One-Week Predictions for Districts of Karnataka

Indian Institute of Science, Bangalore,
Indian Scientists' Response to COVID-19, and
Indian Statistical Institute, Bangalore
Jawarhalal Nehru Centre for Advanced Scientific Research
and Council for Scientific and Industrial Research, Bangalore.

1 Introduction

In this file, we describe four prediction models (INDSCI-SIM Prediction, IISC-ISI Prediction 1, IISC-ISI Prediction 2, (JNCASR-IISC)SAIR-Prediction) to predict the infection counts for each district of Karnataka for the next one week. INDSCI-SIM Prediction, IISC-ISI Prediction 1, are SIR model based predictions. IISC-ISI Prediction 2 is based on regression in log scale. (JNCASR-IISC)SAIR-Prediction is based on a SAIR model CSIR-Model is based on Neural Network algorithm. Please refer to the Section 2 for details of each fit. The file includes-

- In Section 2- A description of the models used for the predictions.
- In Section 3- A list of our data sources and the repository created.

2 Methods

INDSCI-SIM Prediction

The model is a metapopulation epidemiological model that contains 9 compartments for each population. It is defined at the level of each district in India, but has also been used for single cities as well as wards within those cities. The compartments are: susceptible, exposed, asymptomatic, presymptomatic, mildly symptomatic, severely symptomatic, hospitalized, recovered and dead. Each compartment is age-structured into 7 age brackets, transition rates between compartments depend, in general, on these age brackets and contacts between these compartments are accounted for using contact matrices derived from social surveys. Migration between metapopulations can be incorporated into the model in two ways, using a network approach or using a gravity model. The model is fit to data using a joint fit to the time-line of the numbers of deaths and the numbers of cases, up to a multiplicative factor that accounts for the undercounting of cases. The initial numbers of infected, an unknown quantity, as well as the infectivity (for which approximate initial estimates can be obtained using Bayesian methods implemented in the package EpiEstim), are varied to obtain a band in which the best-fit projected time series of infections, deaths etc. lies, thus accounting for uncertainties in model inputs. All parameters such as rates that enter the model are benchmarked to what is known about COVID-19 disease progression from studies in China, Italy and the USA, but can be further refined to represent the Indian situation better once data becomes available. The parameters entering the model can depend on time, accounting for measures such as lockdowns and their relaxation, as well as other non-pharmaceutical interventions such as testing and quarantining at rising intensities. The projections of the model have successfully predicted the course of the pandemic in a number of Indian cities so far and can be trusted to provide accurate short-term estimates for the daily incidence of cases and related quantities.

The toolkit detailing the methods of the second prediction model can be found on : https://cmstest.unipune.ac.in:5000/

IISC-ISI Prediction 1 This prediction is based on a SIR model. In this model, the entire population is categorized into three states- Susceptible, which refers to healthy individuals at a risk of contracting the disease; Infected, which refers to the people infected with the disease; and Recovered, which refers to people who have already contracted and then recovered from the disease. At any time instant, an individual of the population falls in one and only one of the above three categories. The progression from one category to another can be schematically represented by- $S \rightarrow I \rightarrow R$. Once a patient has recovered patient then cannot become susceptible or infected. Hence this model assumes that recovery gives an individual complete immunity from the disease. Let the fraction of susceptible in the population be given by $s(t)$, fraction of infected people by $i(t)$ and recovered fraction by $r(t)$. The dynamics of s, i, r are governed by the following coupled ordinary differential equations:-

$$\frac{ds}{dt} = -\beta i(t)s(t)$$

$$\frac{di}{dt} = \beta i(t)s(t) - \nu i(t)$$

$$\frac{dr}{dt} = \nu i(t)$$

When the fraction of infected and recovered people is much smaller than the number of susceptible people (population is very large), one can assume that $s(t) \equiv 1$. The above differential equations can get simplified to:-

$$\frac{di}{dt} = (\beta - \nu)i(t)$$

$$\frac{dr}{dt} = \nu i(t)$$

The can be solved expicitly now to give the following solution-

$$i(t) = i(0)\exp\{(\beta - \nu)t\}$$

$$r(t) = \frac{i(0)\nu}{(\beta - \nu)}[\exp\{(\beta - \nu)t\} - 1]$$

$$i(t) + r(t) = i(0)\frac{\beta\exp\{(\beta - \nu)t\} - \nu}{\beta - \nu}$$

To fit the data to our model, we consider the time series $i(t) + r(t)$ and perform a one-week ahead prediction at each time t. For each t, the prediction is done as follows. Start with $x = 7$. At time t, we look at data from $t - x$ to t, does a least square fit for the above SIR model (with $\nu = 0.1$ and $s(t) = 1$ assumptions) and predicts the data point at $t + 1$ from this fit (i.e. for times $t + 1$ through $t + 7$). Then choice of x is then further optimised in the following fashion. In each step, we change x to $x + 1$, perform the fit from $t - (x + 1)$ to t, compute the average error on the $x + 1$ training samples. If this error is larger than the error in the previous step, then we stop and pick that value of x. For the final data point we extend the least square fit for seven future days.

The error-band for the seven day prediction is computed as follows. For each data point at time t we use the above method to provide a fit for $t+1, \ldots, t+7$-th days ahead. We then subtract each fit value from the corresponding data at $t+1, \ldots, t+7$,-th days to get the error for each of the seven days. We then compute the root-mean-square error (over t) for each of seven days over all the data points. These seven standard errors are plotted in the error band plot.

IISC-ISI Prediction 2

Here we consider log of number of cases on each day. We use an automated window to find the best least-square fit as described in IISC-ISI Prediction 1. We also use the same procedure for providing the error band for the one week future prediction.

SAIR- Prediction 4

These are based on papers [1] and [2].

We have modelled the spread of COVID-19 using an SIR-type model, wherein the population is divided into Susceptible (S), Asymptomatic (A), Infected (I), Recovered (R) and Deceased (D). The system of equations describing the time dynamics of the population is

$$\frac{dS}{dt} = -\beta(t)S\ (A+I)\,, \quad \frac{dA}{dt} = \beta(t)S\ (A+I) - \delta A - (\gamma - \gamma_d)A, \quad \frac{dI}{dt} = \delta A - \gamma I, \\ \frac{dR}{dt} = \gamma(A+I), \qquad \frac{dD}{dt} = \gamma_D I, \tag{1}$$

wherein β, γ, and γ_D are the rates of transmission, recovery and deaths, and δ is the rate at which asymptomatic individuals begin to show symptoms. To calibrate these parameters we note that

$$R_I(t_2) - R_I(0) = \gamma \int_{t_1}^{t_2} I dt, \quad D(t_2) - D(0) = \gamma_D \int_{t_1}^{t_2} I dt, \quad \log I(t) \sim (\beta - \gamma)\ t, \tag{2}$$

where R_I is the fraction of population who made recovery after being symptomatic. We observed that these relations are piecewise linear in the available data, and hence once appropriately partitioned, it can be used to calculate the parameters. The value of the hidden parameters, δ and $A(0)$ is found using a simple grid search.

CSIR- Prediction 5

The CSIR model is based on six models (variants of LSTM). For the weekly prediction purpose only three models (Bi-directional, CNN and CONVOLUTION LSTMs, decided based on an analysis of prior performance) are used. The predictions are done on weekly basis. (Saturday-Friday). Each week the models are re-trained using the newly available data (additional 7 days each week). The predictions are purely based on the observed data and hence, are not expected to predict peaks. However, these can be an essential input for supply chain management and logistic planning.

3 Data

We have sourced all the district time line data from the Daily Media Bulletins of Government of Karnataka: https://karnataka.gov.in/common-10/en (till April 27th, 2020) and https://covid19.karnataka.gov.in/govt_bulletin/en (post April 27th, 2020).

We have converted the rest of the information from the bulletins into usable CSV format and made them publicly available for use at the Data Repository at https://www.isibang.ac.in/∼athreya/incovid19/data.html. We have organised the bulletins into the following files:

- Karnataka Trace History : This csv file contains information on each patient (age, sex, city, history) who was confirmed before 21st July.
- Karnataka Hospitalization information : This csv file contains information on the patients regarding their hospitalization. The entries in the cells are either H [implying they were hospitalized on that day], C [Cured], D [Deceased], ICU [required an ICU], ICUO [ICU and required Oxygen] or ICUV [ICU and required Ventilator support]. The ICU information is available only till 8th May.
- Karnataka Hospitalization information - Consolidated : This csv file contains consolidated counts of the total counts of Active, Recovered, Discharged and ICU patients.
- Karnataka Testing information : This csv file contains information on the testing and screening done by the Karnataka government.
- Karnataka District timelines : This csv file contains the infected, recovered and deceased timelines.
- Karnataka Deceased data : This csv file contains about the deceased as provided in the bulletins.

For the CSIR modle the data used is collected from https://www.covid19india.org using an automated API.

4 Acknowledgements:

The prediction time line has been prepared in a collaborative effort by students(undergraduate and post graduate) and faculty at Indian Statistical Institute: COVID-19 India-Timeline an understanding across States and Union Territories , Indian Institute of Science: Modeling of epidemic spread in Indian urban conditions, Indian Scientists' Response to COVID-19: INDSCI-SIM DASHBOARD, SAIR Model- Region-specific COVID-19 predictions for 200 Indian Districts, and https://csir4pi.in.

References

[1] Prakash, M.K., Kaushal, S., Bhattacharya, S., Chandran, A., Kumar, A. and Ansumali, S., 2020. Minimal and adaptive numerical strategy for critical resource planning in a pandemic. *Physical Review E*, 102(2), p.021301.

[2] Kaushal, S., Rajput, A.S., Bhattacharya, S., Vidyasagar, M., Kumar, A., Prakash, M.K. and Ansumali, S., 2020. Estimating Hidden Asymptomatics, Herd Immunity Threshold and Lockdown Effects using a COVID-19 Specific Model. *arXiv preprint* arXiv:2006.00045.