SEIRV MODELING OF COVID 19

Team Name: Omicrons

Team Members: 1. Shubham Wasnik (17894) 2. Palash Kamble (17853)
3. Aakash Singh (18136) 4. Biswajit Mondal (18022)

Project Name: Covid 19 Prediction

1. INTRODUCTION

Corona viruses are a large family of viruses that are known to cause illnesses ranging from the common cold to more severe diseases such as severe acute respiratory syndrome (SARS). The severe acute respiratory syndrome 2 (SARS-Cov-2) was identified as the cluster of pneumonia cases in Wuhan, China at the end of 2019. It subsequently spread throughout China and elsewhere, becoming a global health emergency. In February 2019, the World Health Organization (WHO) designated the coronavirus disease 2019 (COVID-19) a global pandemic. In the absence of a ready-to-use vaccine, and besides medical and biological research, mathematical models can play an important role in understanding and predicting disease transmission. Moreover, it helps to implement appropriate measures and efficient strategies to control the pandemic's spread and mitigate its impact. Several mathematical models like SIR, SEIR have been developed to study transmission of a disease in a population. However the models suffer from various sources of uncertainties like incomplete description of biological processes governing the disease spread and also due to some parameters being poorly known. Our work aims to study the more detailed mathematical modeling from the time-series data to make future predictions. In the modeling we incorporate the effect of vaccinations and mobility amongst 38 districts.

2. METHODS

SEIRV Model: In this study, a mathematical model of the spread and transmission of SARS-CoV-2 was formulated. The model subdivides the total human population size at time t denoted as $N(t)$ into susceptible $S(t)$, exposed $E(t)$, infectious $I(t)$, and recovered as $R(t)$. Hence for the human population we have $N(t) = S(t) + E(t) + I(t) + R(t)$. If an individual is in the susceptible state, we can assume they are healthy or their antibodies are waned. If they are in the exposed state, they have been infected with the virus but are not infectious. If they are infectious they can transmit the disease, and we assume that they ultimately recover. Once an individual is in the recover state, we assume they go the susceptible state due to antibody waning either by taking vaccine or not taking vaccine.

The disease transmission flow of the proposed model is sketched in Figure 1.

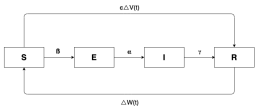

Fig. 1.

The model is then governed by the following set of differential equations:

- $\triangle S_i(t) = -\beta_i(t) S_i(t) \Sigma_{j \in U} M_{(i,j)}(t) \frac{I_j(t)}{N_j} - \epsilon \triangle V_i(t) + \triangle W_i(t)$
- $\triangle E_i(t) = \beta_i(t) S_i(t) \Sigma_{j \in U} M_{(i,j)}(t) \frac{I_j(t)}{N_j} - \alpha E_i(t)$
- $\triangle I_i(t) = \alpha E_i(t) - \gamma I_i(t)$
- $\triangle R_i(t) = \gamma I_i(t) + \epsilon \triangle V_i(t) - \triangle W_i(t)$

  where, α^{-1}=Incubation period, β=Contact Rate, γ^{-1}=Recovery period, ϵ=Efficacy, N_j = Total number of population in district j

2.1. Data Pre-processing

2.1.1. Handling missing vaccinations

We are given vaccination details for 31 districts. In that except Bengaluru-Urban, all the vaccination details are kept as it is. To handle the missing data for 9 districts, we added the population of those 9 districts in Bengaluru-Urban and split them in proportion for 9 districts. Further we are given vaccination details upto 2nd September, so we are projecting the vaccination details further as follows: Suppose the vaccination numbers of last five days are v1,v2,v3,v4,v5, then we take the average of difference of the ith day vaccinations and

(i+1)the day vaccinations. We add this average to v5 to get v6, and so on.

2.1.2. *Initialization of S0, E0, I0, R0*

We got S0, E0, I0, R0 for each district from sero-prevalence and timeseries data as follows
$S0$: (Total population of district) - E(0) - I(0) - R(0)
$E0$: %Active Infection * (Population of District) * 0.1
$I0$: %Active Infection * (Population of District) * 0.9
$R0$: %IgG * (Population of District)

2.2. Modeling Antibody Waning

Antibody waning is modeled by weibull distribution. Let the daily recovered fraction is given by $\triangle R(t)$ at time t, pdf of weibull distribution is given by $p(t)$ (modeling the decay of antibodies in an individual with respect to time 't') and the population with the antibodies less than detectable levels can be given as $\triangle W(t)$ then:
$\triangle W(t) = \sum(\triangle R(t - tau) * p(t))$, where $t > tau >= 0$
$p(t) = \frac{k}{\lambda} * (\frac{t}{\lambda})^{k-1} * e^{-(\frac{t}{\lambda})^k}$, where $t >= 0$, $\lambda = scale$, $k = shape$

2.3. Loss Function

- For 'per day': We take the mean squared error of (seven day averaged) $\triangle I$ from the daily infection data and (seven day averaged) predicted $\triangle E$ times alpha, in the log domain, where $\triangle E$ is equal to $\frac{E(t)}{CIR(t)}$. Averaging is done over the asked time period for all the districts separately.
  loss = mean($(\log(true\overline{\triangle I}) - \log(\alpha * predicted\overline{\triangle E}))^2)$))

- For 'last day': We find the percentage error in the true and predicted values.
  loss = $|\frac{(true\overline{\triangle I} - \alpha * predicted\overline{\triangle E})}{true\overline{\triangle I}}|$

2.4. Optimization

First we use grid search to get a coarse estimate of the contact rate parameters. Then we did the gradient descent to get the final optimized values of the contact rate parameters. Because of the mobility matrix M, there was coupling between the betas, i.e. modifying the beta of one district was modifying the loss of other districts. We noticed that changing the beta of one district was affecting the losses of not just its own district but also other districts. We assume that the coupling was not so strong that hence we optimized each loss independently. The update rule is given as:
$p(j+1) = p(j) - \frac{1}{j+1} * \partial_p loss(p(j))$, where $p \in P$, $j >= 0$

3. RESULTS

In part1 we were initializing the initial values of S0, E0, I0, R0 from the round1 seroprevalence data and running forward the simulator (evolving the SEIR equations) from 11 oct 2020 to 31 oct 2020 such that the daily infected predicted cases matches with the actual time series data within 10% absolute error. We tried some novelties on mobility matrix (scaled diagonal entries population wise) and the plots for the actual daily infected cases from the time series data (called as true) and predicted daily infected cases vs days ranging from (11 oct 2020=0 , 31 oct 2020 = 20) are shown in figure 2.

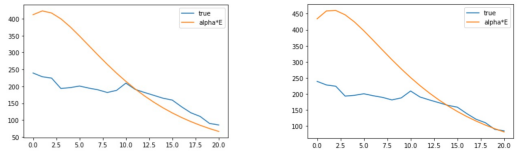

Fig. 2. Left Image: Mobility Matrix as Identity Matrix, Right Image: Modified Mobility Matrix. District: Ballari. On x-axis=days, y-axis=daily infections

In part2 the per day squared error loss is minimized between the predicted daily infections and that of reported from the timeseries data in the duration 1 nov 2020(0) to 28 feb 2021(140) and the plots are shown in figure 3.

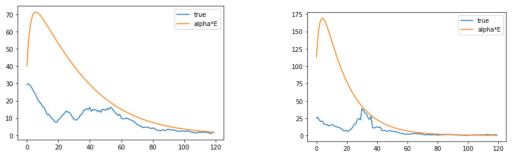

Fig. 3. Left Image: Bidar, Right Image: Hassan. On x-axis=days, y-axis=daily infections

4. CONCLUSION

We wanted to fit the SEIRV model to the time series data by tuning the contact rate parameters for 38 different units in Karnataka. We extracted the initial parameters from round1 seroprevalence data and ran the predictions forward.

In the first part of the project we successfully reduced the loss to below 10% for all the districts. We can see from the plots that the predictions on 1 nov 2020 are very close to actual confirmed cases i.e. on 1 nov 2020.

We thought that mobility within one unit may depend on the population of that unit. Hence we tried scaling the diagonal entries of mobility matrix by the proportion of the populations of individual units. Let the mobility matrix be M, then : $M(i,i) = p(i)/P$, where P is maximum population across all units. We can see the results have improved and the

predictions are more closely following the actual confirmed cases.

For the second part of the project we had to minimize the per day squared error below 0.1. For many of the districts, the loss indeed converged below 0.1, but for some districts it did not converge.That might be because of the coupling of the contact rate parameters among the units.