Journal of Engineering and Applied Sciences 14 (21): 8055-8060, 2019
ISSN: 1816-949X

A (Near) Real-Time Traffic Monitoring System using Social Media Analytics

Goboitshepo Ororiseng Leroke and Manoj Lall
Department of Computer Science, Tshwane University of Technology, Pretoria, South Africa
Lallm@tut.ac.za, 0027-83-262-5772

Abstract: Dealing with traffic congestions is an integral part of a city life. Many hours are spent in traffic congestion leading to great cost and time losses. Normally, traffic conditions are monitored by the government agencies using electronic sensors or CCTV cameras. Undoubtedly maintaining a large networks of sensors and cameras to monitor every street in a city is both impractical and very expensive. However, since, the evolution of social media in all its forms, including blogs, online forums, Facebook and Twitter it is possible to treat social media as a human sensor network. In this study, we present a more cost effective and near real-time traffic monitoring alternative, based on Twitter data analytics which not only reports on the current traffic congestion conditions but also on the reasons causing the traffic congestion. Knowing the cause of the traffic congestion is important as it gives an indication of the severity of the problem. We demonstrate the feasibility of our solution through the use of Twitter data obtained for the city of Pretoria, South Africa. From the data collected, the location and the potential traffic related topics, such as vehicle accident or road construction are extracted. Public sentiments are calculated using a lexicon dictionary based approach and visualized using open street map. This system is aimed at assisting commuters in making an informed decision on route selection.

Key words: Social media, traffic congestion, sentimental analysis, natural processing language, Twitter, Facebook

INTRODUCTION

The road transportation network is unquestionably one of the largest revolutions in this present century (Niaraki and Kim, 2009; Elsafoury, 2013 and Musakwa, 2014). It has made significant inroads in the daily commute of the population around the globe. In spite of the benefits of this revolution, the problems associated with its emergence, such as traffic congestion has the potential to impacts everely on a country’s economy. To deal with this problem, urban communities have adopted the use of hardware sensors, cameras and radars to monitor the flow of traffic (Gong *et al*., 2015). These tools have certain limitations they are costly to maintain can focus on limited area of the transportation network and are designed to collect a specific type of information like the count of vehicles (Elsafoury, 2013; Gong *et al*., 2015).

With the incorporation of Web 2.0 technologies, the social media platforms such as Facebook, Twitter and LinkedIn are able to provide location based information and have successfully shifted social media platform from just being a cyberspace to real places (Kurashima *et al*., 2010; Somwanshi *et al*., 2015 and Mahmood *et al*., 2017). Social media, a worldwide phenomenon that has become a part of people’s daily lives could be viewed as a virtual sensor network full of human-agents with high mobility, quick response time and greater flexibility. This view of the social media has led several researchers to make use of the contextual information available in the network to explore a variety of scenarios and application fields. For instance, social media have been used for product recommendations, news article, video clips and audio music (Esparza *et al*., 2012; Ho *et al*., 2012). Disaster and emergency response systems using Twitter has been explored by Hua and Guang (2016). Featherstone (2013) explored the use of Twitter data for the prediction of crime whereas Kurashima *et al*. (2010) analysed photographic evidence of locations posted on Twitter to identify tourist locations. A majority of these systems are based on extracting information from the Tweets and applying natural language processing techniques for sentiment classifications.

Twitter is a micro-blogging platform that was launched in 2006. Twitter allows users to write messages called Tweets which permits a maximum of 140 characters. Frequently, Tweets are used to express an individual’s views, thoughts and emotional state (Musakwa, 2014). Additionally, Tweets often have geospatial information such as the user’s location profile. Combining these two characteristics provides possibilities to mine patterns and

Corresponding Author: Manoj Lall, Department of Computer Science, Tshwane University of Technology, Pretoria, South Africa, Lallm@tut.ac.za, 0027-83-262-5772

phenomenon associated with what is trending at any given time and place or in establishing the positive or negative sentiment on a particular topics across the population (Gong *et al*., 2015; Kharche and Bijole, 2015; Mahmood *et al*., 2017 and Khatri, 2018). In comparison to other social media platforms such as Facebook and Blogs, Twitter give a more real time feed and can be updated very often due to limitations imposed on the number of characters allowed in a Tweet. Our research makes use of this characteristic of Twitter for the proposed near real-time traffic monitoring system.

Transport and traffic related research based on social media platform often use natural language processing techniques to identify related keywords and correlates them with traffic congestion events. Sentiment analysis pertaining to the related event is performed to determine the sentiment associated with the traffic related event. The confidence level of such a system is based on the timeliness of the information and the number of Tweets received in a particular time period (Kosala and Adi, 2012). The other frequently suggested methodology for traffic related research is the use of network traffic simulation. However, a major challenges with this approach is that the accuracy of data of such simulations is not guaranteed to reflect the true transportation issues of cities. For example, the day-to-day commuting patterns can fluctuate greatly for a given city, based on a variety of factors: real-time transport incidents, sporting events, flooding and bush fires. This study complements the work of other researchers by makes use of Twitter platform to minimise problems affecting daily lives of people due to traffic congestion in the city of Pretoria, South Africa. Besides calculating and displaying the sentiment of the road users related to the traffic congestion on a particular route, the causes of the traffic congestion is also displayed. Having both of these information is essential in making an informed decision with regard to the route to be taken and when to start the trip. For example, if the cause of the traffic congestion is due to a truck having lost its load Fig. 1. It would be advisable to delay the trip, if no alternate route is available.

A number of transport management system have since been introduced in South African to alleviate the challenge posed by with traffic congestions. Some of the popular techniques are outlined below.

Traffic pointsmen project: The traffic pointsmen project, a private company undertakingis a way of trying to help the community to alleviate traffic congestion resulting from incidences such as car accidents or traffic lights not working. In such situations, a pointsman is dispatched to the location to assist in improving the traffic flow. A major limitation of this initiative is the availability of the pointsmen and the timely availability of the information relating to the location experiencing the traffic congestion (Fig. 2).

Fig. 1: Traffic disrupted due to truck losing its load

(a)

(b)

Fig. 2: a, b) Traffic pointsmen at work

Google maps: Google maps are a popular tool to provide information about traffic conditions on the roads. The routes are colour coded to display the density of traffic at a particular moment in time Fig. 3. A limitation of the Google maps are that it does not give the reasons behind the traffic congestion. Recognizing the reasons behind the traffic congestions would give a commuter some indication of how long the current traffic situation is likely to persist and therefore being able to plan their commute.

Fig. 3: Google map Pretoria traffic updates

For example, traffic congestion due to a traffic light (traffic robot) being temporarily out of order is not going to last as long as it would if the problem is due to a truck having lost its load. In our proposed system, besides displaying the sentiment of the Tweets (colour coded), it also displays the reasons behind the present traffic condition at a particular point on the roads.

Radio and television updates: Radio and television channels disseminates information about traffic events focusing largely on the main roads of the metropolitans. This information is disseminated periodically, making it difficult for the user to get information at a particular moment in time. Another limitation of this technique is that it is largely limited to major cities of this country. Our approach, presented in this study is not confined to major cities alone but can easily be applied to even smaller towns.

Sensors: At present, a great number of traffic monitoring systems use sensors and other infrastructure that are costly, requiring high maintenance and restricted to few geospatial locations. Therefore, countless incidents and congestion events go unreported, on the street where there are no cameras or other sensors. Our approach outlined in this study provides a more cost-effective and geographically unconstrained mechanism for near real-time traffic monitoring system.

MATERIALS AND METHODS

Sentiment analysis, also referred to as opinion mining is a type of natural language processing for tracking sentiments, attitudes and emotions of the public about a certain topic. Fundamentally, sentiment analysis aims to understand the sentiments and opinions and distribute them into categories like positive, negative or neutral. The sentiment analysis tasks can be done at various levels of granularity, namely, word level, phrase or sentence level, document level and feature level (Kumar and Sebastian, 2012a, b). As Twitter allows its users to share short pieces of information, limited to 140 characters the word level granularity is most appropriate for our research. A survey through the literature supports that the methods for automatically annotating sentiment at the word level follows either the dictionary-based approaches or the corpus-based approaches. Additionally, to automate sentiment analysis, different approaches have been applied to predict the sentiments of words, expressions or documents. These include Natural Language Processing (NLP) and Machine Learning (ML) algorithms (Kumar and Sebastian, 2012a, b). In dictionary-based approach, firstly, the opinion word from review text are found which is followed by finding their synonyms and antonyms from dictionary. The dictionaries like WordNet, SentiWordNet and SenticNet may be incorporated for mapping and scoring. Corpus-based method helps to find opinion word in a context specific orientation. Beginning with a list of opinion word, the corpus-based approach finds other opinion word in a huge corpus. In machine learning techniques various classification methods like Support Vector Machine (SVM), Naive Bayes (NB) and Maximum Entropy (ME) are used for sentiment classification. Machine learning methods maintain two datasets, namely the training dataset and the testing data set (Hardeniya and Borikar, 2016). For the purpose of conducting sentiment analysis we have applied dictionary based approach in conjunction with NPL parsing. The tools used are WordNet and Natural Language Toolkit (NLTK) libraries with python 3 and TextBlob 0.7. TextBlob integrates NLTK's WordNet interface to make interaction with WordNet easier. The details of the approach taken is presented below.

System development methodology: Opinion words are the words used by people express their opinion (positive, negative or neutral). To access the opinion words, Twitter supports two programmatic Representational State Transfer (REST)-based interface: a streaming API and a search API. The former is used for Tweets that are pushed to the end user clients whilst the latter is used for requesting specific Tweets (Gong *et al*., 2015). For this research, search API were used. Twitter supporting specification of geospatial coordinate systems where harvested from the Tweets. The Tweet text were then processed to get the opinions expressed in the Tweet (e.g., bad traffic or slow moving traffic) and possible reasons for traffic congestions(e.g., car crash, truck lost its load). The system development flow diagram is depicted in Fig. 4.

Data retrieval and pre-processing: The Twitter streaming API was used to get access tothe Tweets. The streaming API encompasses of three endpoints that provide access to streaming content from one user, a group of users or all public users (Nguyen *et al*., 2016; Giridhar *et al*., 2015). For the purpose of this study, we used Tweepy to gain access to the Twitter APIs. In Tweepy, an instance of Tweepy. Stream establishes a streaming session and routes messages to StreamListener instance. The default StreamListener can classify most common Twitter messages and routes them to appropriately named methods. In addition to providing steaming sessions, Tweepy makes authenticating a user an uncomplicated process. The authentication process is handled by the Tweepy. AuthHandler class, various filters were set on

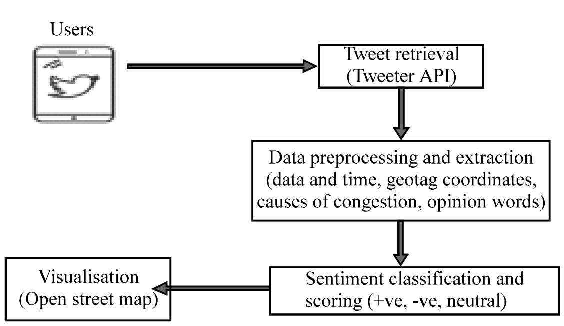

Fig. 4: System development flow diagram

these data streams to capture Tweets within a geographic area. Data for this study was filtered to contain keywords that relate to traffic congestion, such as, "accident", "crash", "traffic", "road", "freeway", "highway", "lane closed"," lane blocked", "wreck, "car", "truck"," protests", "strikes" and "marches".

RESULTS AND DISCUSSION

In addition to filtering Tweets based on traffic congestion related keywords, information related to geographical location of the origin of the Tweets are obtained. Essentially, a geotagged post consists of fields to identify the user who has posted a particular post a location identifier showing the location from where the post was sent, the time when the post was sent and the textual content of the post. Sentiment analysis may be viewed as the process of linking the text to the emotions (such as happiness, sadness, anger).

Sentiment analysis: In this research, sentimental analysis were conducted by the use of NLTK (a leading platform built for Python programs to work with human language data). NLTK offers easy-to-use interfaces to over 50 corpora and lexical resources like WordNet, including a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing and semantic reasoning. For classifying Tweets in the different class (positive, negative or neutral) we have used the TextBlob sentiments module. The TextBlob sentiment module consists of two sentiment analysis implementations the pattern analyser and the Naive Bayes analyser. Being the default implementation, we have used the pattern analyser. The Python code snipper shown in Fig. 5 demonstrate the application of TextBlob to a Tweet. The Tweet and the resulting sentiments are shown in Fig. 6. Figure 7, shows the sentiment of each Tweet received during a time interval of 1 h on a particular day from a particular location (Fig. 8).

Visualization: After classification, the traffic sentiment of the Tweets for a particular location, the Tweets are plotted on the street map. This step was achieved by using

```
all_data=json.loads(data)
tweet=all_data["text"].encode("utf-8")
tweet=" ".join(re.findall("[a-zA-Z]+",tweet.decode('utf-8')))
blob=TextBlob(tweet.strip())
location=all_data["coordinates"]
geo=all_data["geo"]
place=all_data["place"]
```

Fig. 5: Code snippet for determining the sentiment of a Tweet

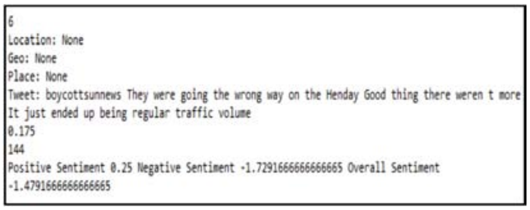

Fig. 6: Sentiment score of the Tweet

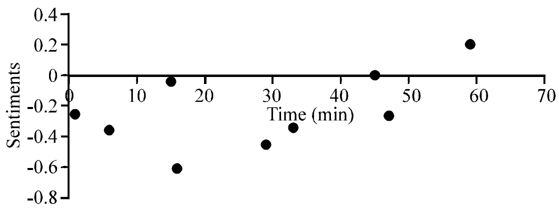

Fig. 7: Live sentimental analysis results

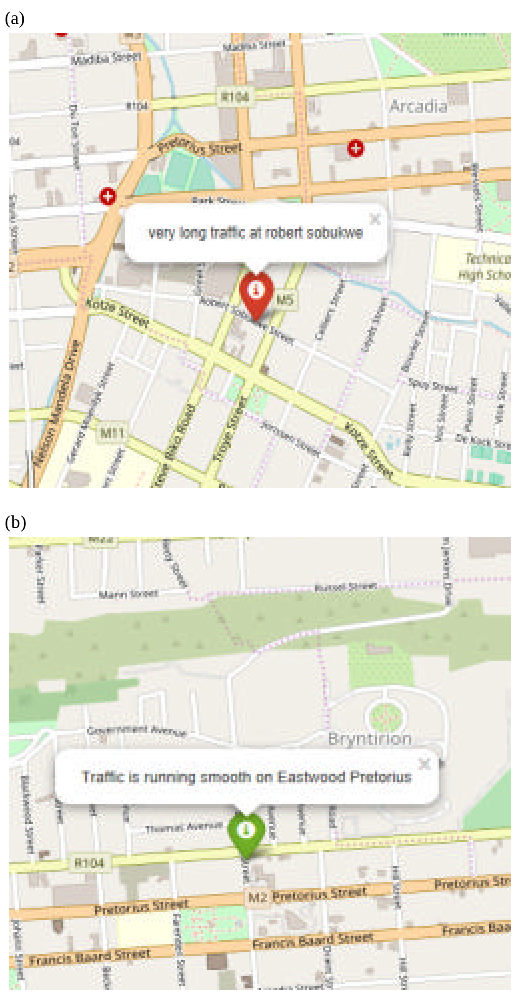

Fig. 8: Sentiments: a) Negative and b) Positive

Folium-a tool for visualizing data that has been manipulated in Python on an interactive leaflet map. It enables both the binding of data to a map for choropleth visualizations as well as passing rich vector/raster/HTML visualizations as markers on the map. Figure 8a, b display the sentiments of Tweets and possible causes of traffic congestion in the central business district of Pretoria, South Africa. The red info-icons marks the Tweets with a negative sentiment, green info-icons are for Tweets with positive sentiments.

CONCLUSION

This study has expounded on an innovative approach of using Twitter as a platform, for near real-time traffic monitoring system. Making use of this social networks platform, provides a rather inexpensive alternative to the numerous traffic monitoring mechanisms currently in place. Besides analysing the sentiments of the Tweets originating from a particular location and within a certain time period our system also looks for possible causes of congestion. Knowing the possible causes of congestion, assists a road user in making an informed decision regarding the persistent nature of the traffic condition. The focus of this study is related to traffic in and around the city of Tshwane in South Africa, however, the system and methodologies used are generic and could easily be applied to other cities in the country or to other countries in the world.

The proposed system relies heavily on the number of Tweets posted related to the traffic conditions. When this number is small, it makes the system unreliable. Such situations are likely to exist in smaller towns with low density of Twitter users. As an extension to this research, we intend to improve on the harvesting of Tweet text for more elaborate traffic congestion causes and validation of our results to the reality on the ground.

Significance statement: In this study, we present a more cost effective and near real-time traffic monitoring alternative, based on Twitter data analytics which not only reports on the current traffic congestion conditions but also on the reasons causing the traffic congestion. Knowing the cause of the traffic congestion is important as it gives an indication of the severity of the problem.

ACKNOWLEDGEMENT

The researchers would like to thank the National Research Foundation, South Africa for providing the necessary funding (Grant number: CPT160509164502).
ORCID ID: 0000- 0001- 9507-6566.

REFERENCES

Elsafoury, F.A., 2013. Monitoring urban traffic status using Twitter messages. MSc Thesis, University of Twente, Enschede, Netherlands.

Esparza, S.G., M.P. O'Mahony and B. Smyth, 2012. Mining the real-time web: A novel approach to product recommendation. Knowl. Based Syst., 29: 3-11.

Featherstone, C., 2013. The relevance of social media as it applies in South Africa to crime prediction. Proceedings of the 2013 IEEE International Conference and Exhibition on IST-Africa, May 29-31, 2013, IEEE, Nairobi, Kenya, ISBN:978-1-905824-38-0, pp: 1-7.

Giridhar, P., T. Abdelzaher, J. George and L. Kaplan, 2015. Event localization and visualization in social networks. Proceedings of the 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), April 26-May 1, 2015, IEEE, Hong Kong, China, ISBN:978-1-4673-7131-5, pp: 35-36.

Gong, Y., F. Deng and R.O. Sinnott, 2015. Identification of (near) real-time traffic congestion in the cities of Australia through Twitter. Proceedings of the ACM 1st International Workshop on Understanding the City with Urban Informatics (UCUI '15), October 19, 2015, ACM, Melbourne, Australia, ISBN:978-1-4503-3786-1, pp: 7-12.

Hardeniya, T. and D.A. Borikar, 2016. An approach to sentiment analysis using lexicons with comparative analysis of different techniques. IOSR. J. Comput. Eng., 18: 53-57.

Ho, S.S., M. Lieberman, P. Wang and H. Samet, 2012. Mining future spatiotemporal events and their sentiment from online news articles for location-aware recommendation system. Proceedings of the 1st ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems (MobiGIS '12), November 06, 2012, ACM, Redondo Beach, California, ISBN:978-1-4503-1699-6, pp: 25-32.

Hua, B. And Y. Guang, 2016. A weibo-based approach to disaster informatics: Incidents discovery in post-disaster situation via weibo text negative sentiment analysis. Nat. Hazards, 83: 1177-1196.

Kharche, S.R. and L. Bijole, 2015. Review on sentiment analysis of Twitter data. Intl. J. Comput. Sci. Appl., 8: 53-56.

Khatri, C.P., 2018. Real-time road traffic information detection through social media. MCS Thesis, Georgia Institute of Technology, Atlanta, Georgia, USA.

Kosala, R. and E. Adi, 2012. Harvesting real time traffic information from Twitter. Procedia Eng., 50: 1-11.

Kumar, A. and T.M. Sebastian, 2012b. Machine learning assisted sentiment analysis. Proceedings of the International Conference on Computer Science and Engineering (ICCSE 2012), July 4-6, 2012, London, UK., pp: 123-130.

Kumar, A. and T.M. Sebastian, 2012a. Sentiment analysis: A perspective on its past, present and future. Intl. J. Intell. Syst. Appl., 4: 1-14.

Kurashima, T., T. Iwata, G. Irie and K. Fujimura, 2010. Travel route recommendation using geotags in photo sharing sites. Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10), October 26-30, 2010, ACM, Toronto, Ontario, Canada, ISBN:978-1-4503-0099-5, pp: 579-588.

Mahmood, T., G. Mujtaba, L. Shuib, N.Z. Ali, A. Bawa and S. Karim, 2017. Public bus commuter assistance through the named entity recognition of twitter feeds and intelligent route finding. IET Intell. Transp. Syst., 11: 521-529.

Musakwa, W., 2014. The use of social media in public transit systems: The case of the Gautrain, Gauteng province, South Africa: Analysis and lessons learnt. Proceedings of the 2014 REAL CORP 19th International Conference on Urban Planning, Regional Development and Information Society, May 21-23, 2014, Vienna, Austria, ISBN:978-3-9503110-7-5, pp: 721-727.

Nguyen, H., W. Liu, P. Rivera and F. Chen, 2016. Trafficwatch: Real-time traffic incident detection and monitoring using social media. Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2016), April 19-22, 2016, Springer, Cham, Switzerland, ISBN:978-3-319-31752-6, pp: 540-551.

Niaraki, A.S. and K. Kim, 2009. Ontology based personalized route planning system using a multi-criteria decision making approach. Expert Syst. Appl., 36: 2250-2259.

Somwanshi, Y., V. Salegaonkar and S. Sharma, 2015. Understanding social media phenomenon, diversity and research. Intl. J. Comput. Appl., 129: 5-8.