques applied to a sample tweet in Elaboration of SUM’s module.

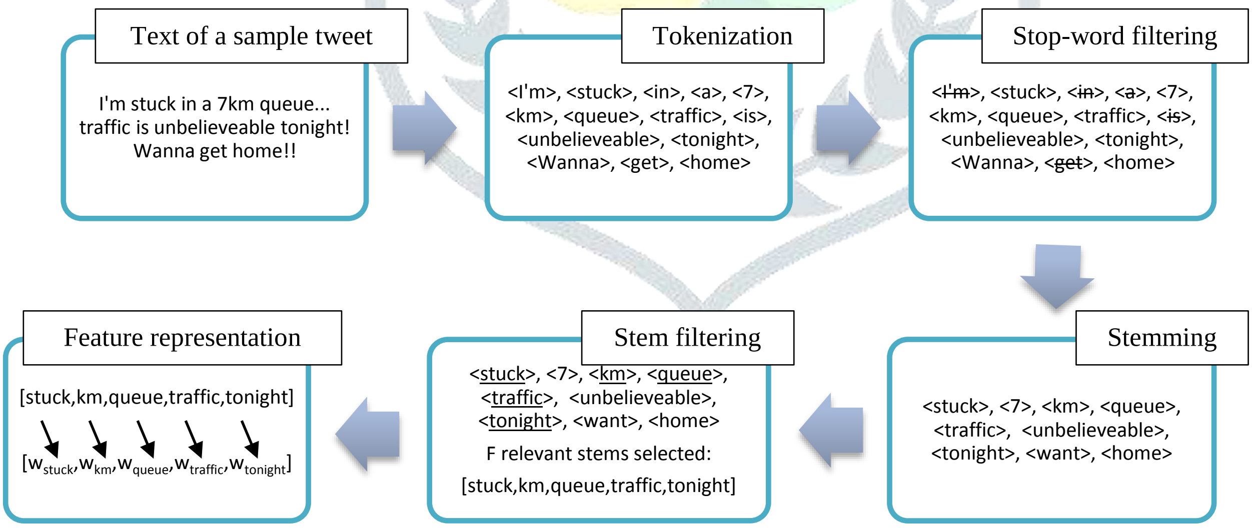

Figure 2. Text mining elaboration applied to a sample tweet (SUM)

The block diagram in Figure 3 shows a detailed overview of the proposed system. The proposed system is divided into two parts: First is the Traffic Notifier application which will be an android application and Second is the Traffic analyser module which will act as a server to the android application and where all the backend processing related to tweets will take place. From Traffic Notifier application the user can login to the application and post a tweet related to traffic. The notifications related to traffic status will be displayed on the Traffic Notifier application. On the server side, tweets will be fetched from twitter dataset. On fetched tweets NB classifier will be performed so as to get the tweets related to traffic event. The shortlisted traffic tweets will be further processed using NLP. The preprocessed tweets will be classified into various classes of traffic and saved into database. The notifications regarding traffic status will be sent to the Traffic Notifier application so that users can view the status.

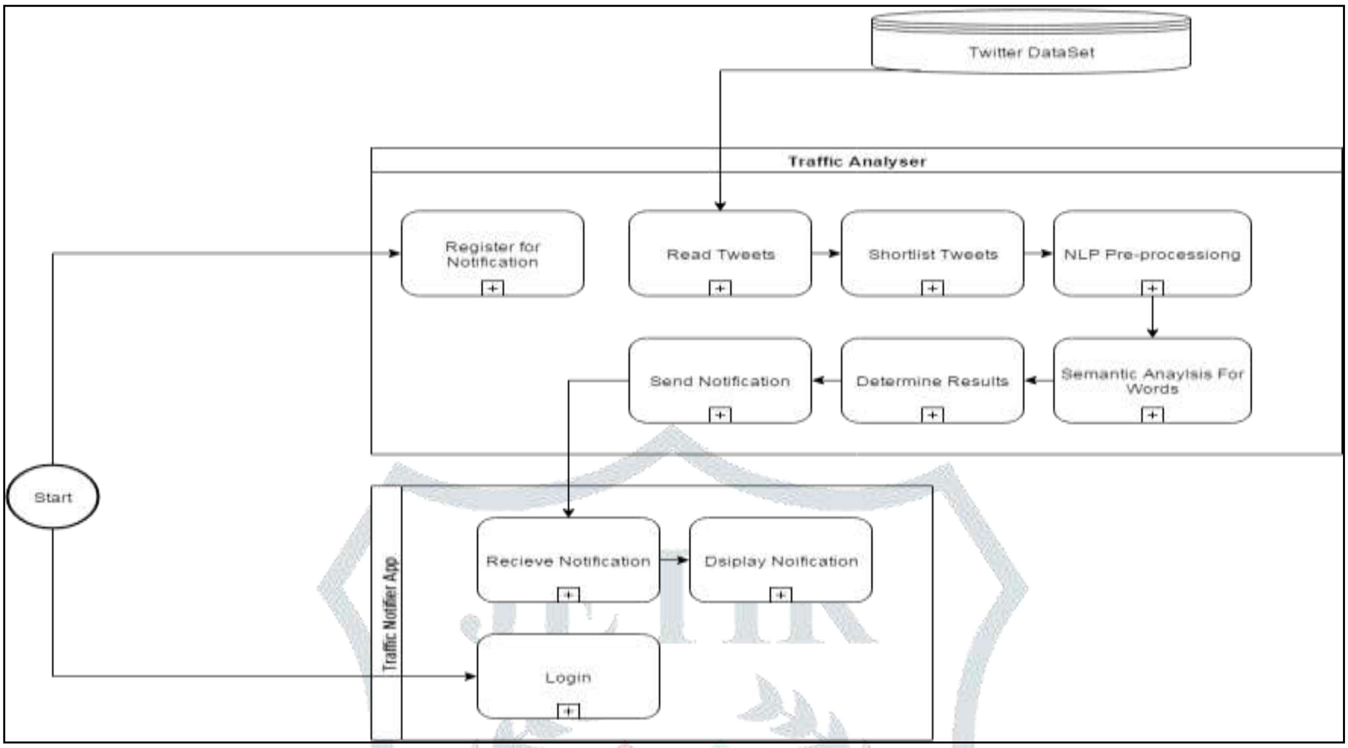

Figure 3. Block Diagram of Proposed System

The proposed system uses two different datasets, i.e., a 2-class dataset, and a 3-class dataset.

2-Class Dataset: The first dataset consists of tweets belonging to two possible classes, namely i) road traffic-related tweets (traffic class), and ii) tweets not related with road traffic (non-traffic class). The tweets were fetched in a time span of about four hours from the same geographic area.

3-Class Dataset: The second dataset consists of tweets belonging to three possible classes. In this case we want to discriminate if traffic is caused by an external event (e.g., a football match, a concert, a flash-mob, a political demonstration, a fire) or not. Even though the current proposed system is not designed to identify the specific event, knowing that the traffic difficulty is caused by an external event could be useful to traffic and city administrations, for regulating traffic and vehicular mobility, or managing scheduled events in the city.

The proposed system makes use of NB Classifier for classifying the traffic tweet into a particular class of traffic.

In general, the Bayesian probability can be represented in following as in Eq. (1):

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}} \quad \ldots(1)$$

Using Bayes' theorem, the conditional probability can be decomposed as in Eq. (2):

$$p(C_k|\mathbf{x}) = \frac{p(C_k)\ p(\mathbf{x}|C_k)}{p(\mathbf{x})} \quad \ldots(2)$$

Where,

$p(C_k)$ is probability of features of a particular category.

$p(\mathbf{x}|C_k)$ is probability of original review divided by features of particular category.

$p(\mathbf{x})$ is probability of review given by the user.

VI. CONCLUSION

Thus in this paper, a system for real-time detection of traffic-related events has been proposed. The proposed system is able to fetch and classify streams of tweets and to notify the users of the presence of traffic events. Further the system discriminates if a traffic event is due to an external cause, such as cricket match, procession and manifestation, or not. The users can get a clear idea about the traffic from the proposed application and can plan the road journey accordingly. This application will provide an ease to the users and will save their time and money.

REFERENCES

[1] F. Atefeh and W. Khreich, “A survey of techniques for event detection in Twitter,” *Comput. Intell.*, vol. 31, 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.

[3] 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.

[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. 919–931, Apr. 2013.

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

[6] P. Agarwal, R. Vaithiyanathan, S. Sharma, and G. Shro, “Catching the long-tail: Extracting local news events from Twitter,” in *Proc. 6th AAAI ICWSM*, Dublin, Ireland, Jun. 2012.

[7] F. Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao, “Twitcident: fighting fire with information from social web streams,” in *Proc. ACM 21st Int. Conf. Comp. WWW*, Lyon, France, 2012.

[8] 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.

[9] 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.

[10] 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.

[11] A. Schulz, P. Ristoski, and H. Paulheim, “I see a car crash: Real-time detection of small scale incidents in microblogs,” in *The Semantic Web: ESWC 2013 Satellite Events*, vol. 7955. Berlin, Germany: Springer-Verlag, 2013.

[12] The Smarty project. [Online]. Available: http://www.smarty.toscana.it/