THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Digital Epidemiology: Forecasting Epidemics in the Era of Social Media and Mobile Devices Big-Data

Mohamed Khalil
Assistant Professor, Department of Mathematics,
October University for Modern Sciences and Arts (MSA), Egypt
Mohamed Said
Associate Professor, Department of mathematics,
October University for Modern Sciences and Arts (MSA), Egypt
Hesham Osman
Associate Professor, Department of mathematics,
October University for Modern Sciences and Arts (MSA), Egypt
Belal Ahmed
Ph.D. Student, Department of Electrical and Computer Engineering,
University of Victoria, Canada
Dalia Ahmed
B.Sc. Student, Department of Electrical Systems Engineering,
October University for Modern Sciences and Arts (MSA), Egypt
Naheel Younis
B.Sc. Student, Department of English and American Studies,
University of Vienna, Austria
Beshoy Maher
B.Sc. Student, Department of Electrical Systems Engineering,
Modern Sciences and Arts University (MSA), Egypt
Mohamed Osama
B.Sc. Student, Department of Electrical Systems Engineering,
Modern Sciences and Arts University (MSA), Egypt
Mohamed Ashmawy
B.Sc. Student, Department of Electrical Systems Engineering,
Modern Sciences and Arts University (MSA), Egypt

Abstract
The revolution of digital big data in the era of advanced technology has led to the rise of the field of digital epidemiology. The rapid development of social media platforms and mobile devices which are considered as the main sources of the digital data can produce huge data daily. Such big data can be used to track and monitor infectious diseases and to give a better understanding of diseases outbreaks. So, in this paper, we study the impact of social media and mobile devices on the field of digital epidemiology.

Keywords: *Digital epidemiology, big data, machine learning*

1. Introduction

In the last few decades, Infectious diseases have become a global threat which has terrible impact on human societies and global economy. Despite the great efforts done to control the spread of infectious diseases, pandemics like Zika and Ebola outbreaks remain in many countries. During the last decade, digital Epidemiology has become a significant field that can be used to handle the epidemiological big data (Salathé, M., 2018). Digital epidemiology is fueled by the rise of the social media, the rapid development of machine learning, and the widespread of smart mobile phones (Salathe, M. et al., 2015). Traditional epidemiology is based on collecting data by doctors and personnel in hospitals and public health organizations. On the other hand, digital epidemiology is based on collecting individuals' data through social media traces and mobile phones data to track the spread and dynamics of infectious diseases through studying people movement (Salathe, M. et al., 2015). The number of social media users is increasing exponentially. A huge amount of big data which contains crucial epidemiological information is produced daily by social media users. Such social media platforms can provide public health agencies with potential information about the spread and dynamics of diseases outbreaks in human societies (Chunara, R. et al., 2012). Social media like Twitter, Facebook and Google plus big data help decision makers in public health organizations to monitor the spread of infectious diseases and to detect outbreaks (Tripathi, M. et al., 2018). The most important challenge in digital epidemiology is the big data analytics to extract information from the gathered big data. It is a complex process to analyze millions of tweets and posts per day in social media networking platforms.

2. Infectious Diseases Detection Using Social Media Platforms

Social media platforms are essential part of our daily activities. People share information about the spread of infectious diseases and do serious discussions about public health issues. For example, the authors in (Chunara, R. et al., 2012) used Twitter data to track Cholera Outbreak in Haiti after the earthquake in 2010. They tracked Tweets containing the word "cholera". In (Hayate, I.S.O. et al.,2016) the authors tracked flu outbreaks using Twitter big data. They capture tweets containing words like "fever," "flu" and "coughing." Twitter big data gives a better understanding of the dynamics of infectious diseases and helps to predict the number of the future new cases.

3. The Impact of Mobile Phone Digital Big Data in Tracking Infectious Disease Transmission

The big data generated by cell phones can be used to track the movement of individuals to monitor the dynamics and spread of infectious diseases in human societies and to predict the number of infected people in an outbreak. In other words, Mobile phone data produces provides big data about the movement millions of people human to describe fluctuation in population density in order to forecast the diseases risk. For instant, the mobile data forms like routing tower location, call records, time of calls and SMS are used in (Wesolowski, A., et al., 2015) to study the spread of rubella disease. Another examples are In (Buckee, C.O., et al., 2013)and (Sacks, J.A., et al., 2015) at which mobile phone big data have been used to track the spread of malaria and Ebola virus respectively.

4. Machine Learning in Digital Epidemiology

Big data analytics are used to uncover the hidden information in big data (Mooney, S.J. et al., 2018). In other words, epidemiological big data analytics can extract smart data from the collected big data to give a better understanding of the impact of infectious diseases on human societies (Fung, I.C.H. et al., 2015). One of the big data analytics methods is the machine learning (Nagar, R. et al., 2014). It is defined as the computer's ability to forecast the future data based on the existing collected big data. So, in the field of digital epidemiology, machine learning gives computers the ability to learn from the existing data to forecast the future behavior of the epidemics outbreaks without being explicitly programmed. There are several algorithms that can be applied in machine learning filed like clustering and regression. For instant, regression which is a supervised machine learning is very important to predict future big data based on the existing data. For instant, the authors in (Hayate, I.S.O., et al., 2016) present a surveillance system to track influenza by forecasting words in Twitter. They implemented the following model to predict the number of infected individuals by influenza:

$$\hat{y}^{(t)} = x_1^{(t-\hat{\tau}_1)}\hat{\beta}_1 + x_2^{(t-\hat{\tau}_2)}\hat{\beta}_2 + \cdots + x_{|V|}^{(t-\hat{\tau}_{|V|})}\hat{\beta}_{|V|}$$

Where $\hat{y}^{(t)}$ presents the estimated number of infected people at time t, $x_{|V|}^{(t)}$ stands for the count of the word V at time t and $\hat{\beta}$is the weight estimated in the training phase, $\hat{\tau}_V$ represents the most suitable time shift width $\hat{\tau}_V$ for each word V.

5. Conclusion

Digital epidemiology is a promising field which is based on the digital big data generated by social media platforms and mobile phones. In this paper, we show how digital epidemiology can help to understand the dynamics of infectious diseases outbreaks through analyzing the produced big data to uncover the hidden information in the collected big data. From our point of view, public health organizations can find a better understanding of the dynamics of infectious diseases by developing new tools for big data analysis.

6. References

i. Buckee, C. O., Wesolowski, A., Eagle, N. N., Hansen, E., & Snow, R. W. (2013). Mobile phones and malaria: modeling human and parasite travel. Travel medicine and infectious disease, 11(1), 15-22.

ii. Chunara, R., Andrews, J. R., & Brownstein, J. S. (2012). Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. The American journal of tropical medicine and hygiene, 86(1), 39-45.

iii. Fung, I.C.H., Tse, Z.T.H. and Fu, K.W., (2015). The use of social media in public health surveillance. Western Pacific surveillance and response journal: WPSAR, 6(2), p.3.

iv. Hayate, I.S.O., Wakamiya, S. and Aramaki, E., (2016). Forecasting word model: Twitter-based influenza surveillance and prediction. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers (pp. 76-86).

v. Tripathi, M., Singh, S. and Soni Ghimire, S., (2018), Effect of Social Media on Human Health., Virology & Immunology Journal

vi. Mooney, S.J. and Pejaver, V., (2018). Big data in public health: terminology, machine learning, and privacy. Annual review of public health, 39, pp.95-112.

vii. Nagar, R., Yuan, Q., Freifeld, C.C., Santillana, M., Nojima, A., Chunara, R. and Brownstein, J.S., (2014). A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. Journal of medical Internet research, 16(10).

viii. Sacks, J.A., Zehe, E., Redick, C., Bah, A., Cowger, K., Camara, M., Diallo, A., Gigo, A.N.I., Dhillon, R.S. and Liu, A., (2015). Introduction of mobile health tools to support Ebola surveillance and contact tracing in Guinea. Global Health: Science and Practice, 3(4), pp.646-659

ix. Salathé, M., (2018). Digital epidemiology: what is it, and where is it going?. Life sciences, society and policy, 14(1), p.1.

x. Salathe, M., Bengtsson, L., Bodnar, T.J., Brewer, D.D., Brownstein, J.S., Buckee, C., Campbell, E.M., Cattuto, C., Khandelwal, S., Mabry, P.L. and Vespignani, A., (2012). Digital epidemiology. PLoS computational biology, 8(7), p.e1002616.

xi. Wesolowski, A., Metcalf, C.J.E., Eagle, N., Kombich, J., Grenfell, B.T., Bjørnstad, O.N., Lessler, J., Tatem, A.J. and Buckee, C.O., (2015). Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. Proceedings of the National Academy of Sciences, 112(35), pp.11114-11119.