5-2018

Malaria Surveillance System Using Social Media

James Boit
Dakota State University, James.Boit@trojans.dsu.edu

Hassan Alyami
Dakota State University, hmalyami@pluto.dsu.edu

Malaria Surveillance System Using Social Media

James Boit
Dakota State University
James.boit@trojans.dsu.edu

Hassan Alyami
Dakota State University
hmalyami@pluto.dsu.edu

ABSTRACT

Social media, for example, Twitter has increasingly provided opportunities for massive data collection of topical issues affecting today's society. Opinions and data in public health issues are very prevalent on Twitter and provide an invaluable source of interesting information that can be mined for decision making in public health organizations. This paper discusses the existing malaria surveillance system and proposes a malaria surveillance system (MSS) that leverages social media with a view to enhancing decision making by public health professionals. The MSS system comprises of a data collection module, analysis engine, and a metrics module. The practical contribution of this research is the construction of a conceptual architecture for the MSS.

Keywords

Malaria, Twitter, sentiment analysis, public health surveillance, design science research.

INTRODUCTION

Social media such as Twitter gives public health providers avenues to mine vast amounts of data for purposes of gaining insights on mitigation measures towards emerging epidemic scenarios (Chew, 2010; Thomson et al., 2006). Also, Twitter can be described as a microblogging website or service launched in 2006 where users publish messages of up to 140 characters. Millions of tweets (short messages on twitter) are published and read daily on a wide gamut of topics ranging from personal to public thoughts (Pak & Paroubek, 2010; Tumasjan, Sprenger, Sandner, & Welpe, 2010).

Surveillance systems can be described as systems which detect and track disease outbreaks and trends thereby giving health organizations and professionals ample time to initiate intervention measures of distributing the necessary resources to affected areas within an effective and timely manner (Asur & Huberman, 2010; Runge-Ranzinger, Horstick, Marx, & Kroeger, 2008). In comparison, public health surveillance can be defined as a systematic process of data collection, analysis, and interpretation of health data with the goal of aiding government agencies and the public to enhance public health interventions (Olson, Konty, Mathes, & Paladini, 2013).

Traditional surveillance methods are known to be slow in responding to real-time incidences (Chew, 2010). Similarly, studies on internet-based surveillance systems are mostly sentinel site surveillance systems or use passive reporting mechanisms (Chan, Sahai, Conrad, & Brownstein, 2011).

Bureaucracy and lack of adequate resources continue to hinder effective detection and communication of infectious diseases in prone countries (Littrell et al., 2013). In fact, statistics indicate that over 90% of data stored today has been generated in the last five years with social media such as YouTube, Facebook, Twitter, and crowdsourcing information contributing a big chunk of the online data (Kass-Hout & Alhinnawi, 2013).

The objective of this study is to discuss the existing disease surveillance system from extant literature and propose an MSS system that leverages on social media. To the best of our knowledge, no study has focused on detecting and monitoring of malaria in sub-Saharan Africa that leverages social media.

The remainder of the paper is organized as follows. Next, we look at the background section, then the existing malaria surveillance systems, and present an existing conceptual framework that supports public health surveillance. Subsequently, we discuss the proposed artifact. Finally, we conclude and share the future direction of this research.

BACKGROUND

Malaria is an infectious and lethal disease affecting most parts of the world with high prevalent incidences occurring in sub-Saharan Africa with an 81% of worldwide cases and 90% of reported deaths (Khosa, Kuonza, Kruger, & Maimela, 2013). Most prior work has focused on infectious diseases such as influenza, and Ebola cases in the United states supported by data from Google trends on infectious diseases and statistics from CDC (Lazard, Scheinfeld, Bernhardt, Wilcox, & Suran, 2015). Malaria contributes to over 500 million infections worldwide with an outcome of more than 1 million deaths reported yearly primarily in sub-Saharan Africa (Thomson et al., 2006).

Similarly, previous studies on flu trends have proposed near real-time models for detection and prediction of influenza epidemic (Achrekar, Gandhe, Lazarus, Yu, & Liu, 2011; Chen, Achrekar, Liu, & Lazarus, 2010). These studies relied on the Google flu trends to provide real-time data on flu epidemics across multiple countries. Currently, other non-traditional sources of data exist like dedicated websites and social media platforms such as Facebook, and twitter where individuals readily share their ailments and experiences before deciding to seek medical services (Sharpe, Hopkins, Cook, & Striley, 2016).

Studies have demonstrated the application of text mining in social media as an important tool in the extraction of timely and efficient information for use by public health practitioners and providers (Lazard et al., 2015). Other similar studies have proposed a conceptual framework for Dengue surveillance system in Malaysia (Othman & Danuri, 2016).

Existing malaria surveillance systems

According to (Ohrt et al., 2015), limited studies currently exist that advance the use of an information system (collect, store, and analyze components) that leverage on real-time data. The following are examples of few comparisons of existing systems as categorized in their geographical regions (Ohrt et al., 2015).For instance, in China, Solomon Islands and Vanuatu, Zanzibar-Tanzania, there was a lack of an integrated mobile technology, lack of location or map to a household based on the origin of each case, and inability to capture new interventions (Cao et al., 2014; Kelly et al., 2013; Ohrt et al., 2015). In Swaziland, Thailand and Zambia, the system lacked capability to capture new interventions or what is referred to as intervention quality (Cohen et al., 2013; Littrell et al., 2013; Ohrt et al., 2015).

Conceptual framework

CDC developed a framework for use to support all public health surveillance systems (Buehler, Hopkins, Overhage, Sosin, & Tong, 2004). The framework is comprised of four categories mainly the systems description, outbreak detection, experience, and, conclusion and recommendation. Figure 1 gives a visual breakdown of the early detection model prescribed for development of public health surveillance systems (Buehler et al., 2004).

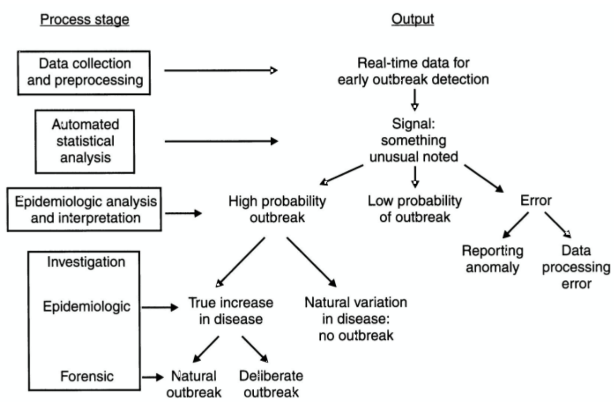

Figure 1. Process model for early outbreak detection

ARTIFACT

The artifact in this study is the construction of a conceptual architecture for the malaria surveillance system (MSS). The MSS leverage twitter data using key words such as malaria, outbreaks, impacts, interventions among others for data extraction. The results from the analysis engine will be displayed visually in a web portal and intuitively through social network graphs.

This section describes the proposed components of the system design process.

System design

The different components of the system architecture comprise of a data collection module, analysis engine and the metrics module. The data collection module as depicted in figure 2., involves data collection from twitter website through a twitter search interface (API). The extracted data will be stored in a MongoDB database for further processing. Consequently, the analysis engine will primarily focus on text classification and representation problems. Sentiment analysis will be performed on the tweets to determine the impact of valence in the streams of malaria tweets. Further, topic modeling will be performed using Latent Dirichlet Allocation (LDA) technique (Blei, Ng, & Jordan, 2003) to categorize emergent words according to the keywords. Lastly, the metrics module will comprise a website interface component integrated to the analysis engine.

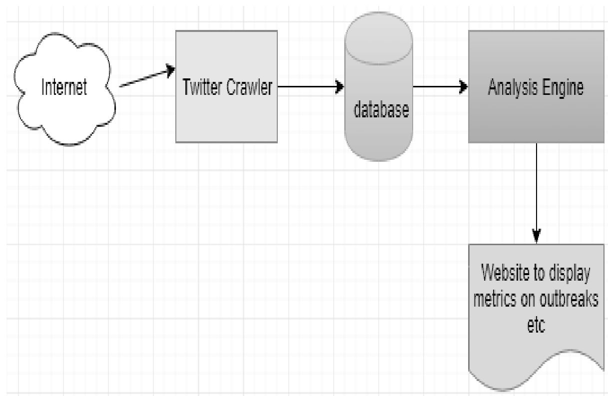

Figure 2. Proposed architecture of the MSS

Datasets

The World Health Organization (WHO) data is comprised of two datasets, namely; total number of malaria cases and total number of malaria deaths by country from the year 2000 to 2014. Twitter streaming API will be used to collect tweet feeds from 2006 to 2014. The WHO data will be used as the baseline data during the evaluation process.

CONTRIBUTION

The practical contribution of this study is the construction of a conceptual architecture for the MSS. Moreover, the output of this study will give public health care providers and professionals a tool to efficiently channel their resource allocation to appropriate malaria prone areas.

CONCLUSION

Social media presents a source of untapped unstructured data covering different subject areas that if mined, can provide gainful insights to public health practitioners for their decision-making activities. To this end, social media can provide a robust platform for public health practitioners to detect incidences and manage intervention measures in malaria prone areas.

FUTURE DIRECTIONS

This study is largely based on the analyses of literature review and provides a basis for the next steps of text mining and data collection process. In the next phase, we aim to develop an architecture model for collecting twitter feeds and proceed with data collection of twitter feeds related to malaria incidences. Further, we aim to use the design science research methodology by Hevner and Chatterjee (2010) and evaluate the results using positivist and interpretive methods of evaluation.

REFERENCES

1. Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H., & Liu, B. (2011). *Predicting flu trends using twitter data.* Paper presented at the Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on.
2. Asur, S., & Huberman, B. A. (2010). *Predicting the future with social media.* Paper presented at the Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01.
3. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*(Jan), 993-1022.
4. Buehler, J. W., Hopkins, R. S., Overhage, J. M., Sosin, D. M., & Tong, V. (2004). Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR Recomm Rep, 53*(Rr-5), 1-11.
5. Cao, J., Sturrock, H. J., Cotter, C., Zhou, S., Zhou, H., Liu, Y., . . . Gao, Q. (2014). Communicating and monitoring surveillance and response activities for malaria elimination: China's "1-3-7" strategy. *PLoS medicine, 11*(5), e1001642.
6. Chan, E. H., Sahai, V., Conrad, C., & Brownstein, J. S. (2011). Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS neglected tropical diseases, 5*(5), e1206.
7. Chen, L., Achrekar, H., Liu, B., & Lazarus, R. (2010). *Vision: towards real time epidemic vigilance through online social networks: introducing SNEFT--social network enabled flu trends.* Paper presented at the Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond.
8. Chew, C. (2010). *Pandemics in the age of twitter: A content analysis of the 2009 h1n1 outbreak.*
9. Cohen, J. M., Dlamini, S., Novotny, J. M., Kandula, D., Kunene, S., & Tatem, A. J. (2013). Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. *Malaria journal, 12*(1), 61.
10. Hevner, A., & Chatterjee, S. (2010). *Design Research in Information Systems: Theory and Practice*: Springer Publishing Company, Incorporated.
11. Kass-Hout, T. A., & Alhinnawi, H. (2013). Social media in public health. *Br Med Bull, 108*(1), 5-24.
12. Kelly, G. C., Hale, E., Donald, W., Batarii, W., Bugoro, H., Nausien, J., . . . Taleo, G. (2013). A high-resolution geospatial surveillance-response system for malaria elimination in Solomon Islands and Vanuatu. *Malaria journal, 12*(1), 108.
13. Khosa, E., Kuonza, L. R., Kruger, P., & Maimela, E. (2013). Towards the elimination of malaria in South Africa: a review of surveillance data in Mutale Municipality, Limpopo Province, 2005 to 2010. *Malaria journal, 12*(1), 7.
14. Lachlan, K. A., Spence, P. R., & Lin, X. (2014). Expressions of risk awareness and concern through Twitter: on the utility of using the medium as an indication of audience needs. *Computers in Human Behavior, 35*, 554-559.
15. Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B., & Suran, M. (2015). Detecting themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. *American journal of infection control, 43*(10), 1109-1111.
16. Littrell, M., Miller, J. M., Ndhlovu, M., Hamainza, B., Hawela, M., Kamuliwo, M., . . . Steketee, R. W. (2013). Documenting malaria case management coverage in Zambia: a systems effectiveness approach. *Malaria journal, 12*(1), 371.
17. Ohrt, C., Roberts, K. W., Sturrock, H. J., Wegbreit, J., Lee, B. Y., & Gosling, R. D. (2015). Information systems to support surveillance for malaria elimination. *The American journal of tropical medicine and hygiene, 93*(1), 145-152.
18. Olson, D., Konty, K., Mathes, R., & Paladini, M. (2013). Syndromic Surveillance from a Local Perspective: A Review of the Literature. *Online journal of public health informatics, 5*(1).
19. Othman, M. K., & Danuri, M. S. N. M. (2016). *Proposed conceptual framework of Dengue Active Surveillance System (DASS) in Malaysia.* Paper presented at the Information and Communication Technology (ICICTM), International Conference on.
20. Pak, A., & Paroubek, P. (2010). *Twitter as a corpus for sentiment analysis and opinion mining.* Paper presented at the LREc.
21. Runge-Ranzinger, S., Horstick, O., Marx, M., & Kroeger, A. (2008). What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends? *Tropical Medicine & International Health, 13*(8), 1022-1041.
22. Sharpe, J. D., Hopkins, R. S., Cook, R. L., & Striley, C. W. (2016). Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using Bayesian change point analysis: a comparative analysis. *JMIR public health and surveillance, 2*(2).
23. Thomson, M., Doblas-Reyes, F., Mason, S., Hagedorn, R., Connor, S., Phindela, T., . . . Palmer, T. (2006). Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature, 439*(7076), 576-579.
24. Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Icwsm, 10*(1), 178-185.

25. Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a framework for evaluation in design science research. *European Journal of information systems*, *25*(1), 77-89.