International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Extract User Travel Habits, Road Conditions & Road Traffic Using Twitter and Pothole Detection

Miss. Ashwini Gaikwad[1] and Prof. Pravin Nimbalkar[2]

[1] PG Students, [2]Assistant Professor, Department of Computer Engineering, JSPM ICORE, Wagholi, Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Poorly maintained roads are a fact of life in most developing countries including our India. A well-maintained road network is a must for the well-being and the development of any country. Our work main focuses to create an effective road surface monitoring system with help of one of social media platform. Twitter is a social networking service with more than three hundred million users, producing a large amount of facts each day. Twitter's most important characteristic is its ability for users to tweet(message) these tweets are all about events, situations, feelings, opinions, or even something totally new, in real time. The social media tweet has textual content which are mined that allows to perceive the lawsuits regarding various avenue transportation issues of site visitors, such as accidents, and potholes. In order to pick out and segregate tweets associated with exceptional troubles, keyword-based methods have been used formerly, these strategies are entirely depending on seed key phrases which can be manually given and these sets of keywords aren't sufficient to crawl through all tweet's posts. So, to conquer this issue, a singular approach has been proposed that captures the semantic context via dense word embedding by employing the word2vec model. However, the system of tweet segregation on the idea of semantic comparable key phrases may additionally suffer from the problem of pragmatic ambiguity. To take care of this, Word2Vec model has been implemented to shape the semantically similar tweets. Furthermore, the hotspots had been identified similar to each category. As there is scarcity of geo tagged tweets, we have proposed a hybrid method which amalgamates Named Entity Recognition (NER), Part of speech (POS), and Regular Expression (RE) to extract the vicinity statistics from the tweet text. This work will assist to avoid injuries and can use to become aware of trouble areas ahead of happening. The government may be alerted to take preventive movements and these preventive actions can save money. Our work contributes to automation of pothole detection using Android mobile, Google map, microcontroller, GSM modem, GPS, Ultrasonic sensor making it liable to be used for social welfare.

Keywords: *Travel Habits, Road Conditions Detection, Road Traffic detection, Incident Detection, Social Media, Named Entity Recognition, Word Embedding, Transportation, Pothole Detection.*

1. Introduction

In social media, posts analysis has always been considered as the most challenging task for twitter analyst/data scientist. In India, four major tier-1 cities (Mumbai, Delhi, Kolkata, and Bengaluru) annually losses 22 billion dollar due to congestion. It mainly induced from non-recurrent events such as accident, adverse road conditions, construction on roads, potholes, adverse weather condition, and inadequate drainage. Due to this individual has to spend more than one and half hour longer during the peak hour to cover the same distance as on non-peak hour. Furthermore, it's only one significant challenge in-front of infrastructural manager and commuters as these events would take most of the time as well as causes a number of deaths. The report published by MORTH (Ministry of Road Transport & Highways) shows that the number of fatalities in India due to potholes from the last five year is 14,296 which is much higher than the casualties due to terrorist or Naxal attacks. Whereas death due to road construction is increased by 50% in 2017 (i.e. 4250). So to overcome, it's essential to identify these events in a timely and efficient manner.

In this work, we identify these nonrecurrent events effectively and inexpensively, by leveraging the potential of social networking sites (like Twitter, Facebook etc). From last few years, peoples interests are more inclined towards these sites to express their opinion, feeling and suggestion regarding any problem or event as short text. Twitter is one of such platforms which has more than 335 million monthly active users over the globe, where users interact with each other through a textual/visual post that is known as "tweet". That results in a vast amount of data records as posts which are very informative and can be used in a

number of applications. As a case study, we consider tier-1 cities in India (Mumbai, Delhi, Hyderabad, Chennai, Kolkata, and Bengaluru) to show the city characteristics, i.e. (traffic congestion, injuries, commuters tour behavior and road circumstance) by harnessing Twitter statistics. We broadly categories the non-recurrent events into three categories i.e. (accident, traffic, and potholes).

Previously, some researchers have dedicated their time to identify the traffic incident by developing an algorithm to spot the event in real time by using the physical sensors. However, these algorithms work well over the highways, but not on local arterials because it is costly as well as difficult to cover every locality under the physical sensor. So in these paintings, our number one motivation is to set up an efficient and price-effective machine to perceive the non-recurrent incidents in each highway in addition to local arterials. Recently, it has been found that Twitter records have turn out to be a rich supply of statistics pertaining to injuries, congestion, terrible lighting, potholes. But it is very hard to identify occasions from the tweet texts due to the fact tweets publish is generally informal, quick, unstructured and regularly includes grammatical mistakes, misspelling, and quite a few noises. That makes a challenging task for researchers, to identify linguistic features for building NLP (Natural Language Processing) based application. It might be due to the restriction imposed by Twitter over tweet post length, i.e. 140 character limits. Thus, it makes text classification and information extraction a challenging problem. So we have performed various data prepossessing steps to convert the text into a readable form.

As clever motors have become greater ubiquitous, the capability now exists to locate environmental road features (e.g., potholes, road incline perspective, etc.) from their embedded sensor information. By aggregating statistics from multiple automobiles, crowdsourcing may be leveraged to discover environmental statistics with improved accuracy. I consciousness on the usage of such data to locate and localize potholes on multilane roads. Extracting data from aggregated vehicle facts is hard because of undersampling sensors, sensor mobility, asynchronous sensor operation, sensor noise, car and avenue heterogeneity, and GPS position error. GPS position error is particularly problematic in multi-lane environments since the position error is generally larger than standard lane widths. In this paper, I look at those troubles and expand a crowdsourced machine to detect and localize potholes in multi-lane environments the usage of accelerometer data from embedded vehicle sensors. The crowdsourced system reduces the required network bandwidth via figuring out street incline and financial institution attitude statistics in each automobile to filter acceleration additives that do not correspond to pothole situations. I evaluate our system on simulated and real-world statistics, examine tradeoffs within the range of motors and the amount of bandwidth required for correct detection, and examine the effects to the less complicated unmarried lane detection situation.

2. Review of Literature

EvenTweet: Online Localized Event Detection from Twitter, Microblogging offerings together with Twitter, Facebook, and Foursquare have turn out to be foremost assets for data approximately real-world events. Most methods that purpose at extracting occasion facts from such resources usually use the temporal context of messages. However, exploiting the place information of georeferenced messages, too, is crucial to locate localized activities, consisting of public occasions or emergency situations. Users posting messages which are close to the area of an event serve as human sensors to explain an event. In this demonstration, we gift a unique framework to stumble on localized events in real-time from a Twitter circulate and to song the evolution of such activities over the years. For this, Spatio-temporal characteristics of keywords are constantly extracted to discover significant applicants for occasion descriptions. Then, localized occasion facts are extracted with the aid of clustering keywords in keeping with their spatial similarity. To decide the most essential events in a (current) time frame, we introduce a scoring scheme for occasions. We demonstrate the functionality of our gadget, known as Even-Tweet, using a circulate of tweets from Europe all through the 2012 UEFA European Football Championship.[1]

Twitcident: Fighting Fire with Information from Social Web Streams, this study, we shown Twitcident, a framework and Web based totally system for filtering, looking and reading facts approximately real-global incidents or crises. Twitcident connects to emergency broadcasting offerings and mechanically begins tracking and filtering records from Social Web streams (Twitter) whilst a new incident occurs. It enriches the semantics of streamed Twitter messages to profile incidents and to constantly decorate and adapt the information filtering to the modernday temporal context. Faceted search and analytical gadget permit customers to retrieve precise records fragments and evaluate and examine the cutting-edge state of affairs as stated at the Social Web.[2]

Getting There First: Real-Time Detection of Real-World Incidents on Twitter, Social networking and microblogging offerings such as Twitter has to turn out to be a precious supply of records on modern-day events. Widespread use of Twitter on mobile gadgets and personal pc systems permits customers to percentage short messages on any assignment in actual-time, for that reason making it suitable for early detection of unexpected activities wherein a speedy response is vital. In this paper, we gift a web technique for the detection of actual-international events in Twitter records, including herbal disasters or man-made catastrophes, by way of studying Twitter information. Our method combines one-of-a-type

textual and frequency components that represent or approximate interesting semantic elements of an occasion. We use visualization as a validation car, which permits us to recognize which components are relevant and what effect the parameters have on the results of our event detection
algorithm.[3]

Event Detection in Twitter, Twitter, as the shape of social media, is fast rising in current years. Users are using Twitter to report reallife events. This paper specializes in detecting those activities with the aid of reading the textual content stream on Twitter. Although event detection has lengthy been a study topic, the traits of Twitter make it a non-trivial venture. Tweets reporting such events are usually beaten by means of hifood niess ". Moreover, occasion detection set of rules desires to be scalable given the sheer amount of tweets. This paper tries to tackle these challenges with EDCoW (Event Detection with Clustering of Wavelet-based Signals). EDCoW builds indicators for person words by applying wavelet analysis on the frequency based uncooked signals of the phrases. It then filters away the trivial phrases by means of searching at their corresponding sign vehicle correlations. The closing words are then clustered to form activities with a modularityprimarily based graph partitioning technique. Experimental results show promising end result of EDCoW.[4]

I See a Car Crash: Real-Time Detection of Small Scale Incidents in Microblogs, Microblogs are increasingly more gaining attention as an essential records source in emergency control. Nevertheless, it is nonetheless tough to reuse this facts supply during emergency conditions, because of the sheer quantity of unstructured statistics. Especially for detecting small scale activities like car crashes, there are simplest small bits of facts, for this reason complicating the detection of relevant records. We gift an answer for a real-time identity of small scale incidents using microblogs, thereby allowing to increase the situational focus via harvesting extra information about incidents. Our method is a device learning set of rules combining text class and semantic enrichment of microblogs. An evaluation based totally shows that our answer permits the identity of small scale incidents with an accuracy of 89% further to the detection of all incidents posted in actualtime Linked Open Government Data.[5]

Real-Time Detection of Traffic From Twitter Stream Analysis, Social networks have been in recent times employed as a source of information for event detection, with precise reference to avenue visitors congestion and automobile accidents. In this paper, we gift a real-time monitoring machine for traffic event detection from Twitter circulate assessment. The device fetches tweets from Twitter in keeping with numerous seek requirements; strategies tweets, by using way of utilizing textual content mining strategies; and in the long run, performs the type of tweets. The aim is to assign the ideal class label to every tweet, as related to a traffic occasion or not. The traffic detection system changed into employed for actual-time monitoring of numerous regions of the Italian street community, bearing in mind detection of visitors occasions nearly in real-time, regularly before online site visitors information internet web sites. We hired the support vector gadget as a class version, and we performed an accuracy cost of ninety five. Seventy five% with the aid of solving a binary classification problem (traffic as opposed to nontraffic tweets). We had been also capable of discriminate if visitors is because of An external occasion or no longer, by means of fixing a multiclass classification problem and obtaining an accuracy price of 88.89%.[6]

Potholes detection based on SVM in the pavement distress image, present a system in which upcoming technology Wi-Fi based architecture for pothole detection and warning system which guide the passenger in ignore the pothole on the roads by prior warning information. The system has many access points which is located on the road sides for broadcasting information, which can be collect by Wi-Fi enable transport vehicles as they enter the area blanketed by the access points influence. The application is developed a combination in the vehicle so as to alarm the driver as form of a visual signal, audio signal. Metrology and visualization of potholes using the microsoft Kinect sensor, describe each individual vehicle is equipped with special radio frequency identification (RFID) tag which makes it impossible to remove or destroy. If RFID-tab examine belongs to the stolen vehicle, then a message is sent using GSM SIM to the police manage room. In addition, while an ambulance is passing thru the junction, it'll tell to the site visitors controller in the junction to show at the inexperienced mild. In this system the wireless get admission to factor collect the statistics approximately potholes, it bypass this statistics to BMC the use of wi-fi broadcast system. Traffic controller gather the environmental facts also has an accelerator which could measure each the vertical and the horizontal acceleration. It is on the whole used in twist of fate detection device.[7][8]

Implementing intelligent traffic control system for congestion control, proposed a paper classification of potholes are given explain. Potholes are summarized taking the help of location, length, shape and depth. There are multiple literature who gives us multiple technique to pothole detection and gives better survey and pavement quality with prior exploration and immediate action. For particular method it Need for growing a class tenet for supporting decision-making system of pothole repair.[9]

The system shows traffic monitoring system. The work completes with a flowchart. Because of focus on mobile application, it has made design of database, interface, and information's classification
based on Indonesia's regulation.[10]

This work[11] has shown that tweets alone can be used to estimate road traffic congestion condition as tweets density and hours of day are useful attributes for building a congestion severity prediction model.

This system has discussed the development of application software for the analysis and monitoring of traffic on the highway or toll road. Information on state on the traffic also provides by the system via Twitter social media network as well as statistical data which can be used by the traffic authority. The software is made for Windows-based desktop, and includes the friendly GUI for ease of operation[12]

3. Problem Statement

Propose system implementing a real time usage of Twitter for extract User Travel Habits, Road Conditions and Road Traffic. Proposed system gives the original pothole location in user route. Also users are able to see updated safe location on their device. In this way, the proposed system provide security alert to the driver.

4. Proposed System

To segregate the tweets, we proposed semantically similar adaptive keyword generative approach through leveraging the semantic context through dense word embedding the use of Word2vec version. The proposed technique overcomes the inability of traditional techniques i.E. Key-word primarily based segregation, and classification by means of using a gadget getting to know algorithm. This undertaking provides a technique to move slowly, pre-method and clear out freely available tweets. These tweets submit then analyzed to extract non recurrent activities data via using deep mastering and NLP strategies. Proposed pothole detection device uses accelerometer facts from embedded automobile sensors and is therefore reliant on some preliminary vehicles being not able to avoid the potholes. The statistics received with the aid of the ones vehicles that could not avoid riding over the potholes can be shared with subsequent automobiles for his or her advantage.

5. System Architecture

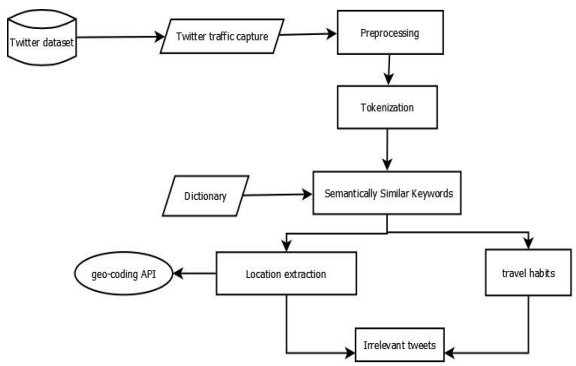

Figure 1. System architecture

6. Algorithm

Algorithm1 : Semantically Extended Keywords Generation

- Input: Clean tweets Ƈ = {δ1, δ2, δ3, ... , δn}
- Output: Expended keyword list W2V_KD = E(SD_KD)

Step 1: res1= [' ']
Step 2: res3= [' ']
Step 3: res2= [' ']
Step 4: for each tweet in Ƈ do
Step 5: T = nltk. tokenize(tweet)
Step 6: end for
Step 7: model=models.gensim.Word2Vec(T)
Step 8: SD_KD = {s1, s2, ..., sn}
Step 9: res3=SD_KD
Step 10: for each keyword in SD_KD do Step 11: res1=model.most_similar(Keyword, topn=5) Step 12: res2.append (res1)
Step 13: res2=unique (res2)
Step 14: if (res2 - SD_KD) = ϕ then
Step 15: Stop
Step 16: Exit
Step 17: else
Step 18: res3= res2-SD_KD
Step 19: SD_KD = SD_KD U res3
Step 20: goto step 10
Step 21: end if
Step 22: end for

Algorithm 2: K-Means Algorithm

- Input: K- the number of clusters
- D: A data set containing n objects
- Output: A set of k clusters

Steps 1: Randomly pick out okay statistics gadgets from dataset D as preliminary cluster middle. Steps 2: Repeat.
Steps 3: Calculate the space among every facts object di (1 <= i <= n) and all k cluster middle cj (1 <= j <= k) and assign facts item di to the nearest cluster.
Steps 4: For every cluster j (1 <= j <= ok),recalculate the cluster middle. Steps 5: Until no changing in the center of clusters.
The computational complexity of the algorithm is O (nkt) Where, n: the total number of objects k: the number of clusters t: the number of iterations Until no converting in the center of clusters.
The computational complexity of the algorithm is O (nkt)
Where, n: the whole quantity of items k: the variety of clusters t: the range of iterations **Algorithm3 :** KNN Algorithm Steps:

- Determine parameter K = variety of nearest neighbor.
- Calculate the gap between the query-instance and all training sample.

• Sort the space and determine nearest neighbor based at the kth minimum distance. Gather the class of the nearest neighbor.

• Use simple majority of the class of nearest friends as prediction fee of the question instance

7. System Requirements

A. Software Requirement

1) OS- Microsoft Windows 7 or Above
2) Programming Language- JAVA, Embedded C
3) Database- MySQL
4) Tools- Netbeans, Android Studio, Arduino IDE

B. Hardware Requirement

1) Processor- Core Intel 3 or Above
2) RAM- 2GB or Higher
3) Hard Disk- 100GB (min)

8. Result and Discussion

The system automatically detects the potholes and humps and sends the records regarding this to the vehicle drivers, in an effort to keep away from injuries. This is a value green solution for detection of humps and potholes. This device is powerful even in rainy season when roads are flooded with rain water in addition to in winter at some point of low visibility, as the alerts are sent from the stored facts inside the server/database. This gadget facilitates us to keep away from dreadful potholes and humps and consequently to avoid any tragic injuries because of bad road conditions. The information can also be used by the Government for the upkeep of the roads.

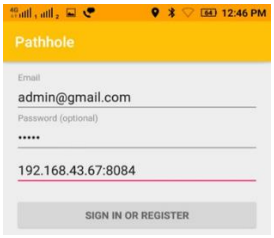

Figure 2: User Registration

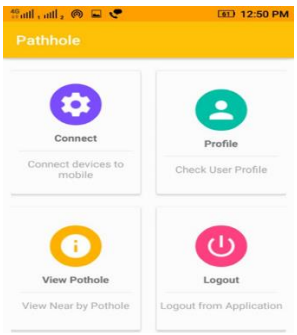

Figure 3: Home Page

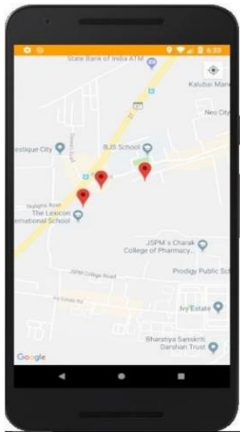

Figure 4 : Pothole Detection on map

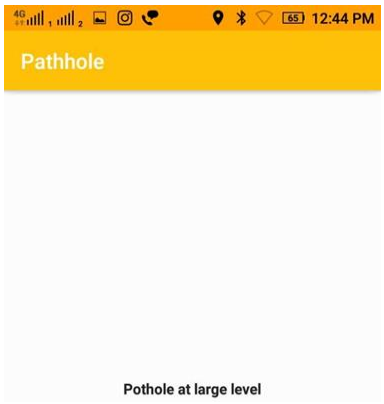

Figure 5 : Pothole at large level detected

The system works in a way to benefit the driver, Figure 1 shows the user registration page wherein the user is able to login and specify his own location ,the next page on list is home page (figure 2) where the user is able to connect and see the potholes through the pathway which he will be travelling.Figure 3 specify potholes.
Thus the work is developed to gather the information through twitter and specify the traffic congestion or potholes.

Conclusion

This will help to avoid accidents and can use to identify pothole problem areas early. In countries where updated economic growth and excellent technology have increase to gives impact on the quality of traditional transport system over Intelligent transportation system. This machine may be similarly stepped forward to do not forget the above fact and update server database consequently. Also, Google maps can be incorporated inside the proposed machine to improve consumer enjoy. This system introduced a framework that identifies incidents caused by nonrecurrent events (accident, potholes, and traffic) from the social media platform. The proposed framework will be distributed into five main components which include collecting data from multiple sources (i.e., hashtags, handle, and bounding box), data preprocessing, identification of similar

semantic keywords corresponding to the different categories, removing the pragmatic ambiguity and content based location identification for finding the vulnerable areas. The predominant findings of this work are as follows:

- Introduce a robust method to classify the tweets into different categories by leveraging dense vector embedding to generate similar semantic keywords.

- The location information from textual content was efficiently extracted by implementing a hybrid approach which is an amalgamation of NER, POS and Regular Expression (RE).

- The temporal and spatial analysis have been performed to determine user mobility patterns through their tweeting behaviour which can be used effortlessly by the government traffic agencies.

- Furthermore, we also identified the top 25 hotspots with respect to each category and listed some of the reasons (water logging, huge rain, absence of sign boards, non functional street lights, traffic rules violation, littering on roads, road construction and illmaintained roads, potholes and improperly parked vehicles) due to which people largely face traffic congestion and accidents.

References

[1]. Hamed Abdelhaq, Christian Sengstock, and Michael Gertz, "EvenTweet: Online Localized Event Detection from Twitter", Proceedings of the VLDB Endowment, 2013.

[2]. Fabian Abel, Claudia Hauff, Geert-Jan Houben, Ke Tao TU Delft, "Twitcident: Fighting Fire with Information from Social Web Streams", WWW Demos Track, 2012.

[3]. Milos Krstajic, Christian Rohrdantz, Michael Hund, Andreas Weiler, "Getting There First: Real-Time Detection of RealWorld Incidents on Twitter", 2nd IEEE Workshop on Interactive Visual Text Analytics, 2012.

[4]. Jianshu Weng, Bu-Sung Lee, "Event Detection in Twitter", Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011.

[5]. Axel Schulz1, Petar Ristoski, and Heiko Paulheim, "I See a Car Crash: RealTime Detection of Small Scale Incidents in Microblogs", Springer-Verlag Berlin Heidelberg, 2013.

[6]. Eleonora D'Andrea, Pietro Ducange, Beatrice Lazzerini, and Francesco Marcelloni, "Real-Time Detection of Traffic From Twitter Stream Analysis", IEEE Transaction on Intelligent Transportation Systems, 2015.

[7]. J. Lin and Y. Liu, Potholes detection based on SVM in the pavement distress image, in Proc. 9th Int. Symp. Distrib. Comput. Appl. Bus. Eng. Sci., Aug. 2010.

[8]. I. Moazzam, K. Kamal, S. Mathavan, S. Usman, and M. Rahman, Metrology and visualization of potholes using the microsoft Kinect sensor, in Proc. 16th Int. IEEE Conf. Intell. Transp. Syst., Oct. 2013.

[9]. R. Sundar, S. Hebbar, and V. Golla, Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection, IEEE Sensors J., vol. 15, no. 2, pp. 1109 1113, Feb. 2015.

[10]. G. Elkana, I. G. B. B. Nugraha and S. H. Supangkat, "Road traffic information for daily drive with level of traffic condition," *International Conference on ICT for Smart Society*, Jakarta, 2013, pp. 1-4.

[11]. S. Wongcharoen and T. Senivongse, "Twitter analysis of road traffic congestion severity estimation," *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Khon Kaen, 2016, pp. 1-6.

[12]. P. H. L. Tobing and Hendrawan, "Application of system monitoring and analysis of vehicle traffic on toll road," *2014 8th International Conference on Telecommunication Systems Services and Applications (TSSA)*, Kuta, 2014, pp. 1-5.