EasyChair Preprint

№ 5194

Using Twitter Extract and Identified Road Traffic, Road Condition And Road Accident Location

Pratik Jambhale and Manisha Singh

March 23, 2021

Using Twitter extract and Identified Road traffic,Road Condition And Road Accident location

Pratik C.Jambhale
Department of Computer Engineering
Dhole Patil College of Engineering, Wagholi, Pune
Savitribai Phule Pune University
pratikjambhale6939@gmail.com

Prof.manisha singh
Department of Computer Engineering
Dhole Patil College of Engineering, Wagholi, Pune
Savitribai Phule Pune University
manishasingh4314@gmail.com

***Abstract*—In most developed nations, like our India, poorly managed roads are a reality in existence. For the well-being and prosperity of each nation a well-maintained road network is a must. Our focus is on the implementation of an efficient tracking system for road surfaces using one of the social media channels. Twitter is a social networking site that provides a vast number of facts daily, involving more than billion people. The most critical feature of Twitter is the potential for tweeters, which include incidents, scenarios, thoughts, views or even something completely different in real time. The social media tweet contains textual contents which enable users, including accidents and potholes, to view the lawsuits concerning different issues related to the transport of web sites. In order to take out and segregate tweets related to extraordinary problems, the keyword based approaches were formerly employed. These solutions are dependent solely on seed key phrases that can be provided manually. To address this problem, therefore, a special approach has been suggested, which uses the words2vec model to capture the semántic context through dense word embedding. However, the tweet separating method for the concept of equivalent semantic main phrases may still be subject to the question of pragmatic uncertainty. For this reason, the Word2Vec model is built for shaping tweets close to the semantinal ones. In addition, hotspots identical to each group have been established. These tasks help prevent accidents and can be used to identify danger spots prior to occurrence. Preventive movements can be alerted by the government and resources can be saved by preventive measures.**

***Index Terms*—Road Traffic detection, Incident Detection, Social Media, Twitter,Machine Learning**

I. Introduction

Social networking has often found publishing research to be tweet analyst/data scientist's most difficult job. Four major cities of tier 1 (Mumbai, Delhi, Kolkata, and Bengaluru) lose $22 billion per year as a result of congestion in India. This has been largely caused by non-recurrent incidents such as accidents, unfavourable road conditions, road building, potholes, bad weather and poor drainage. Because of the fact that this person needs more than one and a half hour to travel the same distance as for non-peak hours during peak hour. Furthermore, the infrastructure managers and the commuter face only one major obstacle, as these events will often take the time.And a lot of deaths are also causing. The MORTH (Ministry of Highways) study reveals that 14,296 deaths from potholes over the last five years are much higher than those caused by militant attacks in India or Naxals. In 2017, the number of traffic deaths rose by 50 percent (i.e. 4250). In order to resolve these incidents, it is important to recognise them rapidly and efficiently.

In this work the ability of social networking sites is used to classify these non-recent incidents easily and cheaply (like Twitter, Facebook etc). Since recent years, the interests of the people have become more inclined to share their opinions, feelings and recommendations surrounding these locations a short text problem or case. Twitter is one such site with over 335 million monthly active users across the world, with users engaging via a "tweet" textual/visual message. This adds to a huge number of data records as posts that can be used and are very informative in a number of applications. In our research, we see the city features, namely (traffic congestion, tragedies, travellers' tour conduct and road circumstances), as seen in Tier-1 cities in India (Mumbai, Delhi, Hyderabad, Chennai, Kolkata and Bengaluru). The non-recurrence cases are generally classified in three groups, i.e. (accident, traffic, and potholes).

Earlier, some researchers have taken time to classify the traffic occurrence by using physical sensors to create an algorithm to spot the accident in real time. However, these algorithms operate well across roads, but not on local arterials, since each location under the physical sensor is both expensive and difficult to protect. Our number one inspiration in these paintings is to create a reliable, cost-effective machine to perceive, in addition to local artery, the non-recurrent incidents at each lane. Recently Twitter archives proved to be a rich range of injury, pollution, bad lighting, potholes figures. But the fact that tweets are usually casual, quick, unstructured and frequently contain grammatical errors, misspellings and a number of noises makes it very difficult to detect opportunities in tweet texts. For researchers, this is a difficult challenge to classify linguistic characteristics for NLP-based applications. That may be because of Twitter's limitation on the post-length tweet, i.e. 140 maximum characters. It therefore makes the classification of text and the retrieval of information a difficult issue. We have also taken different measures to prepare the details in order to make the text

readable.

As smart motors have become omnipresent, they are now able to determine environmental features of their integrated sensor knowledge (such as potholes, road incline viewpoint, etc.). When added Crowdsourcing can be exploited to detect environmental statistics with increased precision from different vehicles. I know how to spot and find potholes on multi-lane roads with such information. The elimination of data from vehicle aggregation is difficult because of undersampling sensors, sensor mobility, asynchronous sensor activities, sensor noise, heterogeneity of cars and avenues, and GPS location errors. In multi-lane settings, a GPS location error is particularly troublesome because the position error is greater than the normal lane widths. In this article, we examine these problems and extend a crowdsourced machine in order to identify and locate the use of accelerometer data from integrated vehicle sensors in multi-lane environments. The crowdsourced method decreases the bandwidth required by calculating road inclination statistics and the financial institution's attitude in each car to filter acceleration additives that do not match potholes. I test our method with simulated and real-world statistics, analyse engine trade-offs and the bandwidth needed to identify them correctly and look at the impact of less complex unmarried lane detection.

II. REVIEW OF LITERATURE

First: Real-time monitoring[1] of real-world Tweet accidents, social networking and microblogging deals, such as Twitter, would be a valuable supply of data on current activities. Twitter's widespread use of mobile devices and personal PC systems allows users to receive a percentage of short messages for any real-time task, which makes it ideal for early identification of unexpected events where quick responses are essential. In this article, we present a web technique to detect current foreign events on Twitter by means of Twitter content, whether herbal disasters or man-made disasters. Our solution blends Textual and frequency components of one-of-one form which constitute or reflect interesting semantical elements of a case. We use visualisation to detect which components are important to our event prediction algorithms and the impact the parameters are having on the results.

EvenTweet: Twitter[2] Localized Event detection, Twitter, Facebook, and Four-Square offerings, along with microblogging, have become the main assets for data about actual incidents around the world. Typically the temporal meaning of messages is used by most approaches that use information from those resources. However, it is important for localised events, which consist of public openings or crises, to use the place details of georeferenced messages too. Users that post messages close to an incident area act as human sensors to describe an occurrence. In this demonstration, we offer a special context for stumbling over placed incidents on Twitter in real time and songing about how these activities grow over the years. Space-temporal properties of for this reason Significant candidates for periodic descriptions are often derived keywords. Then, clustered events are extracted in line with their spatial similarities by using clusters of keywords. We implement a ranking system for occasions in order to determine the most relevant incidents in a (current) time period. Our Also Tweet gadget is seen during the 2012 UEFA European Football Championship by using a circulation of tweet from Europe.

We also shown[3] that Twitcident, a fully-fledged framework for filtering, looking and reading facts approx. real-life events and emergencies, is battling fire with information from social media sources. Twitcident. Twitcident links to emergency broadcast deals and mechanically continues to monitor and process social web feeds (Twitter) records when a new event happens. It enriches the semantics of streaming Twitter messages to profile events and to continuously decorate and adapt the filtering of information to the contemporary temporal context. Facetted search and analytical gadgets enable users to obtain accurate data, analyse and review state-of-the-art business as set out on the Social Web.

Twitter Case Tracking[4], Twitter is growing increasingly in current years as the medium of social media. People use Twitter to report incidents in their daily lives. The aim of this paper is to detect these events by reading the textual information stream on Twitter. Event identification is a long subject of research, but Twitter's characteristics turn it into an incomprehensible undertaking. These tweets are commonly battered by hifood niess." Furthermore, considering the sheer number of tweets, occasional identification of regulations would like to be scalable. With EDCoW, this paper aims to resolve the challenges (Event Detection with Clustering of Wavelet-based Signals). By using a wavelet analysis on the frequency dependent uncooked signals of a sentence, EDCoW creates indicators for personal terms. It then removes the trivial sentences by looking for their related associations with sign vehicles. The final words are eventually divided into operations focused largely on a methodology of modular graph partitioning. Experimental findings indicate positive EDCoW end results.

I see a car crash: Real-time detection[5] for small scale incidents in microblogs, as a source of records of urgency, microblogs are rapidly receiving coverage. However, because of the sheer amount of unstructured statistics, it is still impossible to reuse this supply of information during emergency situations. Smallest small pieces of information are present particularly for the detection of small-scale events, such as car accidents, which make the detection of relevant data challenging. We have a response to minor accidents in real time Microblogs allow to improve the emphasis of situation through the gathering of additional

incident information. Our approach is a compilation of rules that combine text classes with semantic microblog enrichments. An appraisal fully reveals that, after detecting accidents that are posted in actual time, our response allows the identification of small scale incidents with an accuracy of 89 percent Linked Open Government Information.

Twitter Stream Analysis real-time traffic[6] detection Social networks is Recently used for incident detection, with precise regard to traffic delays and traffic crashes as a source of intelligence. In this text, we offer a Twitter circulate evaluation in real time traffic management machine. The device collects Twitter tweets according to various search requirements; tweets tactics, by using textual information mining strategies; and tweets on the long-term. It is the intention, whether you are trafficking or not, to give any tweet the ideal class mark. In the light of visitors' occasions almost on real-time detection, the traffic detection system was updated to track various Italian street regions periodically before online visitors visit knowledge websites. In order to resolve binary classification problems, we have employed the help vector gadget as a class, and carried out an accuracy cost of 95

Potholes detection based on SVM[7] in the distress picture is an architecture scheme with upcoming Wi-Fi technology, which guides the passenger by prior notice ignoring the potholes on the road Facts. Information. There are several access points on the side of the road that can be collected via Wi-Fi and transportation vehicles enter the area blanked by the impact of access points. The software is designed to alert the driver to a sound stimulus in a combination of the car. Describe each individual vehicle with a special radio frequency (RFID) tag that prevents removal or damage by means of the metrology and visualisation of potholes with a microsoft Kinect sensor. When the RFID-tab exam is a part of the stolen car, it sends a report to the police room using GSM SIM. Furthermore, if an ambulance passes the intersection, the site visitors' dispatcher will be informed of this minor inexperience. This machine, which enables wireless access to the factor to gather data about potholes, bypasses the use of a wi-fi broadcasting system by using these statistics. Traffic control has an accelerator that can also calculate both vertical and horizontal acceleration to capture environmental facts. It is used entirely for twisting the destiny detector.

The suggested paper[8] definition of potholes is explained by the implementation of a smart traffic management scheme for congestion control. Potholes are summarised by position, length, type and depth. There are multiple literatures to provide us with multiple potholes detection techniques and better standard of inspection and pavement with prior exploration and immediate action. For a given technique a class principles must be established to help a pothole repair decision-making method. The following:The traffic control device is shown. The job is finalised with a schematic. It has designed the database because of its reliance on smartphone applications,Interface and classification of details based on the legislation of Indonesia .

III. PROPOSED METHODOLOGY

To separated the tweets we recommended an adaptive keyword solution that would be close to semanticist by utilising the semanticist meaning in a dense term that includes word embedding edition. The proposed technology overcomes the failure to segregate and distinguish conventional methods, i.e. the mostly keyword based on algorithm. This business has a strategy for pushing publicly available Tweets slowly, pre-method and simple. These tweets are then analysed with profound mastering and machine learning techniques in order to extract non-recurring data for events.

A. Architecture

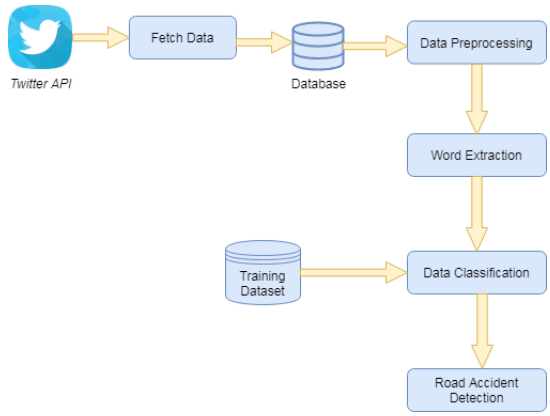

Fig. 1. Proposed System Architecture

B. Algorithm

Preprocessing Algorithm

Data preprocessing removes redundancy and ambiguity inherit in the data and transforms the reviews into sentences to facilitate sentence-level aspect-based classification. First, sentences are extracted by identifying the delimiters (e.g. dot, exclamation or question mark). Next, redundant information, e.g. duplicate sentences, is removed. Finally, ambiguous, vague or misspelled terms are corrected.

1. Stop word Removal-This technique removes stop words like is, are,they,but etc.

Initialize i,j
for i=1 to no of words in documents
for j=1 no of words in stopword list
if
Words(i)==Stopwords(j)
then
eliminate words(i)
end if

end for

2. Tokenization-This technique removes Special character and images.

Initialize feature vector bg feature =[0,0..0] for token in text.tokenize() do
if token in dict then
token idx=getindex(dict,token)
bg feature[token idx]++
else
continue
end if
end for

3. Stemming– Removes suffix and prefix and Find Original words for e.g.- 1. played – play 2.Clustering - cluster

The word w
Input = Normalize(input)
if normalizeValidate(input)
then return input;
for each rule in rules do if input match with rule
then
Stem = ExtractStem(input,rules)
if not TestStemLength(Rule)
then
end for
return input

4. Random Forest Machine Learning Algorithm:
Step 1: Let the number of training cases be N, and the number of variables in the classifier be M.
Step 2: The number m of input variables to be used to determine the decision at a node of the tree; m should be much less than M.
Step 3: Choose a training set for this tree by choosing n times with replacement from all N available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes.
Step 4: For each node of the tree, randomly choose m variables on which to base the decision at that node. Calculate the best split based on these m variables in the training set.
Step 5: Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier).
For prediction a new sample is pushed down the tree. It is assigned the label of the training sample in the terminal node it ends up in. This procedure is iterated over all trees in the ensemble, and the average vote of all trees is reported as random forest prediction.

IV. Conclusion

This can deter injuries and can be used to locate problem areas with potholes early. The efficiency of the conventional transport system by intelligent transportation systems in countries with increasingly upgraded economic development and excellent technology is a major factor. Similarly this computer will be used to not ignore the aforementioned reality and then refresh the server database. This mechanism has developed a structure that defines incidents triggered by the social media platform's non-recurring occurrences. The suggested system would be divided up into five major components, like multi-source data collection (i.e., hashtags, manage and bounding box), preprocessing of data, the recognition of the like keywords that complement the categories, eliminating pragmatic uncertainty, and locations dependent on material, to recognise vulnerable places.

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 Real- World 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.