VE444 — Networks

A Multi-Network Based Framework using SIR pandemic model and Graph Neural Network tracking the COVID-19 Impact in United States

Project Proposal

Instructor: Yifei ZHU

Group 5 — UM-JI (Fall 2020)

Group members

- Liqin Zhang (517370910123)
- Siyin Ma (517370910003)
- Zhejian Jin (517370910167)

1 Introduction

1.1 Motivation

Researches on infectious diseases has always been an important topic in interdisciplinary fields. The spread of infectious diseases not only depends on the attributes of diseases itself, but also depends on the network structure of the crowd. Since late 2019, COVID-19 has become a worldwide pandemic. In order to better understand and control the spread of COVID-19, many researchers collaborated and modeled this process to forecast its influence on the neighbouring areas.

SIR model is a common compartmental model that well-defines disease transmission dynamics. However, it assumes a relatively closed system, and it does not take interactions among regions into consideration. Based on this, we utilize Graph Neural Network using data that describes inter-region interactions, which captures the features of messages among the regions.

In this work, we propose a novel framework for COVID-19 case examination and prediction that uses Graph Neural Network. Different with existing time series models, this model learns from a spatio-temporal graph. We then evaluate this model on the California COVID-19 data set.

2 Project Design

2.1 Tentative Design

In this project, we combine two powerful models (SIR and GNN) into one integrate framework to perform better prediction capabilities in COVID-19 cases with lower cost. We propose this general framework that learns to select relevant edges and graph representations by pre-computing results from SIR model, along with RMSLE and Pearson values. The major flowchart of our framwork is shown in Figure 1.

1. The SIR pandemic model
   SIR model is a simple method describing the epidemic spread that can be applied to any network structures. It divides the popularity into three categories, representing three stages of a person. The three functions with time t is used to simulate the three compartments.
   (a) Susceptible $S(t)$: people before getting infected, but can be infected by neighbors.
   (b) Infectious $I(t)$: people who have been infected, and can infect others.
   (c) Removed $R(t)$: people who endured a complete infection cycle and are immune to the disease.

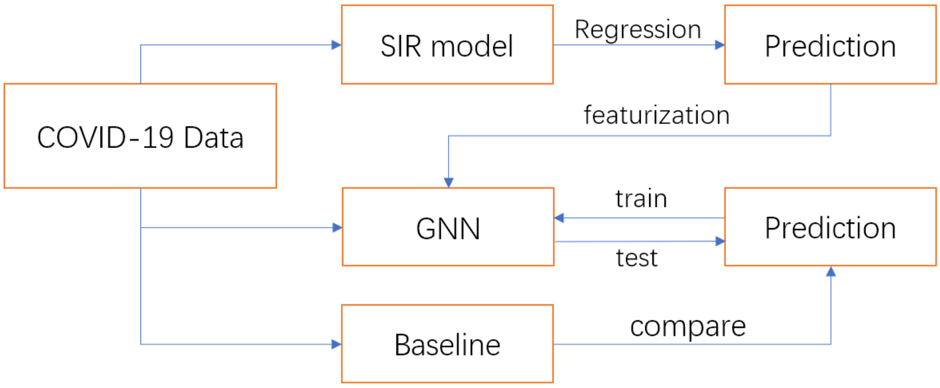

Figure 1: Flowchart of the framwork

The number of population is $N(t)$, where $N(t) = S(t) + I(t) + R(t)$. The SIR model describes the change between the three parts of people. We introduce two parameters β and γ.

- β Effective contact rate
- γ Recovery rate

To be more specific, β describes number of susceptible persons that a patient can infect per unit time is proportional to the total number of susceptible persons in the environment. γ implies that the number of people removed from the infected person per unit time is proportional to the number of patients. Three differential equations are used to sketch the model:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\beta SI}{N}$$
$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\beta SI}{N} - \gamma I$$
$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I$$

We have to estimate the number of β and γ, where gradient descent or other methods would be applied to find the best parameters that fit the data of COVID-19.

2. The Graph Neural Network(GNN)

GNN is a suitable model to disseminate the geographical information as well as the human mobility across location to address the preliminary prediction with a certain amount of factors taken into account. In order to imitate infectious disease modeling, we plan to take create a graph with multiple nodes representing the states of a country and edges defining the spatial and temporal dependencies among states. We present our graph as a multi-layer stack, where each layer manifest the county connectivity graph for one day. Similar as the message-passing framework, we define the update at layer l as

$$\mathbf{m}_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \mathcal{F}^{(l)}\left(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}\right), \quad \mathbf{h}_i^{(l+1)} = \mathcal{G}^{(l)}\left(\mathbf{h}_i^{(l)}, \mathbf{m}_i^{(l+1)}\right)$$

where $\mathcal{F}^{(l)}$ and $\mathcal{G}^{(l)}$ are learned message functions and node update functions respectively, $\mathbf{m}^{(l)}$ are the messages passed between nodes, and $\mathbf{h}^{(l)}$ are the node representations. Notice that the bottom layer representing the date cases began appearing in the US and the top layer representing the newest date in our dataset.

2.2 Method Evaluation

All the methods we use (SIR model, GNN) are trained to predict the change in number of cases on day $t+1$ given the previous date information, since we have data from the first date onwards, we would use first 70% for training and last 30% for testing. Then we will compare the result from the SIR model regression and the forecasting performance of our GNN with a large number of baseline models. We will than calculate the root mean square logarithmic error (RMSLE) and the absolution pearson correlation coefficient to measure the accuracy of our prediction. The final value will be the average across all states for the trainging days of inference.

2.3 Expected Results

- Bottom Line: Implement the SIR model and Graph Neural Network for the Covid-19 epidemic situation based on real data. Compare the two models' realization and their accuracy in prediction.
- Expected: Develop mechanism to combine the two models, SIR and GNN together, where the result of SIR can enhance the robustness of GNN.

3 Dataset

We will make use of two dataset, the first one is the New York Times COVID-19 dataset, which tracks cases of coronavirus in real time and provide information about mask usage in the United States, the other aggregated mobility dataset helps better represent the quantity of movement.

1. Coronavirus (Covid-19) Data in the United States provided by New York Times[3]
2. The Google COVID-19 Aggregated Mobility Research Dataset[4]

4 Task Assignment

Name	Assignment
Zhang Liqin	Build the graph; Train the Spatio-Temporal-GNN model; Data Visualization;
Jin Zhejian	Study and implement the main regression algorithms; build the prediction system.
Ma Siyin	Data Extraction and Data cleaning; SIR model implementation.

Table 1: Task Assignment of VE444 Project, Group 5

5 Sources

1. https://towardsdatascience.com/graph-convolutional-nets-for-classifying-covid-19-incidence-on-states-3a8c20ebac2b
2. Kapoor, Amol, et al. "Examining covid-19 forecasting using spatio-temporal graph neural networks." arXiv preprint arXiv:2007.03113 (2020).
3. New York Times COVID-19 dataset, https://github.com/nytimes/covid-19-data
4. Google COVID-19 Aggregated Mobility Research Dataset, https://www.google.com/covid19/mobility/
5. https://scipython.com/book/chapter-8-scipy/additional-examples/the-sir-epidemic-model/