International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Dynamic Traffic Analyzer Using Twitter

Harshita Rajwani[1], Srushti Somvanshi[2], Anuja Upadhye[3], Rutuja Vaidya[4] , Trupti Dange[5]

[1,2,3,4]B.E., Department of Computer Engineering, RMD Sinhgad School of Engineering, Warje,Pune-58, Savitribai Phule Pune University, India

[5]Professor, Department of Computer Engineering, RMD Sinhgad School of Engineering, Warje,Pune-58, Savitribai Phule Pune University, India

Abstract: ***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. Twitter is a social networking site which allows people to share and read tweets. The system fetches the tweets from twitter; applies natural language processing technique on them; categorizes the tweets related to traffic; notifies the registered users about it. Natural language processing (NLP) focuses on developing efficient algorithms to process text and convert it into machine understandable language. Here, we apply NLP on the tweets to detect the traffic.***

Keywords: Traffic event detection, Twitter, text mining, NLP

1. Introduction

Traffic congestion is a major problem today. One of the important reasons for more traffic congestion is due to increased number of vehicles on the road. The population is increasing and therefore people want their own vehicles to travel. As the number of vehicles increase, congestion also increases. To avoid this problem large cities use hardware sensors like cameras and radars to monitor traffic status. These tools function well but they have some limitations. One of these limitations is the high maintenance costs of these tools. Another limitation is that the tools cover only the certain area of the network. Hence, we implement a system which overcomes the problem of traffic congestion by using twitter.

Twitter is a social networking site which allows people to share and read tweets. Twitter users can broadcast tweets and follow other user‟s tweets. Tweets and their replies can be sent by cell phone application or www.twitter.com website. It‟s a platform where users share their thoughts, news, information and jokes in 140 characters of text or less. Profiles are usually public; anyone in the world can see what you write, unless you make it private.

The user logs in to the twitter account. System fetches the tweets through twitter stream analysis, and then it applies natural language processing technique on them to convert them into machine understandable form. The tweets are differentiated into basic categories such as traffic or non-traffic. Those related to traffic are further differentiated on the basis of their causes like earthquakes, accidents, landslides, fire etc. Further user receives notification about the event detected and also will be able to see the location of the traffic on map.

2. Literature Survey

Eleonora D‟Andrea, Pietro Ducange, Beatrice Lazzerini, and Francesco Marcelloni[1] detected traffic and traffic related events through real time monitoring system using twitter stream analysis. In their proposed paper ,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. As future work, they are planning to integrate their system with an application for analyzing the official traffic news web sites, so as to capture traffic condition notifications in real-time. Further, they are investigating the integration of their system into a more complex traffic detection infrastructure. This infrastructure may include both advanced physical sensors and social sensors such as streams of tweets.

Vikram Singh and Balwinder Saini[2] have proposed an effective tokenization approach which is based on training vector and result shows that efficiency/ effectiveness of proposed algorithm. In the results, it is shown that tokenization with pre-processing generates better tokens, as it is with less number of token generated and less storage space is required and it facilitates more accuracy in results retrieval and it is also responsible for reducing the time of information retrieval model.

Maximilian Walther and Michael Kaisser [3] detect Geo-spatial Event using twitter. In this paper, they describe a new scenario and approach to tackle it, detecting real-world events in real-time in a monitored geographic area.In this approach they gathered tweets for target events that can be defined by a user via keywords. The authors apply classification and particle filtering methods for detecting events. They have chosen MongoDB as the database system because it supports geospatial and temporal indices. They have common theme as if people tweeting from the same place use the same words, it is likely that they talk about the same thing which probably is some noteworthy event.

Varsha Sahayak, Vijaya Shete and Apashabi Pathan [4]discussed about a paradigm to extract the sentiment from a famous micro blogging service, Twitter, where users post their opinions for everything. An approach is introduced that automatically classifies the sentiments of Tweets taken from Twitter dataset. These tweets are classified as positive, negative or neutral with respect to a query term. Further features like emoticons, neutralization, negation handling and capitalization/ internationalization can also be

implemented as they have recently become a huge part of the internet.

Rui LI , Kin Hou Lei , Ravi Khadiwala and Kevin Chen-Chuan Chang [5] proposed a Twitter-based Event Detection and Analysis System (TEDAS), which helps to detect new events, to analyze the spatial and temporal pattern of an event, and to identify importance of events. Technically, they developed an efficient CDE(Crime and Disaster related Events) focused crawler, and explored valuable and novel features from Twitter to classify and rank tweets, and to predict locations from tweets. They implement a real time system that can detect and analyze CDEs based on millions of tweets and users.

3. Proposed System Work

3.1 Modules

Module 1: Fetching
Initially we login into twitter and after successful login we fetch the tweets from that respective account.

Module 2: NLP
Once the tweets are fetched, we perform NLP technique where we perform tokenization, stop word filtering, stemming, stem filtering and feature representation.

Module 3: Tagging
We tag the cause of traffic congestion like earthquakes, accident, landslides etc. Also the location of it is displayed on map.

Module 4: Notification
The user receives the notification about where and why the traffic congestion has occurred.

3.2 System Architecture

Figure 1: Architecture diagram for proposed system

Twitter member logs in with the username and password. Here we are using the AES algorithm for security purpose. It is used for encrypting and decrypting the data. After successful user login the system fetches the data. Next, by applying sentiment analysis we classify into positive and negative events. Further NLP technique is used to retrieve traffic related tweets. Further we perform classification on the result of NLP technique where we identify the causes of traffic congestion and treat each cause as one cluster using k-means algorithm. The traffic location is then displayed it on map.

4. Technology

In this system text mining is used to extract tweets related to traffic, for that we are using Natural Language Processing Technique which includes following steps:

a) Tokenization: It consists transforming a stream of characters into stream of meaningful tokens. For instance, if the post says " I am stuck in traffic from 2 hours…getting bored …want to go home!!!!!!" ,then after performing tokenization we get - I, am, stuck, in, traffic, from, 2, hours, getting, bored, want, to, go, home .
b) Stop word filtering: It consists removing of conjunctions, prepositions, pronouns etc such as a, an, the, is, it etc. On performing stop word filtering on the above result of tokenization we get - Stuck, traffic, 2, hours, getting, bored, want, go, home.
c) Stemming: It consists of removing suffix from tokens and converting them into their root form. After performing stemming on the above result we get- Stuck, traffic, 2, hours, get, bore, want, go, home.
d) Stem filtering: It compares the sequence of stems with the set of relevant stems and removes the stems which do not belong to the relevant stems.
e) Feature representation: It consists the process of identifying the traffic related words from the relevant set of stems.

After processing NLP technique system classify traffic related tweets on the basis of causes for traffic, such as accident, fires, natural disaster. For that we perform K-means algorithm, where we treat each cause as one cluster.

5. Feasibility Study

NP type problem -
A problem is assigned to the NP (nondeterministic polynomial time) class if it is solvable in polynomial time by a nondeterministic Turing machine. A is always also NP.

NP-Complete problem -
A given problem is said to be NP-Complete, if it can be solved within the determined polynomial time. The attribute values should completely satisfy the given polynomial expression. In computational complexity theory, a decision problem is NP-complete when it is both in NP and NP-hard.

This is considered with specifying equipment and software that will successful satisfy the user requirement the technical needs of the system may vary considerably but might include

- The facility to produce outputs in a given time.
- Response time under certain conditions.

- Ability to process a certain column of transaction at a particular speed.

The data processing algorithms are NP type. Because we can get & verify the solution set. Hence the problem statement involving it are NP Complete.

6. Mathematical Module

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 updations}
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.

7. System Features

The system detects traffic events in real-time. Depending on certain search criteria, the system fetches tweets from twitter. It processes these tweets and 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.

8. Advantages

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

9. Limitations

The main limitation of the system is network availability that is the internet connection should be available throughout the process.

10. Conclusion and Future Scope

In this paper, we have proposed a system for dynamic traffic analyzer 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. The system‟s mobile application is limited to android devices but in future it can be developed for iOS , Blackberry, Windows phone OS.

References

[1] Eleonora D‟Andrea, Pietro Ducange, Beatrice Lazzerini, and Francesco Marcelloni ”Real-Time Detection of Traffic From Twitter Stream Analysis”,Ieee Transactions On Intelligent Transportation Systems 2015.
[2] Vikram Singh and Balwinder Saini “An Effective Tokenization Algorithm for Information Retrieval System” CS & IT-CSCP 2014
[3] Maximilian Walther and Michael Kaisser, ”Geo-spatial Event Detection in the Twitter Stream”, P. Serdyukov et al. (Eds.): ECIR 2013, LNCS 7814, pp. 356367, 2013.cS pringer VerlagBerlinHeidelberg 2013.
[4] Varsha Sahayak Vijaya Shete Apashabi Pathan “Sentiment Analysis on Twitter Data” International Journal of Innovative Research in Advanced Engineering (IJIRAE) ,January 2015.
[5] Rui LI, Kin Hou Lei, Ravi Khadiwala, Kevin Chen-Chuan Chang “TEDAS: a Twitter Based Event Detection and Analysis System”.
[6] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo ,”Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development”,IEEE transactions on knowledge and data engineering, vol.25,no.4, april 2013.

[7] Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo," Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors", 2010.
[8] H. Takemura and K. Tajima, "Tweet classification based on their lifetime duration," in *Proc. 21st ACM Int. CIKM, Shanghai, China, 2012*,pp. 2367–2370.
[9] V. Gupta, S. Gurpreet, and S. Lehal, "A survey of text mining techniques and applications," J. Emerging Technol. Web Intell., vol. 1, no. 1,pp. 60–76, Aug. 2009.
[10] L. H. Patil and M. Atique, "A novel feature selection based on information gain using WordNet," in *Proc. SAI Conf.*, London, U.K., 2013,pp. 625–629.
[11] Feasibility Study – NP complete type problems: https://en.wikipedia.org/wiki/NP-completeness
[12] Feasibility Study - NP type problems:http://mathworld.wolfram.com/NP-Problem.html

Author Profile

Harshita Rajwani is currently pursuing the degree of B.E in Computer Engineering from the RMD Sinhgad School of Engineering which is affiliated to the Savitribai Phule Pune University (formerly The University of Pune)

Srushti Somvanshi is currently pursuing the degree of B.E in Computer Engineering from the RMD Sinhgad School of Engineering which is affiliated to the Savitribai Phule Pune University (formerly The University of Pune)

Anuja Upadhye is currently pursuing the degree of B.E in Computer Engineering from the RMD Sinhgad School of Engineering which is affiliated to the Savitribai Phule Pune University (formerly The University of Pune)

Rutuja Vaidya is currently pursuing the degree of B.E in Computer Engineering from the RMD Sinhgad School of Engineering which is affiliated to the Savitribai Phule Pune University (formerly The University of Pune)

Trupti K. Dange received the B.E. and M.E. Degree in Computer engineering. She now is working with the RMD Sinhgad School of Engineering , Warje, Pune as Asst. Professor. She has experience of more than 4 years and her Area of specialization - Web Technoloy, Software Engg./Testing