EPIDEMIC FORECASTING USING MODIFIED SEIRD MODEL

Shubham Jhanwar
Indian Institute of Information Technology,
Design and Manufacturing
Jabalpur, India
2018375@iiitdmj.ac.in

ABSTRACT

Infectious disease outbreaks play an important role in global morbidity and mortality. Real-time epidemic forecasting provides an opportunity to predict geographic disease spread as well as case counts to better inform public health interventions when outbreaks occur. An epidemic is a disease that spread very quickly and covers an entire geographical area in no time that what the world are facing off. It can cause significant economic, social, and political disruption. Before this, mankind has experienced several epidemics like SARS, Dengue, Plague, and Spanish Flue which lead to several death in history. Let alone The Spanish Flu origin in west Asia resulted in death of atleast 17 million people. Owing to the pandemic scenario of COVID-19 disease cases all over the world, the outbreak prediction has become extremely complex for the emerging scientific research. Several epidemiological mathematical models of spread are increasing daily to forecast the predictions appropriately. Currently, we can clearly observe that almost all countries of the world are suffering from this. So by observing the severity of this epidemic we need to inform the right and exact guidelines to every other people..

In this study, we'll consider the Covid-19 disease and analyse its outbreak trend of both the first and the current second wave in India with the classical susceptible-infected-recovered (SIR) modeling approach employed to study the different parameters of this model for India. This approach was analyzed by considering different governmental lockdown measures in India. Some assumptions were considered to fit the model in the Python simulation for each lockdown scenario. The predicted parameters of the SIR model exhibited some improvement in each case of lockdown in India. In addition, the outcome results indicated that extreme interventions should be performed to tackle this type of pandemic situation in the near future

KEYWORDS:

Coronavirus, SEIR model, Logistic Function, Differential Equation, Mean Absolute Error, Mean Square Error

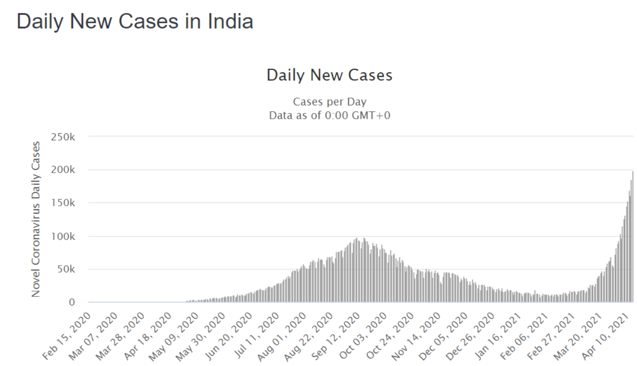

Figure 1: An illustration of the rise of coronavirus cases in India., *India Corona Curve*, Dated April 14.

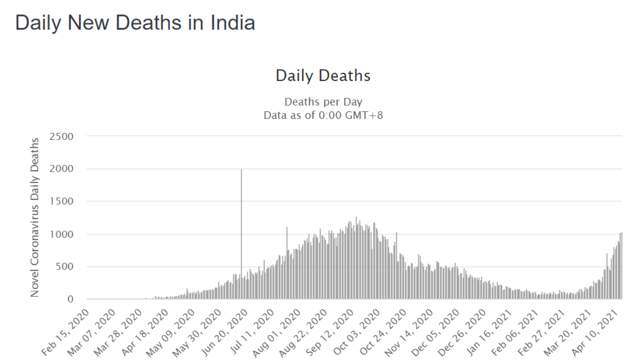

Figure 2: An illustration of reported coronavirus deaths in India., *Corona Reported Deaths, India*, Dated April 14.

1 INTRODUCTION

The COVID-19 pandemic has created a lot of havoc in the world. It is caused by a virus called SARS-CoV-2, which comes from the family of coronaviruses and is believed to be originated from the unhygienic wet seafood market in Wuhan, China but it has now infected around 215 countries of the world. With more than 138.9 million people affected around the world and more than 2,987,189 deaths (As of April 15, 2021), it has forced people to stay in their homes and has caused huge devastation in the world economy.

In India, the first case of COVID-19 was reported on 30th January, which was linked to the Wuhan city of China (as the patient has a travel history to the city). On 4th March, India saw a sudden hike in the number of cases and since then, the numbers are increasing day by day. As of 15th April, India has more than 14 Million cases with more than 174,000 deaths and is the world's 2nd most infected country. It seems like India is facing a second coronavirus wave.

The paper aims to model the disease dynamics by determining epidemic parameters in order to estimate the number of infected individuals over time in India and to predict the pace of COVID-19 disease spreading by using the basic reproduction number Ro. Since control measures in India, such as tracing close contacts, quarantining infected cases, promoting social distancing and self protection using protective equipment, i.e. medical masks and gloves in a public area, started with the first confirmed case on February, 25th, Ro explicitly indicates the effectiveness of the current control measures and facilitate the adoption and implementation of potential additional measures. Thus, the ultimate goal of this paper is to accurately determine R0 and the infected population with the time span of an epidemic as the expected number of secondary infections caused by a single infectious symptomatic or asymptomatic individual in a fully susceptible population.

In order to determine epidemic parameters, stochastic models are well adopted and preferred in the current research, but during the initial stage of the epidemic, the data are sparse and the epidemic dynamics are better described using deterministic modeling. Generalizations of the Susceptible Exposed Infected Recovered (SEIR) model are used for China and heavily affected parts of Europe. However,due to the change of diagnostics and the lack of implementation of control measures in the initial stage of the outbreak, early published models were not close to reality of the situation.

Time series data provided by the Government of India web portal (https://www.mohfw.gov.in/) and John Hopkins University, USA, have been used for the empirical result analysis. The time period of training and test data is from 01/02/2021 to 15/04/2021. The data include confirmed cases, death cases, and recovered cases of all countries. However, this article focuses only on India's data for analysis and prediction of COVID-19 confirmed patients, Reproduction value and stay of the epidemic. The fact that India covers approximately 17.7% of the world's population and to date, the effect of COVID-19 cases per million is less than 1 is the motivation behind this research.

2 OBJECTIVE

The current study use machine learning models to find the rate of spread of the disease in India using SEIR (Susceptible, Exposed, Infectious, Recovered) model. The major objective are: predict the number of cases through Our main objectives are:
a) We try to sketch current scenario of CORONA Infection in India.
b) We model the classic SEIR model with Addition Death Compartment.
c) We further used time dependent Ro(Reproduction Value) to determine peak and span of epidemic using discrete value.
d) We further used time dependent Ro(Reproduction Value) to determine peak and span of epidemic using Logistic Function.

3 LITERATURE SURVEY

Based on the epidemiological characteristics of COVID-19 infection, the SEIR model is more commonly adopted to study the dynamic of this disease. SEIR is a deterministic meta-population transmission model, it simulates each individual in the population as a separate compartment, with the assumption that each individual in the same compartment has the same characteristics.
The entire is divided into 4 compartments.The rate of change over time for each compartment is expressed in terms of a set of 4 coupled ordinary differential equations (ODEs).
Basic SEIR Model(without any interventions and measures):

Time dependent variables are:
N: total population

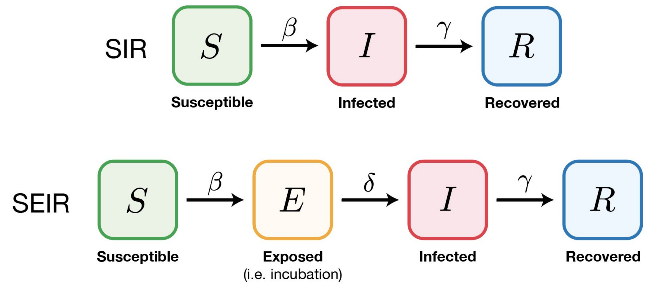

Figure 3: An illustration of SIR/SEIR MODEL

$$\frac{dS}{dt} = -\beta \cdot I \cdot \frac{S}{N}$$
$$\frac{dE}{dt} = \beta \cdot I \cdot \frac{S}{N} - \delta \cdot E$$
$$\frac{dI}{dt} = \delta \cdot E - \gamma \cdot I$$
$$\frac{dR}{dt} = \gamma \cdot I$$

Figure 4: SEIR Differential Equation

S(t): number of people susceptible on day t
E(t): number of people exposed on day t
I(t): number of people infected on day t
R(t): number of people recovered on day t
D(t): number of people dead on day t
β: expected amount of people an infected person infects per day
D: number of days an infected person has and can spread the disease
γ: the proportion of infected recovering per day (γ = 1/D)
Ro: the total number of people an infected person infects (Ro = β / γ)
δ: length of incubation period

4 DATASET DESCRIPTION

Province_State	Country_Region	Last_Update	Lat	Long_	Confirmed	Deaths	Recovered	Active
Andaman and Nicobar Islands	India	2021-04-11 04:20:56	11.225999	92.968178	5175	62	5038.0	75.0
Andhra Pradesh	India	2021-04-11 04:20:56	15.912900	79.740000	921906	7291	895949.0	18666.0
Arunachal Pradesh	India	2021-04-11 04:20:56	27.768456	96.384277	16878	56	16790.0	32.0
Assam	India	2021-04-11 04:20:56	26.357149	92.830441	219958	1117	215907.0	2934.0
Bihar	India	2021-04-11 04:20:56	25.679658	85.604840	279473	1604	265870.0	11999.0
Chandigarh	India	2021-04-11 04:20:56	30.733839	76.768278	30341	396	26680.0	3265.0
Chhattisgarh	India	2021-04-11 04:20:56	21.264705	82.035366	432776	4777	342139.0	85860.0
Dadra and Nagar Haveli and Daman and Diu	India	2021-04-11 04:20:56	20.194742	73.080901	4014	2	3667.0	345.0
Delhi	India	2021-04-11 04:20:56	28.646519	77.108980	714423	11235	674415.0	28773.0

Figure 5: An illustration of Dataset of Corona Cases

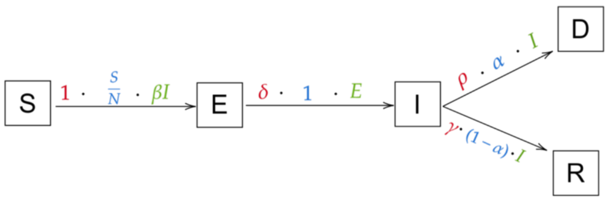

Figure 6: An illustration of SEIRD Model

The dataset that we used here has different attributes like State, confirmed, Death, Recovered and Active status. We have chosen India to visualize the scenario and also perform data analysis and forecasting.

5 DATA VISUALIZATION

The theoretical meaning of Data Visualization is the graphical or pictorial representation of data or information in the form of graphs, charts and maps, etc. In the world of data science, data visualization is used for analysing large amounts of data and information and producing data-driven decisions. Human eyes are subtle to colours and data presented in different colours gives easy grasp to the user and hence can quickly analyse it.

The libraries such as pandas and matplotlib are used in order to plot graphs with a python program.

5.1 Pandas

Pandas is used to create plots from pandas data frames and libraries. It is an open source data structure and data analytics tool used in Python. Pandas can import various formats of files while .csv is the most popular.

5.2 Matplotlib

Matplotlib is a library used in python to create static, animated and interactive visualizations. It comes up with a wide variety of applications using GUI toolkits like Tkinter, , wxPython, etc. This paper uses the pyplot API to plot areas, plot lines and decorate the plot with labels.

6 METHODS

6.1 The Modified SEIR Compartmental Model

For very deadly diseases, this compartment is very important.[Figure:6]
New variables introduced:
α: fatality rate
ρ: rate at which people die (= 1/days from infected until death)

6.2 Time Dependent Ro

As you can see, only the compartments change over time (they are not constant). Of course, this is highly unrealistic! As an example, why should the Ro-value be constant? Surely, nationwide lockdown reduce the number of people an infected person infects, that's what they're all about! Naturally, to get closer to modelling real-world

$$
\begin{aligned}
\frac{dS}{dt} &= -\beta \cdot I \cdot \frac{S}{N} \\
\frac{dE}{dt} &= \beta \cdot I \cdot \frac{S}{N} - \delta \cdot E \\
\frac{dI}{dt} &= \delta \cdot E - (1-\alpha) \cdot \gamma \cdot I - \alpha \cdot \rho \cdot I \\
\frac{dR}{dt} &= (1-\alpha) \cdot \gamma \cdot I \\
\frac{dD}{dt} &= \alpha \cdot \rho \cdot I
\end{aligned}
$$

Figure 7: Differential Equation of SEIRD Model

$$R_0(t) = \frac{R_{0_{start}} - R_{0_{end}}}{1 + e^{-k(-x+x_0)}} + R_{0_{end}}$$

Figure 8: Reproduction Value Logistic Function

developments, we have to make our variables change over time.

6.2.1 Discrete Ro value. First, we implement a simple change: on day L, a strict "lockdown" is enforced, pushing Ro suddenly to other value.
Eg. After 40 days lockdown has been imposed in a region such that the Ro value decreases for 2.3 to 1.

6.2.2 Logistic Ro value Function. In reality, Ro probably never "jumps" from one value to another. Rather, it continuously changes. You can choose any function you want for Ro, We just want to present one common choice to model the initial impact of social distancing: a logistic function.
R_0_start and R_0_end are the values of R_0 on the first and the last day
x_0 is the x-value of the inflection point (i.e. the date of the steepest decline in R_0, this could be thought of as the main "lockdown" date)
k lets us vary how quickly R_0 declines

7 EXPERIMENTATION

We applied SEIR algorithms for the prediction of confirmed cases, death cases, and recovered cases on a random value with N= 1000000 and in another case on Maharashtra datacase Provided by govt.

7.1 SEIRD Model on Maharashtra Data using Discrete Ro value change

After 40 days lockdown has been imposed in a Maharashtra considering the Second wave such that the Ro value decreases for 1.8 to 1.2. Observe that few days can make a huge difference in the overall spread of the disease

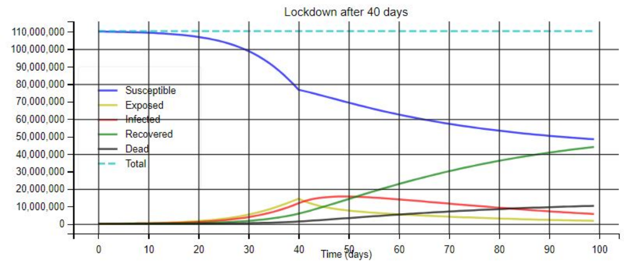

Figure 9: Observation:Due to sudden change is Ro, sudden change in exposed number curve after lockdown

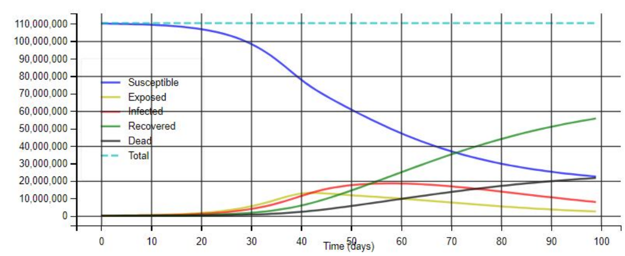

Figure 10: Peak comes soon after the lockdown announced

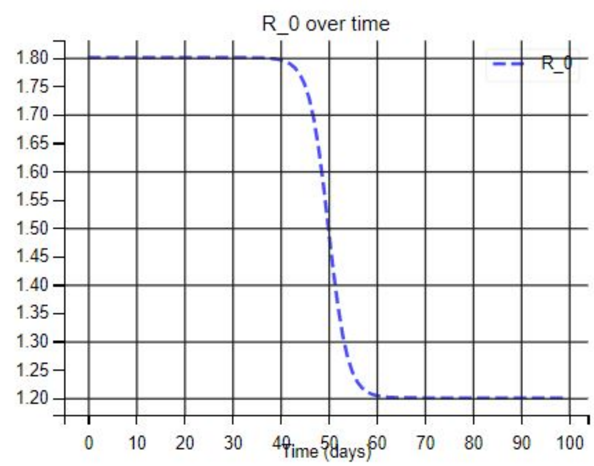

Figure 11: Ro changes smoothly from high to low

7.2 SEIRD Model on Maharashtra Data using Logistic Ro value Function

R_0_start=1.8 and R_0_end=1.1 are the values of R_0 on the first and the last day
x_0 is the x-value of the inflection point=40 DAYS
k=0.5(optimal slope)

8 CONCLUSION

From the study of the modified SEIR model for India, several points could be concluded, which are considered as the betterment of each lockdown measure that is to be incorporated. The recovery rate increased and the transmission rate decreased. Thus, the value of basic reproduction number decreased, flattening the curve of epidemic spread of COVID-19. Similarly, the percentage of susceptible people remaining after the infection had passed decreased, and it should be checked in the near future. In addition, the peak percentage value of infectious people decreased and exhibited better values. In the case of India, further strict governmental interventions should be performed, and, of course, the pandemic cases can be drastically decreased by spreading awareness among the residing people.

9 REFERENCES

[1.] https://www.mohfw.gov.in/
[2.] https://api.covid19india.org/documentation/csv/
[3.] Machine Learning Models for Government to Predict COVID-19 Outbreak RAJAN GUPTA, GAURAV PANDEY and POONAM CHAUDHARY, SAIBAL K. PAL
[4.] Infectious Disease Modelling: Understanding the models that are used to model Coronavirus-Henry Freose
[5.] Modeling the Epidemic Outbreak and Dynamics of COVID-19 in Croatia Ante Lojic Kapetanovi, Dragan Poljak†
[6.] Forecasting the Worldwide Spread of COVID-19 based on Logistic Model and SEIR Model Xiang Zhou1, Na Hong2, Yingying Ma2, Jie He2, Huizhen Jiang3, Chun Liu

ACKNOWLEDGMENTS

This work was done under the guidance of **Dr. Kusum Kumari Bharti**, Assistant Professor in Computer Science and Engineering Department at Indian Institute of Information Technology, Design, and Manufacturing, Jabalpur. I also thank Mr. Apoorv Jain and Mr. Yash Shah for contributing to research and development work.