Prediction of end-time of COVID-19 spread in Andhra Pradesh using machine learning

Anvesh Reddy, Hanesh Koganti, Sai Krishna, Suhas Reddy, Soumyajyoti Biswas
SRM University - AP, Andhra Pradesh - 522502, India

In the current pandemic situation, a pertinent question is an estimate of the time in which the virus spread could be contained and normalcy would return. Here we use a well-established Susceptible(S)-Infected(I)-Removed(R) model [1] for simulating the spread of virus and use the simulation data to train a supervised machine learning algorithm (Random forest regression). Then in the trained algorithm, we feed the actual data for the state of Andhra Pradesh and make an estimate of the end-time for the virus. **We estimate that by July 15, the rate of infection in the state will go below 100 per day.** However, the model is a simplified one and does not consider: effects of lockdown, vaccination or third wave due to possible mutant variants of the virus.

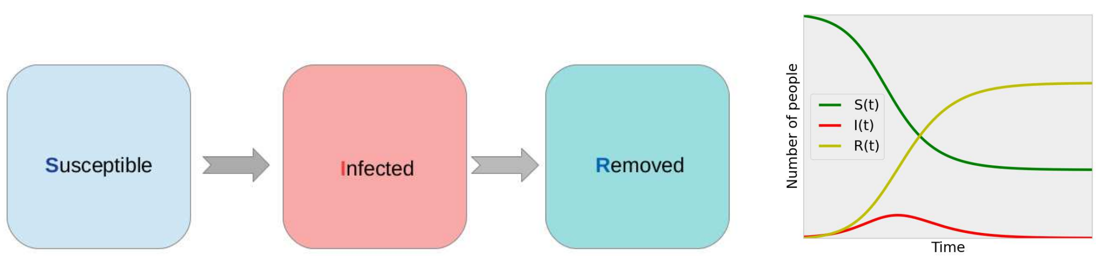

Figure 1: Left: A block diagram of the Susceptible(S)-Infected(I)-Removed(R) model, where the total population is assumed to be constant and divided into these three groups. The right hand side figure shows a schematic representation of how these three quantities vary with time in the model. We focus on the red-line curve showing infected (I) individuals per day.

Fig. 1 shows the schematic diagram of the S-I-R model, where the total population belongs to either of the three categories $N = S(t) + I(t) + R(t) = conts$. A schematic time variation for the three quantities in the model are also shown. It is simulated in a square lattice. We focus on the time variation of the infected population $I(t)$.

The infected population $I(t)$ starts from zero, reaches a maximum and the decreases back to zero (Fig. 2, left). If this total time is T, then the remaining time starts from T and goes to zero (denoted by the straight line in Fig. 2). A supervised machine learning algorithm can predict the remaining time (after training by 100 sets) at each instance. The right hand side of Fig. 2 shows these predictions with the green dots, while the actual times are shown in solid red lines.

Finally, we use the trained ML algorithm to make predictions of the remaining time with data from the state of Andhra Pradesh. The data for the second wave (starting from March, 1) are used. Until May, 3 the remaining time prediction shows the end-time to be July, 15 (Fig. 3), where the infection rate is predicted to go below 100/day. With the additional assumpton that the decay rate of infection is the same as the growth rate, in Andhra Pradesh, the model predicts that the number of infected people could be 10,000 (May 21), 5000 (May 30), 1000 (June 14) 500 (June 14), 500 (June 23) and 100 (July 15).

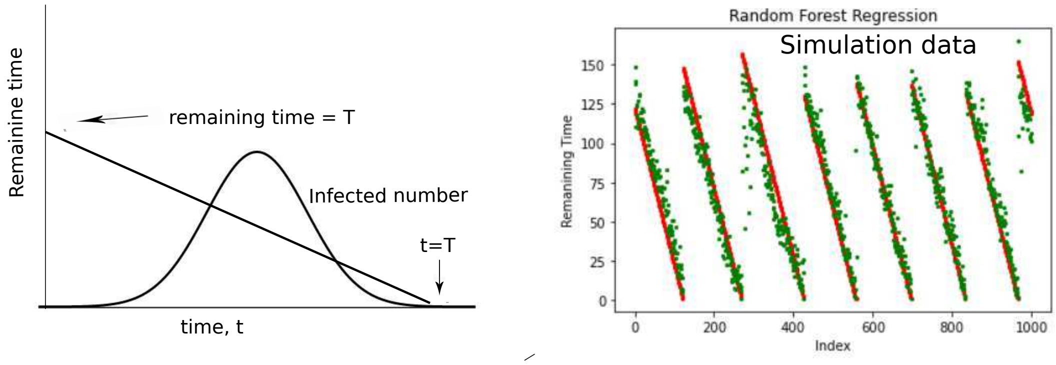

Figure 2: Left: A schematic diagram of the time evolution of the infected number per day ($I(t)$) is shown. The straight line represents the number of days left before the infection rate becomes vanishingly small. If the total duration of the pandemic is T, then the remaining time starts from T when the infection starts and goes to zero when the infection ends. The right hand side figure shows data from simulations. The red lines denote the remaining times found from the simulations of different samples and the green dots denote the predictions provided by machine learning. The goodness of fit is $R2 = 0.90$.

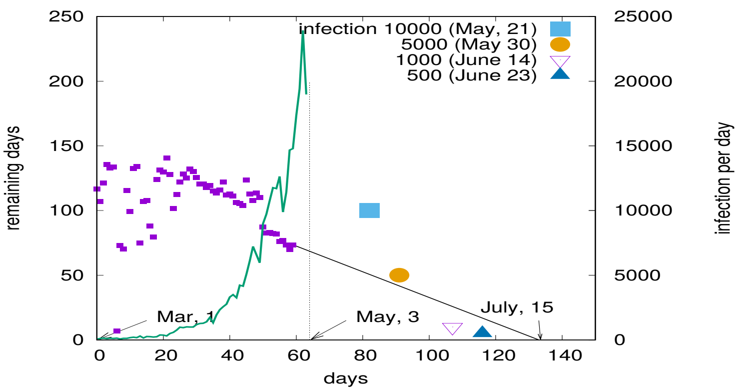

Figure 3: Prediction using actual data: The machine learning algorithm is used to make prediction of the remaining time (denoted by the purple dots) by using the data for infections per day in the state (denoted by green line). Data from March, 1 to May, 3 were used. The final predicted value for the remaining time, as of data till May 3, is July 15 (extrapolated by the straight line). With an additional assumption that the magnitudes of rate of decay in infection is the same as the growth of infection, the decaying values of infections per day at various stages are also indicated.

References

[1] W. Kermack, A. McKendrick, *A contribution to the mathematical theory of epidemics*, Phil. Trans. R. Soc. Lond. A **115**, 700 (1927).

[2] Data source: https://api.covid19india.org/