



# Automatic Event Detection in Smart Cities Using Big Data Analytics

Sugimiyanto Suma<sup>1</sup>(✉), Rashid Mehmood<sup>2</sup>, and Aiiad Albeshri<sup>1</sup>

<sup>1</sup> Department of Computer Science, FCIT,

King Abdulaziz University, Jeddah 21589, Saudi Arabia

sugimiyanto@gmail.com, aaalbeshri@kau.edu.sa

<sup>2</sup> High Performance Computing Center, King Abdulaziz University,

Jeddah 21589, Saudi Arabia

RMehmood@kau.edu.sa

**Abstract.** Big data technologies enable smart city systems in sensing the city at micro-levels, making intelligent decisions, and taking appropriate actions, all within stringent time bounds. Social media have revolutionized our societies and is gradually becoming a key pulse of smart societies by sensing the information about the people and their spatio-temporal experiences around the living spaces. In this paper, we use Twitter for the detection of spatio-temporal events in London. Specifically, we use big data and machine learning platforms including Spark, and Tableau, to study twitter data about London. Moreover, we use the Google Maps Geocoding API to locate the tweeters and make additional analysis. We find and locate congestion around London and empirically demonstrate that events can be detected automatically by analyzing data. We detect the occurrence of multiple events including the London Notting Hill Carnival 2017 event, both their locations and times, without any prior knowledge of the event. The results presented in the paper have been obtained by analyzing over three million tweets.

**Keywords:** Smart cities · Big data · High performance computing  
Social media analysis · Machine learning

## 1 Introduction

Nowadays, having gadgets has become a necessity. Every electronic device is connected to a worldwide network called internet. This large network connection enables us to manage those connected devices. In another hand, the ability of managing the connected devices, and the flowing data are essential to support smart cities. Smart city application helps us to make a better life. It changes the way we work and overcome the problem of urban life, supported by various emerging technologies. Big data technologies would play a key role in support smart city systems and applications due to the need to sense the city at the micro-levels, make intelligent decisions, and take appropriate actions, all within stringent time bounds. Social media has revolutionized our societies and is gradually becoming a key pulse of smart societies by sensing the information about the people and their spatio-temporal experiences around the living

spaces. One of the useful application in smart city is event detection, sub-part of smart societies. For stakeholders, detecting an occurring event is important in finding out what is happening in the city for decision-making or future planning purposes. Compared with sensor-based event detection, analyzing social media data such as twitter is more cost-effective way to detect events. Sensor-based detection mines traffic data of installed sensors and cameras in certain places. It is costly, for hardware procurement and network installation, among other things. In addition, the number of installed measurement instruments limits the detection coverage. While social media event detection has wider coverage, and more efficient in terms of resources. They both have their pros and cons and could complement each other in terms of the convenience of event detection and information coverage.

In this paper, we use twitter for the detection of spatio-temporal events in London. Specifically, we use big data and AI platforms including Spark, and Tableau, to study twitter data about London. We extend our earlier work [1] of using social media data to detect spatio-temporal events in London. In this paper, we use better data analytics by implementing machine learning for contextual analysis awareness using Apache Spark MLlib. The acquired data should be truly related to traffic problem, while the acquired data is not necessary relevant to traffic problem. As well as, integrating the process with HPC to have better analysis performance. We use apache spark for parallel data processing, which is installed on top of HPC cluster. As well as, we utilize FEFS system for high-speed data distribution system. Moreover, we use the Google Maps Geocoding API to locate the tweeters and make additional analysis.

We find and locate congestion around the London city. We also empirically demonstrate that events can be detected automatically by analyzing data. We detect the occurrence of multiple events such as “Underbelly festival” and “The Luna Cinema”. Underbelly festival was located at south bank, while The Luna Cinema was located in some places such as around Greenwich Park, crystal Palace Park, and National Trust-Morden Hall Park. As well as, we detect the London Notting Hill Carnival 2017 event. This is located around Notting Hill as it was the location of Notting Hill carnival [2], the Europe’s biggest street festival which was organized by London Notting Hill carnival enterprises trust. We detect those event’s locations and times, without any prior knowledge of the event. The results presented in the paper have been obtained by analyzing over three million tweets.

We summarize our contributions in this paper as follows:

- We design workflow of big data analytics to detect spatio-temporal events in London using Apache Spark.
- We detect the occurring events automatically by analyzing twitter data without any prior knowledge of events, both its location and time.
- We integrate big data processing with HPC technology to improve performance.

While researcher have studied social media based event detection in the recent past, the use of Apache Spark for social media based event detection has not been found in the literature. The specific data, its analysis, and event detection presented in this paper also make the contributions of this paper unique.

This paper is structured as follows. Literature review is given in Sect. 2. The used design and methodology is explained in Sect. 3. The discussion of our research results

and analysis are described in Sect. 4. Finally, conclusion and future works are discussed in Sect. 5.

## 2 Literature Review

Research for smart cities using big data and social media analysis is becoming increasingly important. Khan et al. developed a prototype analytics as a cloud service, for managing and analyzing big data in smart cities [3]. Herrera-Quintero et al. combined big data and IoT to support transportation planning system for Bus Rapid Transit (BRT) systems [4]. Kolchyna et al. predicted spikes in sales by detecting twitter events of 150 million tweets [5]. Arfat et al. proposed an architecture for smart city as a mobile computing system with big data technologies, fogs, and clouds. In order to enable smarter cities with enhanced mobility information [6]. Other related works in smart cities are exists, such as virtual-reality-based traffic event simulations [7], location based services [8], urban logistics [9–12], and smart emergency management system [13, 14].

In another hand, for spatio-temporal event detection purpose, exploiting social media data is a cost-effective way compared with traditional method using installed sensors and cameras. There have been number of works analyzing social media data for detecting event. Gu et al. developed a real-time detector of traffic incident with five categories, including occurring events. It applies semi-naïve-Bayes classification [15]. Nguyen and Jung, proposed an approach for early event identification, by combining content-based features from the social text data and the propagation of news between viewers [16]. Unankard et al. identified strong correlations between user location and event location to detect emerging hotspot events [17].

Wang combined visual sensor (cameras) with social sensor (twitter feeds) to detect events. Images processing is applied to detect abnormal patterns indicating occurring events. Next, using social data information to derive the high-level semantic [18]. Kaleel and Abhari, proposed an algorithm to detect interesting events by matching its keywords on cluster labels of tweet (clustering). Subsequently, trend it based on time, geo-locations, and cluster size [19].

There are several event detection works related to road traffic [20–22], however, they do not use big data technologies and different analysis technique. Moreover, our work integrates big data technologies and HPC. In substance, this is an extension of our previous work [1]. In this work, we carried out a deeper analysis such as overcoming the limitation of contextual analysis of the status message, in order to ensure that the acquired data is truly related to traffic problem caused by occurring events.

## 3 Methodology and Design

We developed a system architecture to detect spatio-temporal events as shown in Fig. 1. First, we crawl the status message (twitter) according to a predefined keyword set and a set of social media user accounts, which is relevant to traffic. Thereafter, we store the crawled data into data pool. Secondly, we preprocess the acquired raw data before going

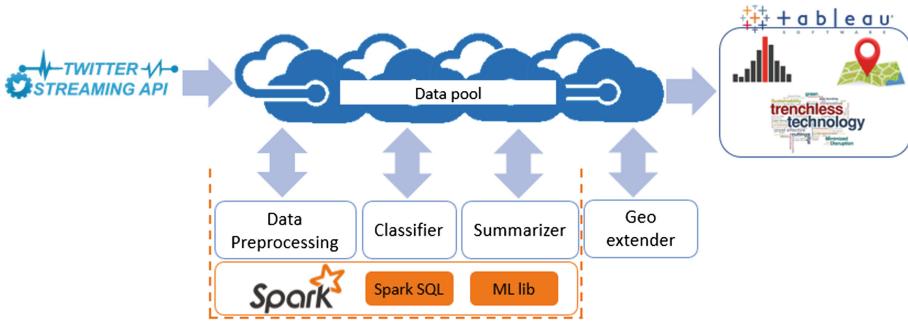


Fig. 1. Workflow of spatio-temporal events detection

to classification learning, where-upon social media data has lots of noises. It is not standardized, and there are plenty of unnecessary characters and words. Third, the status messages are classified into either traffic-related or non-traffic-related message by utilizing a supervised machine learning system. Forth, the classified status messages are extended to get more location information. Finally, the traffic-related status message with spatio-temporal information is visualized by using a map visualization.

We use apache spark platform to do heavy computation with huge data. Since spark is an in-memory computation platform, spark has better speed up to process big data in parallel, compared with other parallel data processing such as Hadoop map reduce [23]. We use spark for data processing and classifier stages. For data pool, where all machine processors take the acquired data for further processing, we utilize the power of Fujitsu Exabyte File System (FEFS). It is a parallel file storage system technology. FEFS is a software for HPC cluster systems, developed by Fujitsu Ltd. It enables high-speed parallel distributed processing of huge amounts of transactions [24]. As well as, it has superior features such as actual operational convenience, system scalability, and high reliability for zero operational downtime during a long computation. Thus, it contributes to significant improvements in system performance. Those FEFS and spark technologies are installed on top of HPC cluster.

### 3.1 Data Acquisition

We use social media data source (twitter) related to traffic. It is done by defining a set of keywords and a set of twitter user accounts which tend to post messages relevant to traffic such as government and media user accounts. Data crawling is performed by invoking twitter streaming API through a java based crawler application. The acquired data subsequently be stored in a data pool as raw data in FEFS system which has been described in Sect. 3. Any further data processing will pull raw data from this pool.

#### 3.1.1 Dataset Structure

The acquired data is in raw JavaScript object notation (JSON) as a Twitter data format. It is stored in a file system as JSON file extension. In raw format, each status message contains a bunch of attributes. For our experiment purposes, we use several selected

**Table 1.** Raw status message data structure

Attribute	Length	Data type
Created_at	30	Time stamp
Latitude		Double
Longitude		Double
Text	140	String

required attributes for spatio-temporal event detection. The structure of raw and extended status message is shown in Tables 1 and 2 respectively.

The illustration of selected fields of raw twitter JSON data is shown in Fig. 2. It is delimited by “|” character for each attribute.

Wed Aug 21 16:54:04 +0000 2017		51.5131392		-0.1397757		Having a lovely day sitting in grid lock traffic...
Wed Aug 22 08:31:09 +0000 2017		51.4943053		-0.1023826		Traffic at this time. that's London for you...
Wed Aug 23 09:01:54 +0000 2017		null		null		stuck in London bcause it outage i called u following...
Wed Aug 23 10:15:18 +0000 2017		null		null		Victoria park hackney hackneytower hamlets london...
Wed Aug 23 10:29:07 +0000 2017		51.3898872		0.0421551		Southborough, a26 London road slow moving traffic...
Wed Aug 24 15:22:41 +0000 2017		null		null		on way home from london, been delayed in traffic...
Wed Aug 25 10:29:27 +0000 2017		51.5083869		-0.1140182		Actually hate london traffic it's taken 30 mins to go...
Wed Aug 26 12:45:32 +0000 2017		51.4623707		-0.5004316		m25 delays near j14/a3113 anticlockwise caused...
Wed Aug 28 05:09:11 +0000 2017		null		null		fun fact london underground's logo was inspired by what is on...
Wed Aug 28 17:00:51 +0000 2017		null		null		did you know the london chatham and dover railway opered...

**Fig. 2.** Illustration of raw twitter JSON data

After the data processing, classification, and geo-extender function are applied, it extends more additional attributes for spatio-temporal purposes, in order to easily plot status message's location on map visualization, as shown in Table 2.

**Table 2.** Extended status message data structure

Attribute	Length	Data type
Created_at	30	Time stamp
Latitude		Double
Longitude		Double
Text	140	String
Postal_code	8	String
Type	100	String

Each attribute is defined as follows:

- Created\_at: the time when the status message is posted by user (timestamp)
- Latitude, Longitude: geolocation of status message

- Text: the message content posted by user
- Postal\_code: the postal/zip code of status message, e.g., “SE6.”
- Type: the location type of detected road name from text attribute, e.g. “route,” “point\_of\_interest.”

The illustration of data after applying data processing, classification and geo-extender function is shown in Fig. 3. It is delimited by “|” character for each attribute.

2017-08-21-16	51.5131392	-0.1397757	Having a lovely day sitting in grid lock traffic...	W1B	route
2017-08-22-08	51.4943053	-0.1023826	Traffic at this time. that's London for you...	SE11	route
2017-08-23-10	51.3898872	0.0421551	Southborough, a26 London road slow moving traffic...	BR2	route
2017-08-25-10	51.5088869	-0.1140182	Actually hate london traffic it's taken 30 mins to go...	SE1	route
2017-08-26-12	51.4629707	-0.5004316	m25 delays near j14/a3113 anticlockwise caused...	TW19	route

**Fig. 3.** Illustration of processed data after applying data processing, classification, and geo-extender function

### 3.2 Data Preprocessing

Data preprocessing is the first action against the acquired data in the data pool. Since it has a significant impact on accuracy and quality of learning the data by machine, hence it is an essential stage in big data analytics workflow [1]. In fact, social media status text contains lots of noise. It has plenty of unnecessary characters and words such as URL, user mention, illegal character, e.g. ‘&’, punctuation, and stop word. Therefore, the raw data should be preprocessed to clean those up from outliers and make it standard. We utilize spark SQL and regular expression function to preprocess the data, by referring to our defined stop word dictionary, which is adopted from stop word list website [25–27]. Data preprocessing also includes data extraction and parsing such as ‘created\_at’ field as the date time posting to get the formatted date time, and ‘coordinates’ field as location precision of status message to get the spatio-temporal information. Furthermore, the preprocessed data is used to feed supervised machine learning for classification. In another hand, we ignore retweet (repost of another user’s post) status message, because, it contains the same information. Thus, it leads to efficient processing. The result of data processing is stored back in data pool as cleaned data.

### 3.3 Classifier and Summarizer

After the status messages are clean and standard, we separate them into two categories, either they are traffic-related or non-traffic-related. At which point, the acquired data is not necessary has a meaning of traffic problem caused by occurring events, although it contains our defined keywords related to traffic. It could be a news about an accident which affected the traffic. For example, “traffic in linbro park, there has been an accident on the n3 south at the London road exit”, or someone’s position like “I am at London underground,” or even absolutely not related to traffic such as “wanted man appeal issued human trafficking inquiry.” We build a classification model by using logistic

regression with stochastic gradient descent, which is available in spark MLlib. First of all, from the data pool, we take sample data to create training data to train the model. It is done by examining the status messages manually to determine its categories. Then labeling each status message by either 1 or 0. Label 1 denotes that it is traffic-related, and label 0 for non-traffic-related. Secondly, we train our model by using the training data. Thirdly, using the trained model, we classify each status message into two categories (0 and 1) iteratively. Finally, we filter out the status messages, which are not related to traffic for further processing. Furthermore, we summarize the data to get the insight by applying several data summaries such as generating buzzword (top mentioned words), counting an hourly number of tweet, and information for map plotting.

### 3.4 Geo-Extender

In order to plot status message's distribution in the form of map, we need to get a specific point/area location of traffic-related status messages in the form of earth cartesian coordinate (latitude, longitude). Afterwards, it will ease the user for analysis processes. This phase is done by passing latitude and longitude fields of status messages into google geocode API to get more location information such as postal code, road name, city, etc. Notably, our work uses a postal code in order to plot the location distribution of traffic-related tweet or relevant topic in the form of a map view.

### 3.5 Analysis Visualization

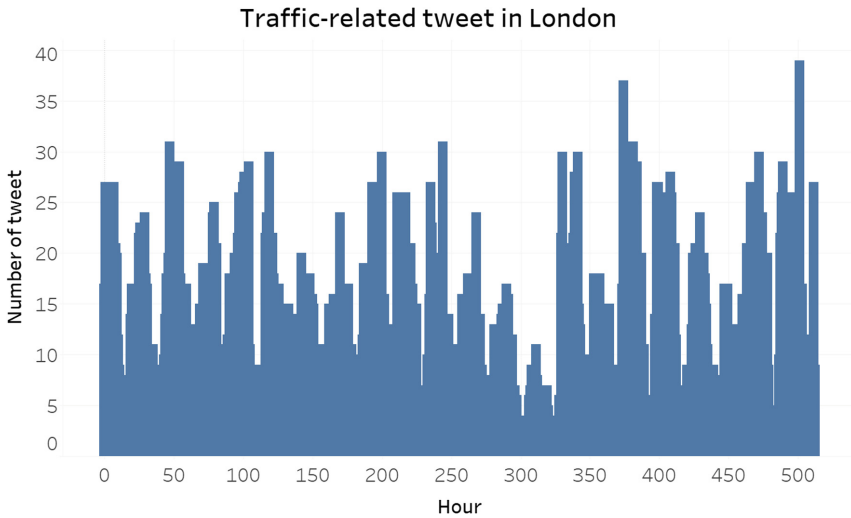
There are many ways to plot geolocation data into a map visualization for analysis phase. One of them is by using Tableau software. Tableau is a business intelligence software which helps people to see and have a better understanding of their data [28]. It enables users to explore their data with limitless visual analytics. As well as, it eases user to perform ad-hoc analysis with just a few clicks. We take the processed and geolocated data as a data source of Tableau; then we generate several visualization summaries such as a graph, word cloud, and map. With a graph, we can see the number of tweets with the time-frequency distribution in order to see the anomaly which indicates more traffic than usual. With word cloud, we can see the most mentioned words which imply the hot topic. With a map, we can depict the dissemination of traffic condition on a particular area.

By applying supervised machine learning which aims to ensure that the status message data is genuinely related to a traffic problem, we overcome the limitation of our previous work which did not consider a contextual analysis of actual traffic related problem caused by occurring events.

## 4 Result and Discussion

For experiment purpose, we gathered twitter data between 21st August–13th September 2017, represented as hourly data in the range (0, 576). In total, it consists of 3 million records of tweet related to our defined keyword. However, after classification process which aims to filter out non-related traffic tweet, the number of tweets are

decreasing. The decrement in the tweet count is reasonable, as these status messages are not genuinely related to traffic. Even though it contains a traffic-related word, it does not mean a road traffic problem. Figure 4 shows that the time-frequency distribution of the number of tweets is varied hourly. The peak hours are around 375th and 500th hours. By looking at this, we can assume that there was an ongoing event occurring at that time which affected the traffic condition in London.

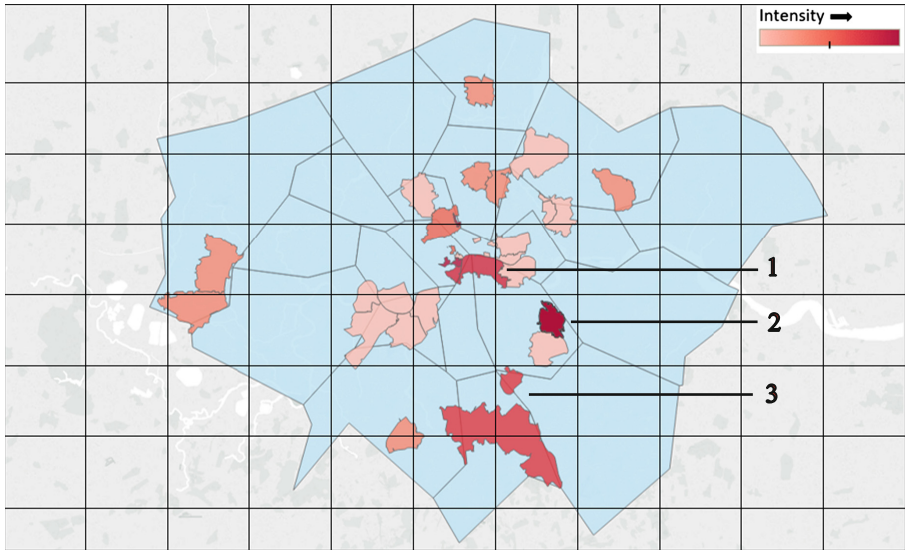


**Fig. 4.** Hourly number of tweet related to traffic in London

The location-intensity distribution of the number of tweets related to traffic in London is shown in Fig. 5. Red color level varies the number of distribution. The higher the intensity of red, the more traffic an area has. The figure shows that the areas around downtown had more traffic. We assume that an increase in a traffic-related tweet points to increased traffic. Grey color denotes the areas where we could not get geotagged tweets since not all acquired tweets are geotagged. There are three main areas with high-intensity red on the map. Those are in South Bank (shown as 1), around Greenwich park (shown as 2), around crystal palace park and National Trust-Morden Hall Park (shown as 3), in Fig. 5. The postal codes are SE1, SE10, SE19, and SM4, respectively.

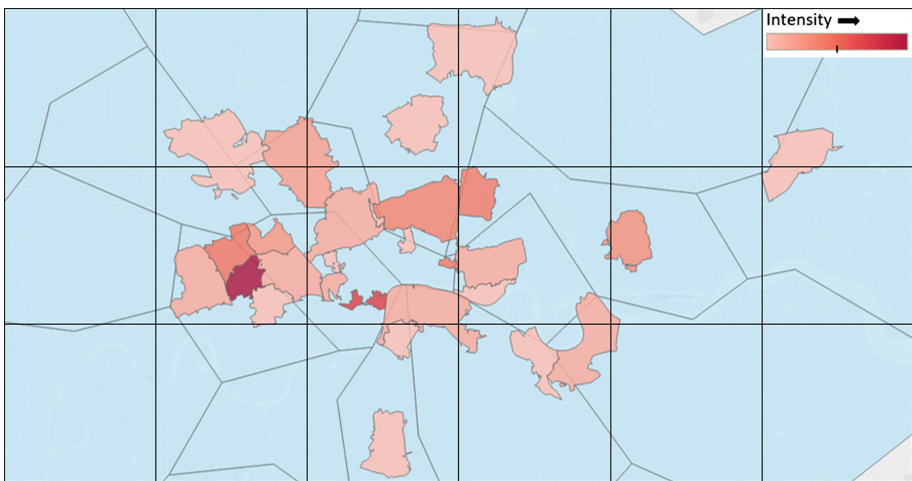
According to London events calendar [29], in south bank, there has been an event held, called “underbelly festival.” It has started from 28th April to 30th September 2017. It was a festival event held by Underbelly, which showing off cabaret, comedy, live circus, and family entertainment [30]. In another hand, around Greenwich Park, Crystal Palace Park, and National Trust-Morden Hall Park, there have been events occurred, called “The Luna Cinema,” running from June to October 2017. It is an outdoor cinema for citizens or tourist to spend their warm summer evening watching a new film release or age-old classic.



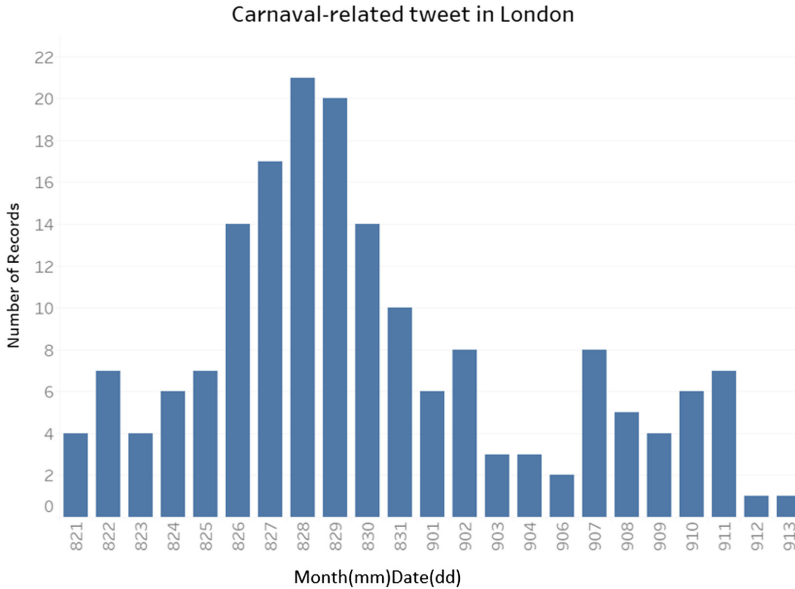


**Fig. 5.** Tweet intensity related to traffic in London (Color figure online)

Moreover, we analyze the hot topic among the whole tweets. Word “carnival” was among the top five most mentioned word. The location spreading of tweet related to carnival takes place is shown in Fig. 6. Most of the tweet comes from around one area, as shown in red color in the figure. By inspecting this circumstance, we can infer that there was an event related to a carnival on that area. This red colour area lies around Notting Hill London, and was the location of Notting Hill carnival [2], the Europe’s biggest street festival which is organized by London Notting Hill carnival enterprises trust.



**Fig. 6.** Tweet intensity related to carnival in London (Color figure online)



**Fig. 7.** Number of tweet in the period related to carnival in London

The daily frequency distribution of the number of tweets about carnival is shown in Fig. 7. We can see that the peak is on 28<sup>th</sup> August, while it is increasing gradually from 26<sup>th</sup> August and decreasing until 30<sup>th</sup> August. By observing this phenomenon, we can conclude that around those dates, there was an event related to a carnival. The carnival event was held on 26<sup>th</sup>–28<sup>th</sup> August 2017 in London [2].

This information about the event was detected automatically without any prior knowledge of the event, its location and time.

## 5 Conclusion

Social media has revolutionized our societies and is gradually becoming a key pulse of smart societies by sensing the information about the people and their spatio-temporal experiences around the living spaces. Analyzing social media data such as twitter has become a cost-effective way to detect events, by utilizing big data technologies such as spark, FEFS, and tableau. In this paper, we use Twitter for the detection of spatio-temporal events in London. Specifically, we use big data and AI platforms including Spark, and Tableau, to study twitter data about London. It is done by mining twitter data related to traffic problem which indicates occurring events. We have better data processing by implementing machine learning for contextual analysis awareness using spark MLlib. While the acquired data should be truly related to traffic problem.

We empirically demonstrate that events can be detected automatically by analyzing data. We detect the occurrence of multiple events such as “Underbelly festival” and “The Luna Cinema”. Underbelly festival was located at south bank, while The Luna

Cinema was located in some places such as around Greenwich Park, Crystal Palace Park, and National Trust-Morden Hall Park. As well as, we detect the London Notting Hill Carnival 2017 event. This is located around Notting Hill as it was the location of Notting Hill carnival [2], the Europe's biggest street festival which was organized by London Notting Hill carnival enterprises trust. We detect those event's locations and times, without any prior knowledge of the event. The results presented in the paper have been obtained by analyzing over three million tweets. We overcome the future work of our prior work, which did not consider textual analysis to ensure that tweets are truly related to traffic problem. As well as, integrating big data technologies with HPC to enhance scalability and computational intelligence.

Although we have improved the data management and processing. It still needs improvement to have better detection accuracy, wider spatio-temporal detection, and better quality of analysis. For better detection accuracy, we plan to develop an algorithm and compared the result with actual information by associating it with events reporting such as news or media websites. For wider detection, we would acquire more social media data such as Facebook. For better quality of analysis, we hope to utilize more AI techniques. Hence, those will be the future work of our research.

**Acknowledgments.** The authors acknowledge with thanks the technical and financial support from the Deanship of Scientific Research (DSR) at the King Abdulaziz University (KAU), Jeddah, Saudi Arabia, under the grant number G-661-611-38. The experiments reported in this paper were performed on the Aziz supercomputer at King AbdulAziz University.

## References

1. Suma, S., Mehmood, R., Albugami, N., Katib, I., Albeshri, A.: Enabling next generation logistics and planning for smarter societies. *Procedia Comput. Sci.* **109**, 1122–1127 (2017)
2. London.gov.uk: Notting Hill Carnival (2017). <https://www.london.gov.uk/events/2017-08-26/notting-hill-carnival-2017>
3. Khan, Z., Anjum, A., Soomro, K., Tahir, M.A.: Towards cloud based big data analytics for smart future cities. *J. Cloud Comput.* **4**, 1–11 (2015)
4. Herrera-Quintero, L.F., Banse, K., Vega-Alfonso, J., Venegas-Sanchez, A.: Smart ITS sensor for the transportation planning using the IoT and Bigdata approaches to produce ITS cloud services, pp. 3–9 (2016)
5. Kolchyna, O., Treleaven, P.C., Aste, T.: A framework for twitter events detection, differentiation and its application for retail brands (2016)
6. Arfat, Y., Aqib, M., Mehmood, R., Albeshri, A., Katib, I., Albogami, N., Alzahrani, A.: Enabling smarter societies through mobile big data fogs and clouds. *Procedia Comput. Sci.* **109**, 1128–1133 (2017)
7. Ayres, G., Mehmood, R.: On discovering road traffic information using virtual reality simulations. In: 11th International Conference on Computer Modelling and Simulation, UKSim 2009, pp. 411–416 (2009)
8. Ayres, G., Mehmood, R.: LocPriS: a security and privacy preserving location based services development framework (2010)
9. Mehmood, R., Graham, G.: Big data logistics: a health-care transport capacity sharing model. *Procedia Comput. Sci.* **64**, 1107–1114 (2015)

10. Mehmood, R., Lu, J.A.: Computational Markovian analysis of large systems. *J. Manuf. Technol. Manag.* **22**, 804–817 (2011)
11. Mehmood, R., Meriton, R., Graham, G., Hennelly, P., Kumar, M.: Exploring the influence of big data on city transport operations: a Markovian approach. *Int. J. Oper. Prod. Manag.* (2016, forthcoming)
12. Graham, G., Mehmood, R., Coles, E.: Exploring future cityscapes through urban logistics prototyping: a technical viewpoint. *Supply Chain Manag.* **20**, 341–352 (2015)
13. Alazawi, Z., Alani, O., AbdLjabar, M.B., Altowaijri, S., Mehmood, R.: A smart disaster management system for future cities. In: *International Workshop on Wireless and Mobile Technologies for Smart Cities, WiMobCity 2014*, pp. 1–10 (2014)
14. Alazawi, Z., AbdLjabar, Mohmmad B., Altowaijri, S., Vegni, A.M., Mehmood, R.: ICDMS: an intelligent cloud based disaster management system for vehicular networks. In: Vinel, A., Mehmood, R., Berbineau, M., Garcia, C.R., Huang, C.-M., Chilamkurti, N. (eds.) *Nets4Cars/Nets4Trains 2012*. LNCS, vol. 7266, pp. 40–56. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-29667-3\\_4](https://doi.org/10.1007/978-3-642-29667-3_4)
15. Gu, Y., Sean, Z., Chen, F.: From twitter to detector: real-time traffic incident detection using social media data. *Transp. Res. Part C Emerg. Technol.* **67**, 321–342 (2016)
16. Nguyen, D.T., Jung, J.E.: Real-time event detection for online behavioral analysis of big social data. *Futur. Gener. Comput. Syst.* **66**, 137–145 (2017)
17. Unankard, S., Li, X., Sharaf, M.A.: Emerging event detection in social networks with location sensitivity. *World Wide Web* **18**, 1393–1417 (2015)
18. Wang, Y.: Tweeting cameras for event detection categories and subject descriptors. In: *International World Wide Web Conferences Steering Committee*, pp. 1231–1241 (2015)
19. Kaleel, S.B., Abhari, A.: Cluster-discovery of twitter messages for event detection and trending. *J. Comput. Sci.* **6**, 47–57 (2015)
20. D'andrea, E., Ducange, P., Lazzarini, B., Marcelloni, F.: Real-time detection of traffic from twitter Stream analysis. *IEEE Trans. Intell. Transp. Syst.* **16**, 2269–2283 (2015)
21. Li, R., Lei, K.H., Khadiwala, R., Chang, K.C.C.: TEDAS: a Twitter-based event detection and analysis system. In: *2012 IEEE 28th International Conference on Data Engineering, Washington, DC*, pp. 1273–1276 (2012). <https://doi.org/10.1109/ICDE.2012.125>
22. Gutierrez, C., Figuerias, P., Oliveira, P., Costa, R., Jardim-Goncalves, R.: Twitter mining for traffic events detection. In: *2015 Science and Information Conference (SAI)*, pp. 371–378. IEEE (2015)
23. Apache: Apache Spark. <https://spark.apache.org/>
24. Fujitsu Ltd.: Fujitsu Releases World's Highest-Performance File System. <http://www.fujitsu.com/global/about/resources/news/press-releases/2011/1017-01.html>
25. Ranks.nl: stopwords. <http://www.ranks.nl/stopwords>
26. Lextek.com: Stop Word List 1. <http://www.lextek.com/manuals/onix/stopwords1.html>
27. Github.com/Alir3z4: stop-words. <https://github.com/Alir3z4/stop-words/blob/master/english.txt>
28. Tableau: What is tableau - make your data make an impact. <https://www.tableau.com/trial/tableau-software>
29. Visitlondon.com: London Events Calendar. <http://www.visitlondon.com/things-to-do/whats-on/special-events/london-events-calendar#KRbiWhui4SMAd9PT.97>
30. Underbellyfestival.com: About underbelly festival. <http://www.underbellyfestival.com/about>