Travel Volume Modelling for Infectious Disease Transmission Risks Assessment

by

Zichuan Huang

School of Information Technology and Electrical Engineering,
University of Queensland.

Submitted for the degree of Master of Data Science
in the division of 2020 – 2022

April 28, 2022

Zichuan Huang

zichuan.huang@uq.net.au

April 28, 2022

Prof Amin Abbosh

Acting Head of School

School of Information Technology and Electrical Engineering

The University of Queensland

St Lucia   QLD   4072

Dear Professor Abbosh,

In accordance with the requirements of the Degree of Master of Data Science in the School of Information Technology and Electrical Engineering, I submit the following thesis entitled

  "Travel Volume Modelling for Infectious Disease Transmission Risks Assessment"

The thesis was performed under the supervision of Prof. Hele Huang. I declare that the work submitted in the thesis is my own, except as acknowledged in the text and footnotes, and that it has not previously been submitted for a degree at the University of Queensland or any other institution.

Yours sincerely

Zichuan Huang

Contents

Chapter 1

Summary

In this proposal, I will introduce how to use travel volume modelling to evaluate the risks of infectious disease transmission. This technique is importance, and this is because of the connection between countries are globalizing the economy and transportation. Thus, the risk of transmission of a series of infectious diseases will become higher. Therefore, it is necessary to use traffic information for risk assessment. My main goal in this project include: firstly, taking other key factors into account, such as the susceptibility of different regions; secondly, visualizing the results for ease of use by passengers. Therefore, in order to implement this topic, some data mining technics will be used, such as page rank, SIR infectious disease dynamics model and stochastic formulations. After a brief introduction to this proposal, the following mainly introduces the research background of this project.

Chapter 2

Background

Currently, because of international travel, the spread of COVID-19 has become the norm. This is because airborne, foodborne, and zoonotic infectious diseases transmitted during air travel, and this raise important public health concerns as air travel becomes more accessible and affordable. Likewise, the SARS outbreak in 2002 showed that air travel can play and important role in the rapid spread of new infections and may even lead to epidemics(Mangili & Gendreau, 2005). Therefore, I explain that it is necessary to estimate the future infection risk through the flight volume in order to prevent the infection during travel. I will introduce two different methods in the next paragraph.

In response to this situation, various countries and institutions have also developed corresponding risk assessment mechanisms. Since 2007, the European Center for Disease Control and Prevention has launched the Risk assessment guidelines for infectious diseases transmitted on aircraft(RAGIDA) (Risk assessment guidelines for infectious diseases transmitted on aircraft, 2009). There is a series of algorithms behind the RAGIDA to support this project to be able to predict and prevent a range of infectious diseases such as anthrax and SRAS. Similarly, a data-driven practical method for predicting the evolution of the new coronavirus pneumonia epidemic have been also developed by some researchers in China. Based on moving average prediction limit(MAPL), this project uses the previous SARS epidemic data to verify the practicability of the MAPL method for epidemic trend and risk prediction(Hao et al., 2020). The official data of COVID-19 from January 16, 2020 was tracked to establish the corresponding MAPL for timely epidemic prediction and risk assessment. In the he next paragraph, I will introduce some of the remaining problems.

Chapter 3

Problem identification

Although several projects have now investigated this issue, there are still two issues that need to be addressed. The first problem is that some models only consider the effect of connectivity between regions on infection risk. However, the spread of the epidemic also depends on the susceptibility of countries and regions with different latitudes and longitudes, and the sensitivity of local people to certain infectious diseases. Moreover, when implementing the actual infectious disease risk prediction method, it is also necessary to evaluate the situation in different regions to predict the risk. Therefore, more factors need to be taken into our consideration. The second problem is that currently, there is a lack of visual implementation of these models. This is very necessary because infectious diseases affect every passenger. Moreover, in today's convenient transportation environment, it is necessary to make it easy to obtain epidemic risks information. To solve these issues, I propose some objectives for the project.

Chapter 4

Objectives

To implement this method, I will build a system that enable to do the following things:

- base on international air-travel volume, design and develop models for predicting the transmission of new infectious pathogens to and from different countries.
- build a user interface dashboard, which enable to update the travel data by the users.
- build an interactive visualization system that will use different data structures and graphic designs to help the users to explore the relationships, which would influent the transmission risks.

In order to achieve several objectives mentioned above, I will introduce the solution below.

Chapter 5

Proposed solution

This project will be implemented in two parts. The first one is the risks assessment model. In this part, each region of the world is regarded as a node, and each node is connected by the air-travelling. In this way, the whole world can be seen as a world-wide airport network(*WAN*). Thus, the page rank algorithm can be used to evaluate the important of each node by the weight of each edge, which is flux of the flight line between two nodes. The reason for using this algorithm is that it takes into account the influence that neighboring countries have on a country. The second part is the visualized dashboard. This would help user to explore the current data by various types of graphs, which could assist to discover insightful results. Moreover, it could also be a real-time epidemic risks map for users as a reference. More detailed steps will be explained in the next few paragraphs.

In order to implement this project, I will take following steps. First, relevant data needs to be collected, including demographic, healthcare, public health, disease dynamics, political domestic, political international and economic(Cadavid, Furuya-Kanamori, Mayfield, Nilles & Lau, 2021). Second, I will use the data as the evaluation criteria to evaluate the risks of different countries. Third, apply PageRank algorithm to the graph network, and make iterations to build the model. Finally, I will build up and test the user interface dashboard. In the next few sections, I will introduce the second and third steps in details.

5.1 SIR infectious disease dynamics model

In each node, I will apply susceptible-infected-removed (SIR) model to evaluate the risk of the local epidemic. Supposed that each node has N population, and the N is given by $N = S(t) + I(t) + R(t)$, where $S(t)$, $I(t)$, $R(t)$ represent the number of people who are in the states of susceptible, infected and recovered at time t. In order to evaluate the epidemic situation, there are two parameters in this model that we need to acquire. The first parameter is β, which represents the probability of disease transmission. The second parameter is γ, which represents the recovery rate(Colizza, Barrat, Barthelemy & Vespignani, 2018). By using the optimize method, we can get the final parameter and

calculate the basic reproduction number $R_0 = \beta/\gamma$. The basic reproduction number refers to the average number of how many other people a person infected with a certain infectious disease will infect without external intervention and no one is immune. This is an important indicator of the contagiousness of the disease. If $R_0 < 1$, the infectious disease will gradually disappear, and if $R_0 > 1$, the epidemic will spread exponentially and cause a contagion outbreak("Compartmental models in epidemiology", 2022). To verify the feasibility of this model, I applied the SIR model on a small dataset and I will explain it in next paragraph.

The dataset used here recorded the dates from January 18, 2022 to January 22, 2022, and the number of recovered, dead, infected number of each date in Wuhan, China(Table 5.1). By calculating the β, we can get a result of $\beta = 0.03155$, $R_0 = 2.2083$, given that each infected individual contacts five people per day(nContact = 5). Also, we can make a prediction with the SIR model by different numbers of people that a infected individual contacts, the result is shown in Fig 5.1 to Fig 5.4.

Table 5.1: The dataset of the epidemic in Wuhan

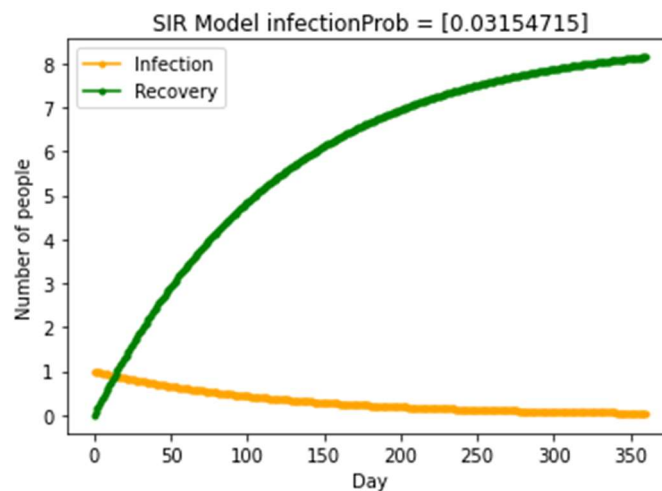| Date	Recovered	Death	Infected
January 18	3	0	4
January 19	4	0	17
January 20	6	2	136
January 21	0	2	60
January 22	3	11	167

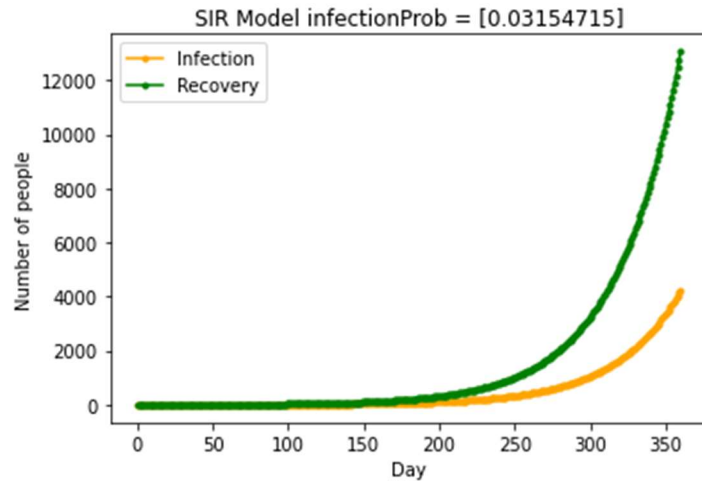Figure 5.1: The prediction results of SIR model given by nContact = 2

Figure 5.2: The prediction results of SIR model given by nContact = 3
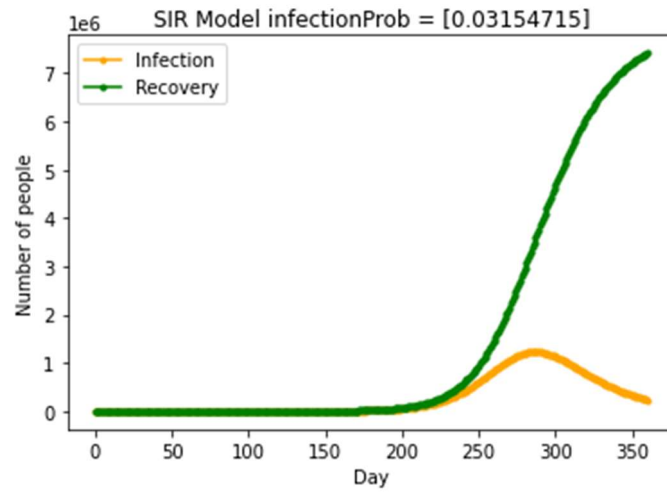

Figure 5.3: The prediction results of SIR model given by nContact = 4
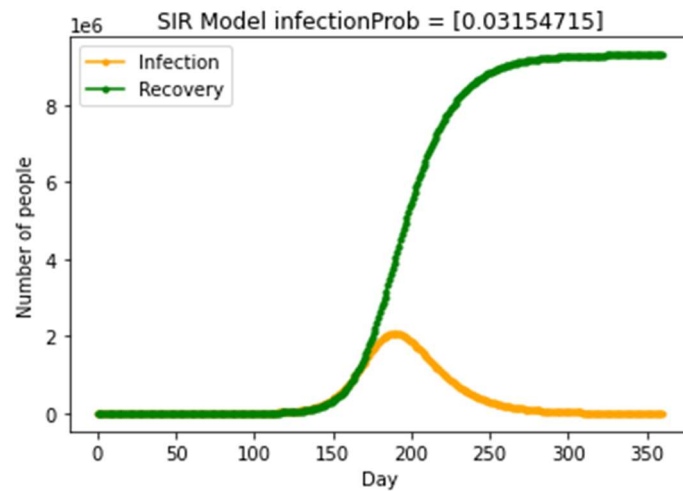

Figure 5.1: The prediction results of SIR model given by nContact = 5

From the result shown above, we can see that with the reduction of nContact, the

number of sick people decreases sharply. When nContact = 2, the number of sick people does not even exceed 100. It can be seen that the adoption of strict control and prevention and isolation measures has a great effect on the reduction of the number of sick people with the new coronavirus. very obvious effect.

5.2 PageRank algorithm

At the global level, I will use the PageRank algorithm to calculate the importance of each region through a network of flight paths. PageRank algorithm is originally used to rank web pages in the search engine results(PageRank, 2022), here we treat each airport as a node, and the flight routes between them constitute edges, with the flux as the weights. Finally, through iteration, we can get the importance of each node.

Chapter 6

Conclusion

In conclusion, this system will give corresponding risk prediction information based on the data of different infectious diseases given by the user and perform it to the user in a visual way. However, this method also got some disadvantages. First, the impact of the global flight network to each region (node) needs more research to figure out. Second, some infectious diseases have the incubation period, so the SIR model may not be able to capture the characteristics of these infectious diseases very well.

Bibliography

Cadavid Restrepo, A., Furuya-Kanamori, L., Mayfield, H., Nilles, E., & Lau, C. L. (2021).
Implications of a travel connectivity-based approach for infectious disease transmission
risks in Oceania. *BMJ open*, *11*(8), e046206. doi:10.1136/bmjopen-2020-046206

Colizza, Barrat, A., Barthelemy, M., & Vespignani, A. (2006). The Role of the Airline
Transportation Network in the Prediction and Predictability of Global Epidemics.
Proceedings of the National Academy of Sciences - PNAS, 103(7), 2015–2020.
https://doi.org/10.1073/pnas.0510525103

Compartmental models in epidemiology. (2022). Retrieved from
https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology#The_SIR_model

He, H., He, Y. T., Zhai, J., Wang, X. J., Wang, B. S. (2020). Predicting the trend of the COVID-19
outbreak and timely grading the current risk level of epidemic based on moving average
prediction limits. *Journal of Shanghai Jiaotong University(Medical Science)*, *4(40)*, 442-
449. doi: 10.3969/j.issn.1674-8115.2020.04.002

Mangili, A., & Gendreau, M. A. (2005). Transmission of infectious diseases during commercial
air travel. Lancet (London, England), 365(9463), 989–996. https://doi.org/10.1016/S0140-
6736(05)71089-8

PageRank. (2022). Retrieved from https://en.wikipedia.org/wiki/PageRank

Robert Koch Institute.(2009). *Risk assessment guidelines for infectious diseases transmitted on
aircraft*. Retrieved from European Centre for Disease Prevention and Control:
https://www.ecdc.europa.eu/en/publications-data/risk-assessment-guidelines-infectious-
diseases-transmitted-aircraft