

SOCIAL MEDIA ANALYTICS FOR SENTIMENT ANALYSIS AND EVENT DETECTION IN SMART CITIES

Aysha Al Nuaimi, Aysha Al Shamsi and Amna Al Shamsi, Elarbi Badidi

College of Information Technology, United Arab Emirates University, Al-Ain,
United Arab Emirates

ABSTRACT

Smart cities utilize Internet of Things (IoT) devices and sensors to enhance the quality of the city services including energy, transportation, health, and much more. They generate massive volumes of structured and unstructured data on a daily basis. Also, social networks, such as Twitter, Facebook, and Google+, are becoming a new source of real-time information in smart cities. Social network users are acting as social sensors. These datasets so large and complex are difficult to manage with conventional data management tools and methods. To become valuable, this massive amount of data, known as 'big data,' needs to be processed and comprehended to hold the promise of supporting a broad range of urban and smart cities functions, including among others transportation, water, and energy consumption, pollution surveillance, and smart city governance. In this work, we investigate how social media analytics help to analyze smart city data collected from various social media sources, such as Twitter and Facebook, to detect various events taking place in a smart city and identify the importance of events and concerns of citizens regarding some events. A case scenario analyses the opinions of users concerning the traffic in three largest cities in the UAE

KEYWORDS

Internet of things, Urban data streams, Stream processing, Big data, Analytics

1. INTRODUCTION

Modern cities use digital technologies to reduce costs, balance budgets, enhance the efficiency of various city systems, optimize city management, improve the quality of services delivered to citizens, create new facilities for the public, reduce energy consumption and thus offer a better quality of urban life. These technologies create new opportunities for cities to make themselves smarter through innovative planning and information-based management and operation. As the amount of data, in structured and unstructured formats, is so huge, new approaches to data management are needed [1]. Besides, social networks (such as Twitter, Facebook, and Google+) are becoming a new source of real-time information in smart cities. Social network users are regarded as social sensors. To become valuable, this massive amount of data, known as 'big data,' needs to be processed and comprehended.

Big data analytics is a recent technology that has an immense potential to permit comprehending city data and, hence, enhancing smart city services. Effective management and analysis of big data is a fundamental component to achieve the goals of the smart city. These goals include tackling the problems, reducing resources consumption and costs, engaging actively with citizens, and making informed decisions that will result in enhancing the environment and improving economic outcomes leading to improved quality of urban life. For the big data scientist, there is opportunity amongst this vast amount and array of data. The analysis of big data has the potential to improve the lives of citizens and reduce costs by uncovering associations and by understanding the trends and patterns in the data.

In this work, we propose to collect and analyze social media data concerning the main cities in the UAE. An increasing number of events is taking place every year in these cities. Therefore, it is paramount to analyze the conversations of the citizens with regards to these events and other concerns. First, we focus on data retrieved from Twitter given the small size of the tweets. Twitter is an exciting source of information for real-time event detection and sentiment analysis. Twitter API is used to get up-to-date tweets. As social media analytics uses several analysis and modeling techniques, we focus primarily on techniques such as sentiment and trend analysis that support the data understanding phase. This quest will contribute to a better understanding of the needs and concerns of the public so that event organizers and municipal governments take appropriate action to address these concerns.

The remainder of the paper is organized as follows. Section 2 provides background information on the benefits of big data analytics in smart cities and the application of social media analytics for events' detection. Section 3 describes the methodology followed to analyze social media data streams. Section 4 describes a case scenario in which tweets concerning the UAE are analyzed by applying sentiment analysis. Finally, Section 5 concludes the paper and highlights future work.

2. BACKGROUND AND LITERATURE REVIEW

2.1 Big data analytics applications

Using advanced analytics techniques such as data mining, machine learning, statistical learning, text analytics, predictive analytics, and visualization tools, city stakeholders and local governments would be able to analyze previously inaccessible or unusable data to gain new insights resulting in significantly faster and informed decisions. These techniques can speed up the analytical investigation, leading to insights from both traditional and non-traditional data sources. Ideally, city and local government would use data analytics to monitor public utilities, alleviate traffic congestion, assess and anticipate crime, follow education trends, and carefully watch public resources.

Another smart city application where big data analytics will play a vital role is crowd control. Indeed, the cities are more and more crowded, and many events are organized with as many people attending the events. As a result, municipalities should provide many services to the participants such as safety, mobilizing police and emergency responders, basic needs such as food and beverages, and many more. Crowds of people mean that massive amounts of data are generated. Big data analytics can be used to predict the movement of the masses to avoid jostling or other severe disasters.

During emergencies, cities equipped with sensors can take advantage of collected data to make informed decisions. Using social media analytics during emergency response is of particular interest. Social media networks offer data streams, which can be used to collect near real-time information concerning an emergency. Though people post on social networks many unrelated messages, any emergency information can be valuable to emergency response teams and can help them to get a good picture of the situation, permitting a more effective and faster response that can reduce overall loss and damage [2].

2.2 Social media analytics

Social network sites, such as Facebook, Twitter, and Google+, have become so popular that they represent a new source of real-time information concerning various topics and events. The availability of social networks sites on numerous devices ranging from personal computers to tablets and smartphones contributed to their popularity [3][4].

People typically use social media networks to post small messages that allow them to express their opinions on a variety of topics or to report events occurring in their vicinity. These networks allow their users to have an identity, build small online communities, find other users with similar interests, and find content published by other users [5]. Shared messages on these networks are called Status Update Messages (SUM). In addition to the text of the message, a SUM contains metadata information such as the name of the user, timestamp, geographic coordinates (latitude and longitude), hashtags, and links to other resources.

The SUMs originating from users in a specific geographical area, such as a city, or discussing a topic or raising a concern can offer valuable information on an issue or event that can help decision-makers make informed decisions. As a result, social network users are considered as social sensors [6][7], and SUMs as sensor information [8].

Social media analysis is about collecting data from social media websites and blogs and processing that data as structured information that can lead decision-makers to make information-driven decisions. The customer is at the center of these decisions. The most common use of social media analytics is to leverage customers' opinions about products and services to support customer service and marketing activities.

2.3 Detection of events from social media analytics

In recent years, social networks have become a new source of information that municipal authorities can use to detect events, such as traffic jams, and get reports on incidents and natural disasters (fires, storms, tremors etc.) in their neighborhood. An event is a real-world occurrence that happens at a given moment and space [3] [9]. In particular, regarding traffic-related incidents, people often share using SUM information about the current traffic situation around them while driving.

However, event detection from the analysis of social networks' short-messages is more challenging compared with event detection from conventional media, such as emails and blogs, where texts are well-formatted [4]. SUMs are unstructured and irregular texts, which might contain simple or abbreviated words, grammatical errors or misspellings [3]. They are usually very brief, which makes them an incomplete source of information [4]. Moreover, SUMs

contain a massive amount of meaningless or not useful information [10], which require filtering. To extract meaningful information from social media networks, it is necessary to use text mining techniques that use methods in the fields of Natural Language Processing (NLP), machine learning, and data mining. [11] [12].

Concerning current approaches for using social media to obtain useful information for event detection, there is a need to distinguish between large-scale events and small-scale events. Large-scale events such as the election of a president, earthquakes or the tsunami are generally characterized by a considerable number of SUMs and a wider temporal and geographic coverage. Instead, small events such as traffic jams, car accidents, fires or local shows are characterized by a small number of associated SUMs, limited geographic and temporal coverage [21]. As a result, small-scale event detection is not a trivial task because of the smaller number of SUMs related to these kinds of events. Several research efforts studied event detection from social networks data. Many of them deal with large-scale event detection [8][13][14][15][16] and only a few of these works investigated small-scale events [17][18][19][20].

3. METHODOLOGY

A three-phase process, which includes data capture, data understanding, and presentation is used in this work (see Figure 1).

Data Capture. This phase helps identify messages on social media networks about the city's activities, events, and concerns. This process is achieved by collecting massive amounts of pertinent data from social media networks using primarily APIs, provided by these sources, or through crawling. The main popular social media platforms are Twitter, Facebook, Google+, LinkedIn, and YouTube. Some preprocessing steps may be performed to prepare a dataset for the data understanding phase. They typically include modeling and data representation, syntactic and semantic operations, linking data originating from different sources, and feature extraction.

Data understanding. In this phase, the meaning of collected data in the previous phase is assessed and metrics useful for decision-making are generated. As collected data might originate from multiple sources, it might contain a significant amount of noise that need to be cleaned before conducting meaningful analyses. This cleaning function may use simple, rule-based text classifiers or more sophisticated classifiers trained on labeled data. To assess the meaning of cleaned data, various techniques such as text mining, natural language processing, social network analysis, and statistical methods can be used. This phase can provide information about user opinions concerning an event or a service. Its results significantly impact the information and metrics in the presentation phase as well as the future decisions and actions a smart city might take.

Presentation. In this last phase, the results of different analyzes are summarized, evaluated and presented to users in an easy-to-understand format. The visual dashboard, which aggregates and displays data from multiple sources, is a commonly used interface design.

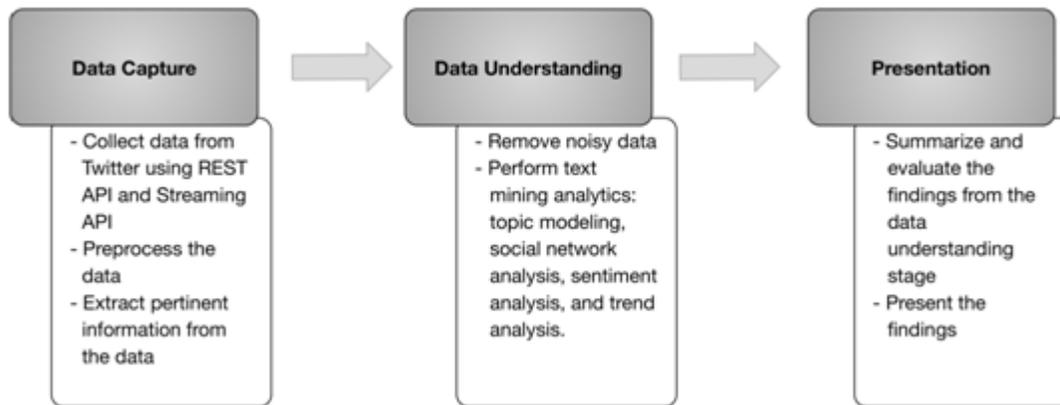


Figure 1. Overview of the proposed method

4. CASE SCENARIO

In this scenario, we implement a real-time application to get the latest Twitter feeds concerning the #Dubai, #AbuDhabi, and #Sharjah hashtags and process them by subjecting them to sentiment analysis.

Data streams (tweets) from Twitter have been recognized as a valuable data source for many smart cities in areas such as law enforcement, tourism, and politics (e.g., US presidential election). The Twitter Streaming API permits extracting datasets that are then used to perform sentiment analysis.

The following workflow describes the different steps we used for the analysis of the tweets:

1. Authenticate and connect to Twitter using Twitter API
2. Use search REST API to gather tweets from an Input file using keywords, hashtags, and Twitter account.
3. Store collected tweets in a CSV file
4. Preprocess the gathered tweets by removing duplicates
5. Use TextBlob to obtain sentiment for the unique tweets. TextBlob is a Python library for processing textual data. The result of a sentiment analysis task is the percentage of positive, negative, and neutral opinions.
6. Compute total, positive, negative and neutral tweets over a number of days.
7. Visualize results

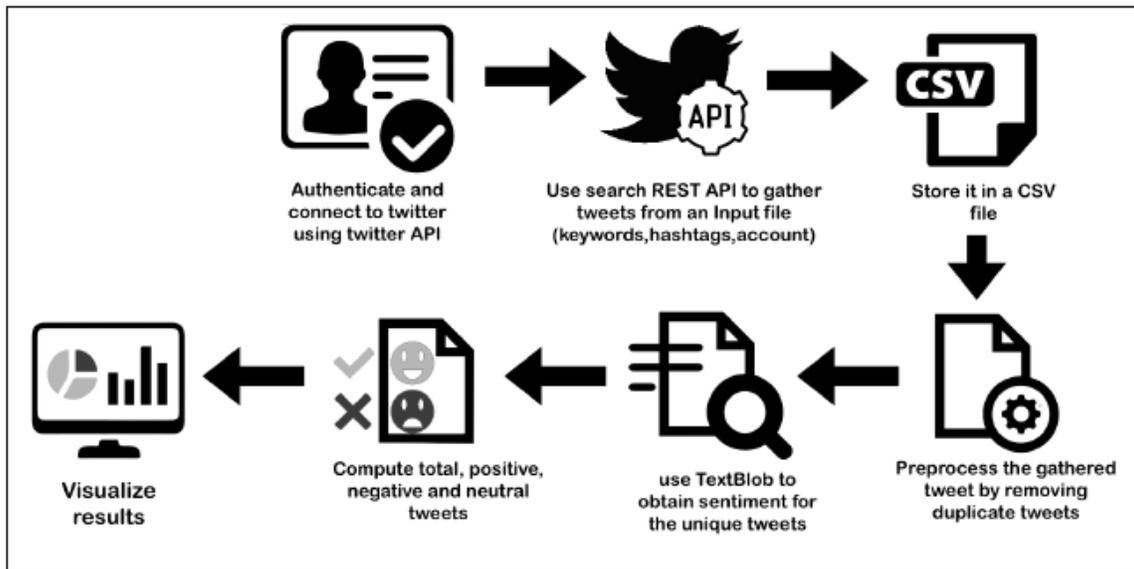


Figure 2.A workflow for the sentiment analysis of tweets

Figure 3 depicts the results obtained by analysing the opinions of the users regarding the traffic in the UAE main cities: Dubai, Abu Dhabi, and Sharjah over a week period from March 7, 2018, to March 12, 2018.

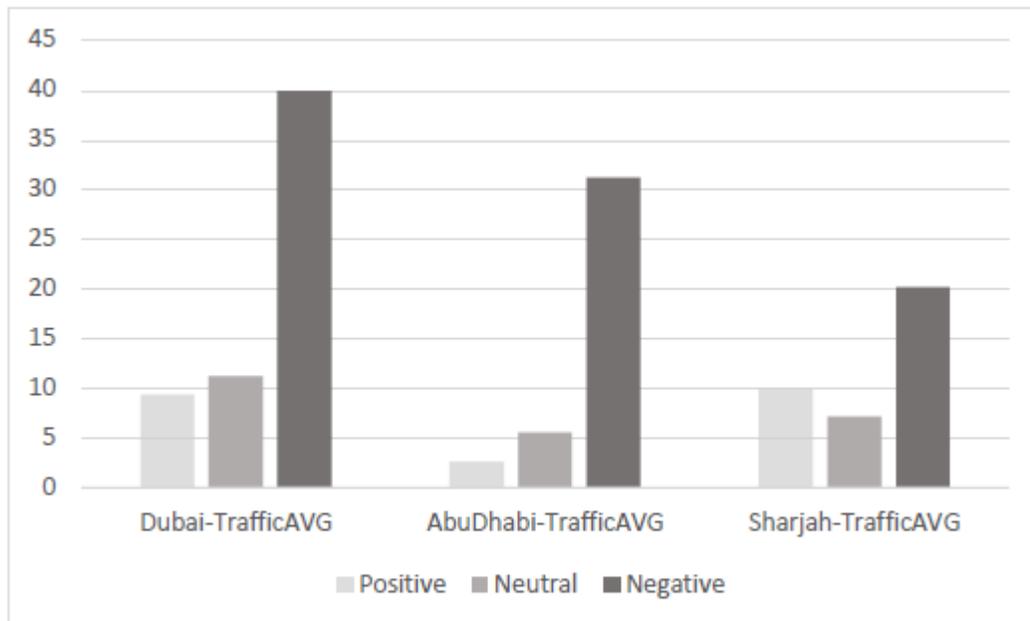


Figure 3. Opinions of the users regarding the traffic in the UAE main cities

The results show that a significant proportion of users have a negative opinion on the traffic in the three large UAE cities with more negative views on the traffic in Dubai. The percentage of users with a positive opinion is relatively low compared with negative opinions in the three cities. However, for Sharjah, that percentage is a little bit higher than the percentage of neutral opinions, which is not the case for Dubai and Abu Dhabi. These results indicate that users are suffering from the traffic conditions in the three cities and that municipal services should work hard to find solutions to the traffic problem, which is one of the challenges that most megacities are facing.

5. CONCLUSION

Modern cities are more and more relying on the usage of the Internet of Things (IoT) devices and sensors to sense various parameters such as temperature humidity, water leaks, sunlight, and air pressure. These devices and sensors generate massive volumes of data. Also, social networks are becoming a new source of real-time information in smart cities. Social network users are acting as social sensors. In this work, we described how social media analytics could help analysing urban data streams collected from popular social media sources, such as Twitter and Facebook, to detect events taking place in a smart city and identify the concerns of citizens regarding some events or issues. We analyse in a case scenario the sentiments of users concerning the traffic in three largest cities in the UAE. The results show how frustrated are the users with the traffic in the cities of Dubai and Abu Dhabi that their municipalities need to work hard to alleviate this issue.

REFERENCES

- [1] I. A. T. Hashem, V. Chang, N. B. Anuar, K. Adewole, I. Yaqoob, A. Gani, E. Ahmed, and H. Chiroma, (2016) "The role of big data in smart city," *International Journal of Information Management*, vol. 36, no. 5, pp. 748–758.
- [2] Tim L. M. van Kasteren, Birte Ulrich, Vignesh Srinivasan, Maria E. Niessen, (2014) "Analyzing Tweets to Aid Situational Awareness," *Advances in Information Retrieval*, Vol. 8416, *Lecture Notes in Computer Science*, pp 700-705.
- [3] F. Atefeh and W. Khreich, (2015) "A survey of techniques for event detection in Twitter," *Comput. Intell.*, vol. 31, no. 1, pp. 132–164.
- [4] P. Ruchi and K. Kamalakar, (2013) "ET: Events from tweets," in *Proc. 22nd Int. Conf. World Wide Web Comput.*, Rio de Janeiro, Brazil, pp. 613–620.
- [5] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, (2007) "Measurement and analysis of online social networks," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas.*, San Diego, CA, USA, pp. 29–42.
- [6] G. Anastasi et al., (2013) "Urban and social sensing for sustainable mobility in smart cities," in *Proc. IFIP/IEEE Int. Conf. Sustainable Internet ICT Sustainability*, Palermo, Italy, pp. 1–4.
- [7] A. Rosi et al., (2011) "Social sensors and pervasive services: Approaches and perspectives," in *Proc. IEEE Int. Conf. PERCOM Workshops*, Seattle, WA, USA, pp. 525–530.
- [8] T. Sakaki, M. Okazaki, and Y. Matsuo, (2013) "Tweet analysis for real-time event detection and earthquake reporting system development," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 4, pp. 919–931.

- [9] J. Allan, (2002) “Topic Detection and Tracking: Event-Based Information Organization”. Norwell, MA, USA: Kluwer.
- [10] J. Hurlock and M. L. Wilson, (2011) “Searching Twitter: Separating the tweet from the chaff,” in Proc. 5th AAAI ICWSM, Barcelona, Spain, pp. 161–168.
- [11] S. Weiss, N. Indurkha, T. Zhang, and F. Damerau, (2004) “Text Mining: Predictive Methods for Analyzing Unstructured Information,” Berlin, Germany: Springer-Verlag.
- [12] Hotho, A. Nürnberger, and G. Paaß, (2005) “A brief survey of text mining,” LDV Forum-GLDV J. Comput. Linguistics Lang. Technol., vol. 20, no. 1, pp. 19–62.
- [13] M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, (2012) “Getting there first: Real-time detection of real-world incidents on Twitter” in Proc. 2nd IEEE Work Interactive Vis. Text Anal.—Task-Driven Anal. Soc. Media IEEE VisWeek,” Seattle, WA, USA.
- [14] C. Chew and G. Eysenbach, (2010) “Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak,” PLoS ONE, vol. 5, no. 11, pp. 1–13.
- [15] B. De Longueville, R. S. Smith, and G. Luraschi, (2009) “OMG, from here, I can see the flames!: A use case of mining location based social networks to acquire spatiotemporal data on forest fires,” in Proc. Int. Work. LBSN, Seattle, WA, USA, pp. 73–80.
- [16] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, (2012) “Using social media to enhance emergency situation awareness,” IEEE Intell. Syst., vol. 27, no. 6, pp. 52–59.
- [17] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, (2012) “Real-time event extraction for driving information from social sensors,” in Proc. IEEE Int. Conf. CYBER, Bangkok, Thailand, pp. 221–226.
- [18] P. Agarwal, R. Vaithyanathan, S. Sharma, and G. Shro, (2012) “Catching the long-tail: Extracting local news events from Twitter,” in Proc. 6th AAAI ICWSM, Dublin, Ireland, Jun. pp. 379–382.
- [19] F. Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao, (2012) “Twitcident: fighting fire with information from social web streams,” in Proc. ACM 21st Int. Conf. Comp. WWW, Lyon, France, pp. 305–308.
- [20] R. Li, K. H. Lei, R. Khadiwala, and K. C.-C. Chang, (2012) “TEDAS: A Twitter- based event detection and analysis system,” in Proc. 28th IEEE ICDE, Washington, DC, USA, pp. 1273–1276.
- [21] A. Schulz, P. Ristoski, and H. Paulheim, (2013) “I see a car crash: Real-time detection of small scale incidents in microblogs,” in The Semantic Web: ESWC 2013 Satellite Events, vol. 7955. Berlin, Germany: Springer-Verlag, pp. 22–33.
- [22] TextBlob. TextBlob: Simplified Text Processing. <http://textblob.readthedocs.io/en/dev/>