COVID-19 Virus Prediction of US

Yu Ren
Jing Bian
Jinnuo Che

1. **Motivation**

   Since the end of 2019, a new type of coronavirus pneumonia has erupted in China, and the epidemic has swept the country. It is another national spread of viral pneumonia after SARS. Now, outbreaks are happening all over the world, we plan to analyze the trend of the epidemic objectively in America so as to prevent and avoid the virus effectively.

2. **Introduction**

   We will use SIR model to model the number changes of confirmed and recovered people in China. This model will contain two important parameters, namely, β (rate of infection per susceptible and per infective individual) and γ (rate of recovery) parameters. We will process "COVID19-China.csv" with Bayes and pseudo maximum likelihood, and calculate Chinese β and γ parameters. Based on social comparisons, we can estimate American β parameter so that we can predict what the situation will be in USA.

3. **Methodology**

   Algorithms and Methods of machine learning, which you plan to use to solve your objectives for your project:

   A. SIR model
      a. *S(t)* are those susceptible but not yet infected with the disease;
      b. *I(t)* is the number of infectious individuals;
      c. *R(t)* are those individuals who have recovered from the disease and now have immunity to it.
   B. Pseudo-Maximum-likelihood (PML)
   C. Gaussian Naive Bayes

4. **Specifications or Description of your dataset**

   "COVID19-China.csv" is a dataset containing 70354 pieces of data about the virus situation changes in different cities and provinces in China. Each data contains province information, city information and the numbers of confirmed, suspected, cured and dead people in this city and this province.

5. **Data Sources**

   https://www.kaggle.com/wang749/china2019ncov