

Towards developing a Healthcare Situation Monitoring Method for Smart City Initiatives: A Citizen Safety Perspective

Shah Jahan Miah

VU Business School,
Victoria University,
Melbourne, Australia
shah.miah@vu.edu.au

H. Quan Vu

Department of Information Systems and Business Analytics,
Deakin Business School,
Deakin University,
Burwood, Australia

Abstract

Research in Smart City development has proliferated over the past few years, focussing heavily on various supporting service sectors, such as healthcare. However, little effort has been made to design health surveillance support systems, which are also important for the advancement of public healthcare monitoring as an essential aspect of smart city initiatives. From an information system (IS) design perspective, this paper introduces a social media-based health surveillance supporting method, which can automatically extract relevant online posts for health symptom prediction and management. We describe and demonstrate an IS design approach in this paper for hay-fever prediction solution concept based on Twitter posts. This concept could be developed into a fully functional solution design by practitioners in this field.

Keywords: healthcare, smart city, social media data, healthcare monitoring, public health surveillance

1 Introduction

Smart city research initiatives have proliferated over the past few years in various service development sectors, since Hollands (2008) introduced its core idea. Many academic and non-academic efforts have been initiated for developing research in smart city. Studies that promote people-oriented needs, especially on issues related to healthy living and environment, such as public safety, transportation, environmental sustainability and health care for active living, are still in high demand (Guedes et al. 2018). Examples of relevant studies that focus on the mentioned aspects in the healthcare domain include Forkan, Khalil, Tari, Foufou, and Bouras (2015), Aziz et al. (2016), and Aborokbah et al. (2018). However, current research into healthcare shows limited understanding on formulating real problems and practical conceptualizations for building solution design options and explorations in actual scenarios. Developments are still required, such as for strategic decision support by smart city authorities.

Social media have created huge opportunities and data sources for developing virtual surveillance systems for healthcare authorities. Social media data offers real-time access to big-

data that contain important details of citizens' lifestyle and personal well-being. Social media big data have been considered both as a valuable and reliable source for identifying concurrent details of human activities and their opinions, (Thelwall et al. 2016). However, appropriate surveillance systems methodology for rigorous analysis of social media data is still under-developed, in particular for monitoring health conditions symptoms, such as hay-fever symptoms among communities. This would support the authorities ability to observe signs or indications of true predictions and prevalence.

In this work, we aim to address this need, especially for designing social media-based health surveillance support for the prediction of public disease and their spreading symptoms for effective management. The current study differs from prior studies above by the use of online social media data (e.g. Twitter), which has been proven a useful and reliable source of data collections in recent studies (He et al., 2015; Chua et al., Miah, Vu, Gammack & McGrath, 2017; 2016; Miah, Vu and Gammack, 2019).

2 Methodology

Twitter data relating health symptoms, such as "hay fever" also known as pollen allergy, are extracted during the spring season in a specific country or region, using Twitter's Application Programming Interface (API) (Twitter, 2018). Keywords, such as "hayfever", "hay fever", "allergic", "allergy" can be provided as parameters to retrieve relevant tweets. The geographical area to perform the surveillance can be defined as a bounding box, whose coordinates can be specified as parameters corresponding to maximum latitude (max_{lat}), minimum latitude (min_{lat}), maximum longitude (max_{lon}) and minimum longitude (min_{lon}). The textual content of the tweets and their meta-data, such as UserID, User Location, Date and Time of tweet, geographical location of tweet, are also extracted where available.

The raw tweet data are unstructured and needs to be converted into suitable format for subsequent analysis. We first start with processing textual content of the tweet following some standard steps for text mining (Sharda et al., 2014). The text is first loaded into a text *tokenizing* algorithm, wherein the string of text is broken into words, phrases, or symbols called "tokens". The tokens are then passed through a *filter*, where capital letters are normalized into lower case. Tokens containing symbols, numbers or stop words (e.g. "the", "that" and "very") are removed because they are not considered useful for analysing the tweet meaning. The remaining tokens then go through a *stemming* processing to reduce words into their stem, base, or root form, such as "sneezing" to "sneez" and "flowers" to "flower". The processed textual contents of each tweet are stored as a list of stemmed words for later analysis. Next, we process location data of the tweets to identify place name. The latitude and longitude are input into Google's Geocoding API (Google, 2018), which return the name of the corresponding location (city and country).

Once the data are collected and pre-processed, we apply several exploratory analysis techniques to the data for extracting meaningful insights, including important word analysis, word relationship analysis and temporal pattern analysis.

3 Case Study

We evaluated the effectiveness of the proposed approach in healthcare situation monitoring for Australia, whose large population proportion suffers from hay-fever. We deployed a Streaming function of Twitter API to cover spring season in Australia. A bounding box, whose

coordinates are $max_{la}=-11.0442$ $min_{la}=-45.3812$, $max_{lo}=153.8456$ and $min_{lo}=110.6787$, was defined to cover the entire geographical area of Australia. All collected tweets were validated, duplicate tweets were removed, and the tweets arranged in temporal order from earliest to latest. After data pre-processing, a total of 1893 tweets relevant to hay fever from 756 users were available for analysis. The majority of the tweets were posted in the six most populated cities in Australia; Melbourne, Sydney, Perth, Brisbane, Adelaide and Canberra.

The first analysis stage is to identify popular words relevant to hay fever in the data set. We computed the frequency of each word with respect to the number of tweets, and then selected most popular ones for examination. From the management perspective, government agencies would like to monitor public health in relation to hay-fever such as condition, cause, effect on the body, treatment and affected season. We examined the remaining frequent words to identify those relevant to hay fever as shown in Table 1. The most frequent relevant words are “allergic” “hay fever”, “fever” and “allergy”, which is consistent with the fact that these words were used as keywords to detect those relevant tweets. We found “asthma”, “cold”, “sick” are also frequently mentioned in the tweet data set. “Sneez” and “itchy” are common symptoms in relation to hay fever. “Sneez” is a root form of the word sneezing after removing the suffix “ing” through the stemming process. Common causes of hay fever were identified in the tweets, including “pollen”, “flower”, “grass” and “tree”, with “pollen” being the most frequent cause (support value of 0.134). Other allergy causes were also included, such as “food”, “cat”, “dog” and “peanut”. The most affected body part was “eye”, “nose”, “face”, “head” and “throat”. We found the word “antihistamine”, drugs that treat allergic rhinitis and allergies, were frequently mentioned in the tweets. “spring” and “summer” are common seasons mentioned posts about hay fever and allergy.

Group	Condition	Cause	Body Part	Treatment	Season
Word (support)	allergic (0.268)	pollen (0.134)	eye (0.025)	antihistamine (0.010)	spring (0.057)
	hay fever (0.241)	food (0.037)	nose (0.018)		summer (0.016)
	fever (0.166)	cat (0.029)	face (0.012)		
	allergy (0.166)	flower (0.016)	head (0.010)		
	reaction (0.051)	dog (0.015)	throat (0.010)		
	asthma (0.024)	peanut (0.013)			
	cold (0.021)	grass (0.011)			
	sneez (0.020)	tree (0.010)			
	sick (0.012)				
	itchy (0.010)				

Table 1: Popular words relevant to hay fever

Subsequently, we analysed the relationships between them for understanding how the condition and causes are linked together. A graph visualization technique was applied to illustrate the connection between terms. Each word is a node, and an edge represents a connection between two words if they appear in the same post. For the ease of analysis, we only draw the edge between words that appear together in at least five tweets. The strength of the connection is represented by the thickness of the edge and the number of tweets appearing together. Figure 1a shows that people tend to use the word “allergic reaction” when mentioning about other type of allergy such as with “cat”, “dog”, “food allergy”. “Pollen” from “flower”, “grass” and “tree” are the main cause of “hay fever” and “asthma”. Common symptoms linked to hay fever are “sneez”, “itchy” and “sick. We found that “hay fever” and “cold” are also mentioned frequently in the same tweets, although they are basically different

type of conditions. This is probably because people sometime get confused between these two conditions due to their similar symptoms. Figure 1b shows that “eye”, “head”, “throat” and “nose” are the body part affected by “hay fever”, and “antihistamine” is commonly mentioned as its treatment. For people have allergic reaction other than hay fever, their “face” is mostly affected.

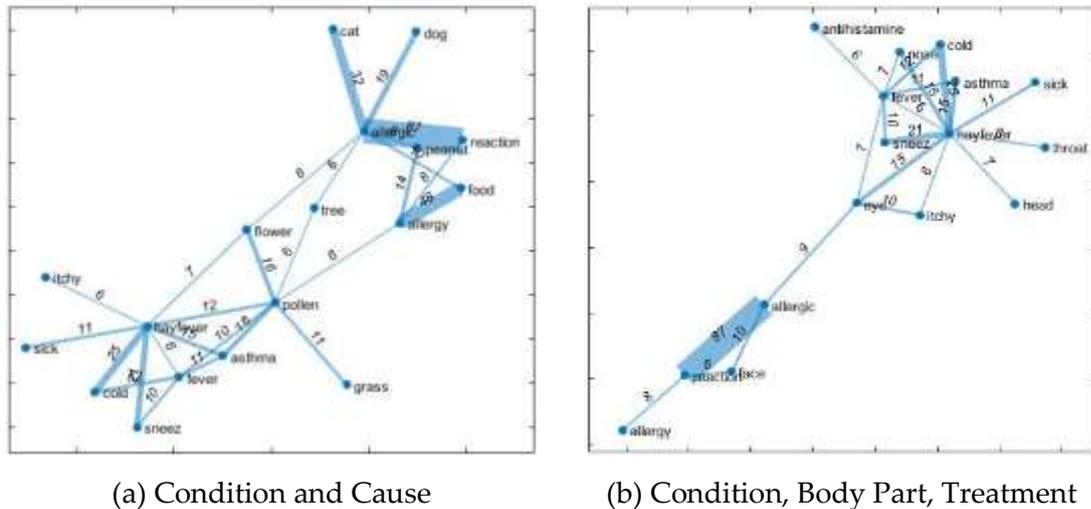


Figure 1: Association between words in tweets

Next, we examine the temporal patterns of the tweets in relation to the diseases. We found different patterns among tweets containing the words “allergic” and “hay fever. The peak period for tweeting about hay fever is in springtime of Australia (September to December). However, tweets containing the word “allergic” were mainly posted at the beginning of the year (January to March) and at the end of the year (October to December), which indicates that the allergic reaction is caused from other sources rather than pollen.

Further temporal analysis of hay fever causes showed that the peak time for “pollen” is August to December, which is aligned with the time for spring and early summer. It should be noted that different people might react with different type of pollens, so detailed information on which pollen type may provide useful information for treatment. We can see that the peak time for “flower” is September and October while, the peak time for “tree” is December. Tree pollens are different from ordinary flower pollens, as they are usually from tree species bearing separate male and female flowers in the same plant such as oak, sweet gum, pine and spruce.

We further examined the patterns with respect to geographical locations (e.g. city) and found that. Melbourne residents are more likely to tweet about their condition around 9:00 to 11:00, while Sydney residents are likely to tweets in early morning 8:00 to 9:00 or in the afternoon around 16:00. More people in Melbourne are having allergy caused by pollen from “grass” and “tree” than those in Sydney. On the other hand, Sydney residents posted tweets containing “cat” and “flower” more than those in Melbourne and more residents having symptoms at their head.

4 Discussions and Conclusion

Under the smart city research paradigm, our study aimed to extend current views of the social media analytics in particular for the purpose of public health-related knowledge discovery.

The suitability of social media data for disease surveillance was demonstrated in a case study of hay fever monitoring in Australia. The analysis, based on the characteristic features of hay fever related Twitter messages, provided useful insights. This paved away for a larger scale investigation into the topic in the upcoming pollen seasons. As a result, the benefits of high-quality input data as well as the visualization tool for patterns exploration considerably facilitated new knowledge discovery.

The pilot method extracted valuable knowledge about this common disease from Twitter data. This study appears to be the first that attempts to utilize social media analytics for hay fever surveillance in Australia. The results have shown that our method is effective in capturing tweets posted by Australian residents about their health condition, especially in relation to hay fever and allergy. Textual analysis with graph visualization was able to reveal interesting relationships between the conditions, cause, and affected body part. Temporal patterns revealed insights into the times when people tweet about different conditions and causes. Comparison between cities suggests that residents at different location could have different health conditions and associated causes.

We attempted to address healthcare authorities' requirements for accurate discovery of public diseases and their spreading symptoms for a healthcare management. Smart technologies such as text processing techniques can produce relevant insights that provide supporting information that may inform planning for surges in patient visits, therapeutic supplies, and public health information dissemination campaigns. These aspects are very important for policy makers or decision makers in government or public welfare organizations. It is important because the pollen allergy, commonly referred to as 'hay fever' induces a range of problematic symptoms that significantly interfere with sufferers' daily activities.

The limitation of this pilot work is that we only ran the data for a single year, and as the data was accessed using a free account with a quota limitation, the collected data may not generalize well. However, we did find some interesting and useful patterns from the current data set. The analysis with respect to location was at large scale of city level. Detailed analysis at suburb level would be useful to provide further insight when there is more data with GPS information available. A further study would aim to develop a fully functional prototype of a social media analytics solution that could other other aspects within community boundaries. A context sensitive design methodology (Miah, 2004; Miah and Gammack, 2014), in particular for decision support or business intelligence design (Ali, Miah and Khan, 2018; Miah & Ahmad, 2013; Miah, 2008), would be applied for designing an innovative social media analytics artifact.

References

- Aborokbah, M.M., Al-Mutairi, S., Kumar, A., & Samuel. O.W., (2018). Adaptive Context Aware Decision Computing Paradigm for Intensive Health Care Delivery in Smart Cities—A Case Analysis, *Sustainable Cities and Society*, 41, 919-924
- Ali, MS, Miah, SJ, & Khan, S. (2018). Antecedents of Business Intelligence Implementation for addressing Organizational Agility in Small Business Context, *Pacific Asia Journal of the Association for Information Systems*, 10 (1), 89-108
- Aziz, K., Tarapiah, S., Ismail, S.H., & Atalla, S. (2016). Smart Real-Time Healthcare Monitoring and Tracking System using GSM/GPS Technologies, In the proceedings of the 3rd MEC

- International Conference on Big Data and Smart City, 2016, URL: <https://ieeexplore.ieee.org/document/7460394>
- Chua, L. Servillo, E., Marcheggiani, A. & Moere, V. (2016). Mapping Cilento: using geotagged social media data to characterize tourist flows in southern Italy, *Tourism Management*, 57, 295–310.
- Forkan, A., Khalil, I., Tari, Z., Foufou, S., & Bouras, A. (2015). A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living, *Pattern Recognition*, 48(3), 628-641.
- Google. (2018). Geocoding API. retrieved May 2018, URL: <https://developers.google.com/maps/documentation/geocoding/intro>
- Guedes, A. L. A. Alvarenga, J.C., Goulart, M.D.S.S., Rodriguez, M.V.R., & Soares, C. A. P. (2018). Smart Cities: The Main Drivers for Increasing the Intelligence of Cities, *Sustainability*, 10(9), 1-19
- He, W., Wu, H., Yan, G., Akula, V. & Shen, J. (2014). A novel social media competitive analytics framework with sentiment benchmarks, *Information and Management* 52, 801–812.
- Hollands, R. G. (2008). Will the real smart city please stand up? Intelligent, progressive or entrepreneurial? *City*, 12(3), 303-320
- Miah, S. J. (2004). Accessibility improvement of multicultural educational web interface by using the User Centred Design (UCD) approach, In the *Proceedings of the 2004 Informing Science and IT Education Joint Conference*, Rockhampton, QLD, Australia, 25-28 June 2004.
- Miah, S. J. (2008). An ontology based design environment for rural decision support, *Unpublished PhD*, Griffith Business School, Griffith University, Australia.
- Miah, S.J., & Ahmad, R. (2013). A cloud-based DSS model for driver safety and monitoring on Australian roads, *International Journal of Emerging Sciences*, 1 (4), 634
- Miah, S.J., Vu, HQ, Gammack, J. & McGrath, GM. (2017). A Big-Data Analytics Method for Tourist Behaviour Analysis, *Information and Management*, 54, 771-785
- Miah, S.J., & Gammack, J. (2014). Ensemble Artifact Design for Context Sensitive Decision Support, *Australasian Journal of Information Systems*, 18 (2), 5-20
- Miah, S.J., Vu, HQ, & Gammack, J. (2019). A Big-Data Analytics Method for capturing visitor activities and flows: the case of an Island Country, *Information Technology and Management*, 20 (4), 203-221
- Sharda, R., Delen, D., & Turban, E. (2014). *Business Intelligence and Analytics: Systems for Decision Support*, 10th edition. Pearson Education Limited, Essex, England.
- Thelwall, M., Goriunova, O., Vis, F., Faulkner, S., Burns, A. et al. (2016). Chatting Through Pictures? A Classification of Images Tweeted in One Week in the UK and USA, *Journal of the Association for Information Science and Technology*, 67 (11), 2575-2586.
- Twitter. (2018). *The streaming APIs*. retrieved May 2018, URL: <https://developer.twitter.com/en/docs>

Copyright: © 2020 Miah & Vu. This is an open-access article distributed under the terms of the [Creative Commons Attribution-NonCommercial 3.0 Australia License](https://creativecommons.org/licenses/by-nc/3.0/australia/), which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and AJIS are credited.

doi: <https://doi.org/10.3127/ajis.v24i0.2551>

