

A Survey on Traffic Detection from Twitter Tweets Analysis

Sandeep G. Panchal¹, Ravindra S. Apare²

¹M. Engineering Student, Smt. Kashibai Navale College of Engineering Pune, Savitribai Phule Pune University, India

²Assistant Professor, Department of I.T. Smt. Kashibai Navale College of Engineering Pune, Savitribai Phule Pune University, India

Abstract: Internet of thing, the social sites have large amount information source. The social site is like twitter, Facebook, google+ and whats app. The social site is used for communication purpose. The user can update him into it status update message. Twitter is the most popular social site. Which having the source of information related traffic detection. This paper is a survey on traffic detection from twitter tweets analysis. First of all find the traffic related tweets from twitter and then process it using text mining techniques. Finally, classify the tweets and assign that tweet with appropriate class labels, as related to traffic or nontraffic. This system is real-time traffic detection and monitoring system from twitter stream analysis.

Keyword: Twitter, traffic detection, text mining and NLP

1. Introduction

Recent survey 66% of world population is living in cities [2]. The continuous global urbanization raises several challenges to city authority in terms of traffic. All the cities are becoming the smart city. In smart city every people having smartphone with internet connection. They are connected to the social site. Like as twitter, Facebook, whats app and google +. Twitter is the most popular micro-blogging social site. The user can update personal or public status as per reason. Twitter has many followers such as political leader, news reporter, radio reporter and fan followers. Today's date twitter is one current update means that it uses propriety algorithm to display trending topics. Twitter is much popular because of its micro-blogs which contain 140 characters. Status update message has been only important and valuable information to be written by the user. The user also writes real life event and their opinion on given topic through the public message. The source of information of the system is twitter tweet. The user message shared in social site is called status update message. It may contain meta-information of message such as time, location and name of the user. The classification of traffic related tweet from twitter use data mining technique, machine learning, statistical and natural language programming (NLP).

2. Literature Survey

Eleonora D'Andrea et al [1]. 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 to classify the tweets from twitter. This system is real-time traffic detection. The author defines system architecture for traffic detection from twitter tweets. 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.

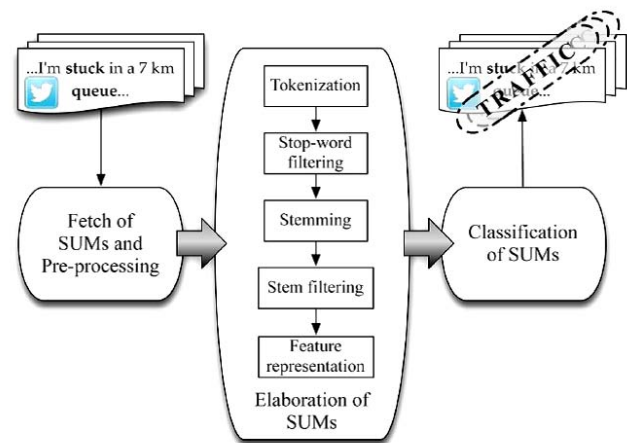


Figure 1: System architecture for traffic detection from Twitter stream analysis

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.

Yuchao Zhou et al [2]. 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.

Table 1: Classification of expected real world events

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 [3]. The system is based on traffic event detection and twitter messages interpretation.

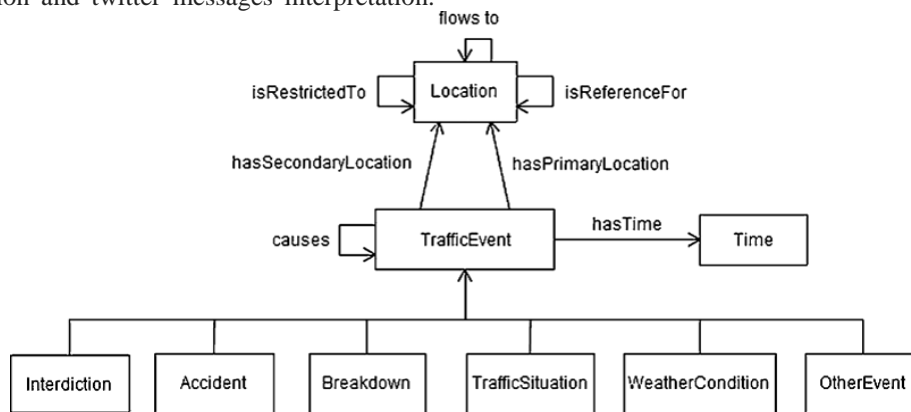


Figure 2: TEDO classes and object properties

Dwayne Henclewood et al [4]. 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 [9].

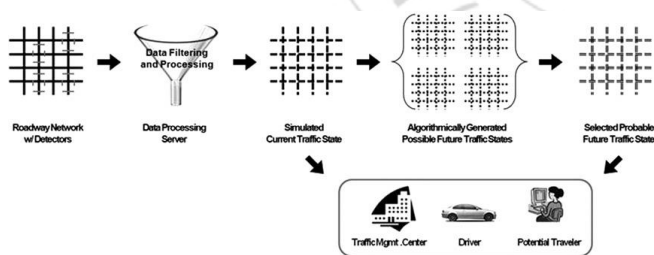


Figure 3: Conceptual framework for proposed methodology

SakkachinWongcharoen et al [5]. 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.

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 otherEvent.

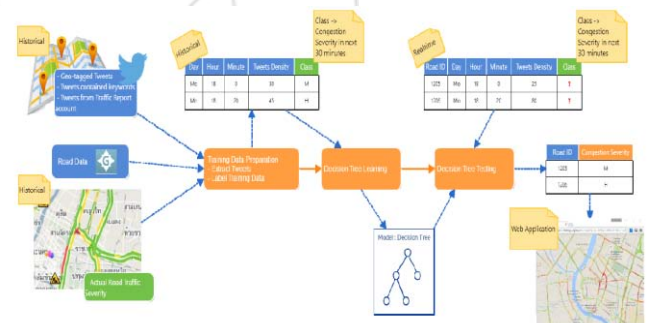


Figure 4: Overall Architecture

TomiJuntunen et al [6]. 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. Figure 6 shows 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

Server side uses PHP API is the layer between the user interface and MySQL database. It receives request from web

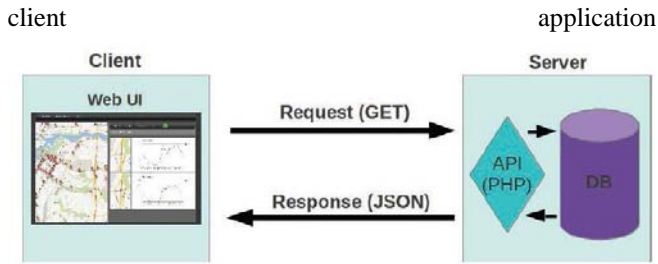


Figure 5: Overall Structure of the tool

SuponKlaithin et al [7]. The system is based on twitter data with data mining. Figure 5 show the data mining technique.

- 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.

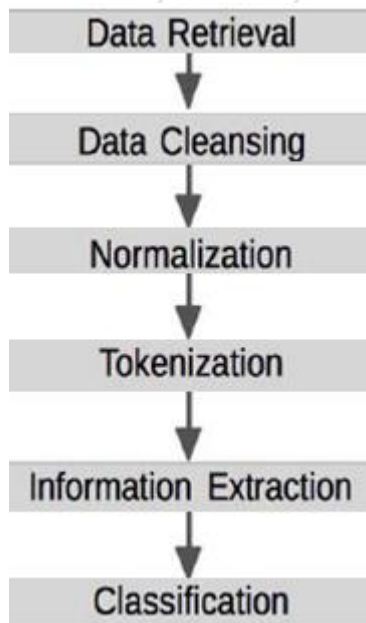


Figure 6: Process of Information Extraction and Classification

SaidiSiuhi et al [8]. In this paper, the author works on the mobile application. Mobile is today's thing that's why author choice mobile application. Smart mobile application in transportation sector such as parking, traffic safety, travel survey and route planning. Traffic-related data is big data on the social site then store on the cloud for fast computing. Analysis data through different server [9].

3. Problem Definition

The user traveling from one location to another location then it will use navigation system application for the find

shortest route. Find the traffic related tweets from the twitter. If traffic in this route, then suggest them an alternative route.

4. Methodology

In metro city traffic is one of the most problems. In the city everyone having a smartphone with the internet. They also connected with the social networks such as twitter, Facebook, and google+. One of them twitter is most popular than other.

Because of every political party or leader, news reporter, radio reporter is attaches twitter. It has massive source of information related to all.

Figure 7 shows the proposed system architecture.

- 1) Get public traffic tweets – In that collect, all traffic related tweets from twitter.
- 2) Show traffic tweets – In that list, all traffic related tweets.
- 3) After list out apply text mining technique in that
- 4) Tokenizing, remove stop words and stemming.
- 5) Database – sorted tweets stored into the database.
- 6) Web services – the web service use because comparing search route with traffic in the route. If traffic finds then give alternate path.

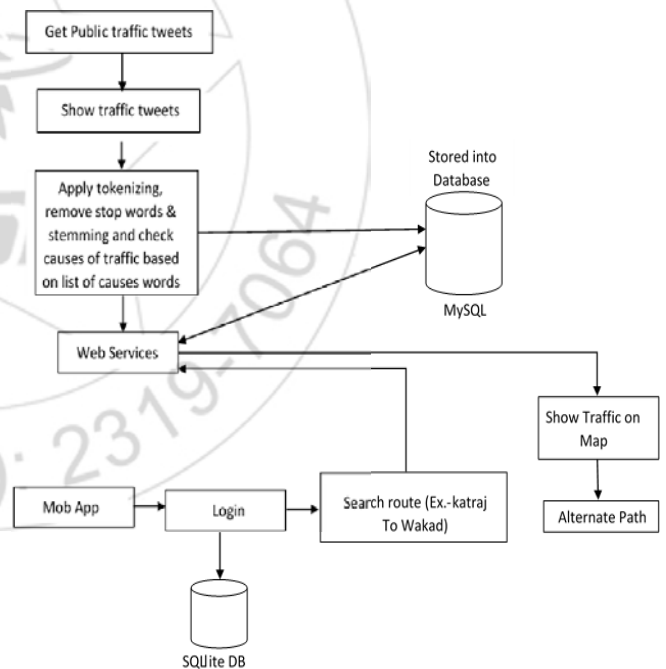


Figure 7: Proposed system architecture

5. Conclusion

In this paper survey on traffic detection system based on twitter tweets analysis. Study of all related to text mining technique. The social network has massive information. The social site is one of the things of all human daily life need. In the previous system only analysis, the traffic related tweets from twitter. Future work proposed system is real time when user finding a route in between source to destination suggest low traffic route.

References

- [1] Eleonora D'Andrea, Pietro Ducange, Beatrice Lazzerini, and Francesco Marcelloni, "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 Klaitin and Choochart Haruechaiyasak, "Traffic Information Extraction and Classification from Thai Twitter" Published By IEEE Journal 2016
- [8] Saidi Siuhi and Judith Mwakalonge, "Opportunities and challenges of smart mobile applications in transportation" published by ScienceDirect 2016.
- [9] Tatiana Sciammarella, Rodrigo S. Couto†, Marcelo G. Rubinstein†, Miguel Elias M. Campista and Lufis Henrique M. K. Costa "Analysis of Control Traffic in a Geo-distributed Collaborative Cloud" Published by IEEE Journal 2016.
- [10] Twitter Usage Statistics [online] Available: <http://www.internetlivestats.com/twitter-statistics>, October 21, 2015.

Author Profile



Sandeep Panchal is currently doing his M.Engg. Degree in Information Technology at Savitribai Phule Pune University, Pune, Maharashtra, India. Sandeep received his B.E Degree in Information Technology From Swami Ramanand Teerth Marathwada University Nanded, Maharashtra, India in 2015. His interest domain is Data Mining and Bigdata Hadoop.