Introduction

In the past year, with the unstoppable global outbreak of COVID-19, many countries have adopted travel bans to prevent the further spread of infectious diseases. Undoubtedly, the movement of people will accelerate the transmission of infectious diseases. However, mobility data describes the mobility behaviour of people in a particular area, such as the journey-to-work data in a country. Otherwise, not only efffectively reflects human behavior and activities, movement data also contains a wealth of information on user behavior patterns and important location attributes. Due to the development of mobile application technology and spatial data acquisition technology, it is not difficult to obtain user’s trajectory data. Zheng (2015) proposed the entire process of trajectory data mining as well as techniques applied in each step [2]. In addition, social interactions are another concerned factor that promotes the spread of diseases. Past research has investigated the attributes of social encounters by conducting a telephone survey [3].

Moreover, great understanding of the relationship between infectious disease transmission and the trajectory of human movement is critical for predicting and preventing future outbreaks of infectious diseases. And, effective public interventions can be designed in a timely manner to control the further spread of infectious disease in this case [1]. Much research has been done to explore the association between the transmission of infectious diseases and geographical or demographic indicators (such as per-head income) at different scale (city, urban, country etc.) [1,4]. Furthermore, the research incorporating trajectory data with social encounters have been conducted to reveal the spatial distribution of infection depending on specific locations [6].

Here, we will combine the trajectory data with social encounters to explore the spatial distribution of infectious disease. In this process, cluster algorithms, which are a group of efficient algorithms to do unsupervised machine learning, will be applied to find general patterns in movement dataset. Then, the social interaction of participants and basic information (such as age etc.) will be combined with movement patterns to reveal the law of disease distribution.

1. Chowell G, Bettencourt LMA, Johnson N, Alonso WJ, Viboud C. The 1918–1919 influenza pandemic in England and Wales: spatial patterns in transmissibility and mortality impact. Philos Trans R Soc B Biol Sci. 2008;275(1634):501–9. <https://doi.org/10.1098/rspb.2007.1477>.
2. Zheng, Y. (2015). Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *6*(3), 1-41.
3. Rolls, D. A., Geard, N. L., Warr, D. J., Nathan, P. M., Robins, G. L., Pattison, P. E., ... & McVernon, J. (2015). Social encounter profiles of greater Melbourne residents, by location–a telephone survey. *BMC infectious diseases*, *15*(1), 1-11.
4. Murray, C. J., Lopez, A. D., Chin, B., Feehan, D., & Hill, K. H. (2006). Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918–20 pandemic: a quantitative analysis. *The Lancet*, *368*(9554), 2211-2218.
5. Viboud, C., Bjørnstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A., & Grenfell, B. T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *science*, *312*(5772), 447-451.
6. Moss, R., Naghizade, E., Tomko, M., & Geard, N. (2019). What can urban mobility data reveal about the spatial distribution of infection in a single city?. *BMC public health*, *19*(1), 1-16.

Techniques