Epidemic Detection System

For Covid-19 Solution Challenge by Nikhil U.S

Abstract: The software aims to collect data from all hospitals across India about patient disease trends (symptoms, diagnosis, date of admittance and release, result of treatment of each patient anonymised). The software will then show the density of hospitalizations and a growth parameter of patients admitted for different critical symptoms across India (e.g. pneumonia, heart diseases, other conditions of important organs which are symptoms of communicable diseases). This data can be used to predict breakout of epidemics like COVID-19 spread across a larger area using the data from hospitalizations. Making such a highly sensitive system with minimum interference can help pinpoint such diseases with accuracy and help in taking early action.

Methodology: An epidemic occurs from diseases which have a Ro value of greater than 1. Ro is a number which refers to how many people one infected person infects. Common Flu has a Ro value of 1.3(meaning 3 people infect 4 more on average) while Covid-19 has an estimated Ro ranging from 2 to as high as 6 in places. Controlling an epidemic means bringing this value under 1 so that on an average infected people infect less than the number already infected, thereby the number of infected people gradually reduce to 0. Understanding epidemics early on can help in taking such measures.

a) Growth Factor

Covid-19 has a incubation period of 3-14 days. Taking the average incubation period as 7 days and the Ro as 3. This means in a period of 7 days 1 person infects 3 more. We need to model this by the equation

$$x \times m^n + c = x \times Ro + c = y + c$$

Where x is the initial number of patients and y is the final number, c is the noise from patients having other conditions and n is the number of days which have passed and m is the multiplicative factor we need to find by which the patients increase daily with respect to that day's number.

Now we can calculate the growth factor (the number by whose power the number of cases increases *daily*) as

$$m = e^{\left(\frac{\ln(Ro)}{n}\right)} = e^{\frac{\ln(3)}{7}} = 1.17$$

For Covid-19 this comes out as 1.17. For common cold with a Ro of 1.3 and incubation period this value is 1.10. It may not seem like a great difference but 1.10^{90} gives around 5,000 patients in 3 months and 1.17^{90} gives 13,70,000 patients in 3 months. Slight differences in this parameter are very important.

Finding out this value from hospitalizations in hospitals while filtering out the noise is important. Using this we can make a heatmap of the rise or decrease in hospitalizations and deaths across districts and cities. This can help in pinpointing outbreaks and help in taking fast measures. This is done with the help of an algorithm which uses least square fitting after getting rid of the noise.

b) Data averaged over periods of (1 week,1 month, 3 months)

Hospitalization and mortality data over the three different periods is averaged over to get the *growth factor* mentioned in the previous section. The parameter in then plotted over map areas corresponding to the hospital or the district. Three such plots are generated and can be used to predict breakouts of different natures.

- The **1 week data** can be used to detect epidemics with **high mortality rate** irrespective of transmission, such diseases will show up on this immediately by increasing number of hospitalizations.
- The **1 month data** can be used to detect those with **moderate mortality** and **moderate rate of transmission** like Covid-19 this shows up gradually and community spread might be hard to understand as some cases may be asymptomatic.
- The **3 month data** can be used to detect epidemics which might have an even **smaller mortality rate** or **transmittivity**.

These are just preliminary ideas and can be built upon later on with more insight. The averages were taken in this project using least square fitting after filtering the noise.

c) Simulations

To test the working of the program simulations were done to obtain the growth factor

Growth Factor=1.5, Bias =10, Noise=0,1,2,3,4 in the order, growth parameter obtained by fitting is shown to the right.

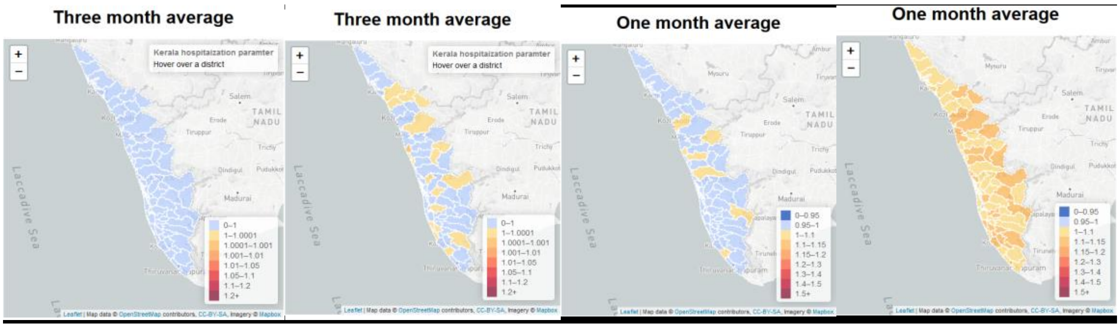
```
11, 12, 12, 13, 15, 18, 21 -> 1.426250
10, 11, 11, 13, 14, 18, 20 -> 1.463570
10, 11, 13, 11, 16, 19, 19 -> 1.324211
14, 14, 10, 13, 15, 21, 19 -> 2.537853
14, 12, 5, 18, 17, 20, 14  -> 0.999999
```

With small noise relative to the increase in patients the algorithm does detect the epidemic quite well.

To generate hospital data. Patients are simulated in hospitals with a symptom list. The severity of the symptoms were used to probabilistically determine if the patient got cured or not. An epidemic was simulated on top of the normal patient inflow to test for epidemic detection.

Data on that is available at the end. The minimum growth factor (gf) of epidemics which were detected by each parameter are shown below. Detection here is defined as when all cases show a growth parameter greater than or equal to 1. This is expected to show higher precision when this is done on just cases with critical symptoms like pneumonia.

Factor	7 day epidemic	30 day epidemic	90 day epidemic
7 day gf	3	1.15	1.06
30 day gf	-	1.1	1.05
90 day gf	-	-	1.07



Images from a simulated epidemic(1.1 gf) are shown above displayed on Kerala state.

Link to the code(github) and the simulations:

- https://github.com/nikhil1998us/Epidemic_Detection_System
- https://github.com/nikhil1998us/Epidemic_Detection_System/tree/master/Simulations

Vision: The software takes hospital data about patients from a Hospital Management Software and then takes the essential data which is sent to a central Facility where this data is compiled to make a *heat map of hospitalizations* for different diseases. It also collects data about hospital capacity and using patient data can be used to predict if there may be an overflow of cases in the area. Average values of hospital occupancy can be estimated and using an exponential fitting algorithm an exponential parameter is calculated which helps in determining epidemic outbreaks.

Steps required for deployment:

1. Integration of this program with a Hospital Database Management Software has to be done. This can be done with some help from one such company by providing the saved data in a folder in a specified format. This program can also be modified to take the information if the format in which data stored is revealed and permissions are given.
2. The data transfer network has to be setup. This will require less than a Kilobyte of data sent from each hospital to the central facility which can be state or nation based.
3. Symptom bases simulation will be added soon. This will reduce the noise and the bias greatly thereby giving more accurate growth paramters.

Acknowledgements:

- I would like to thank the creators of the open source plugin Leaflet whose plugin made my work in displaying the map a lot easier
- I would also like to thank OpenStreetMap and MapBox whose public Map data were used in this project.
- I would also like to thank Sajjad Anwar (https://github.com/geohacker, https://sajjad.in/) from the company Development Seed whose work in providing geojson data of India maps were of great help in my project.

Declarations:

All work used in this project is open source and had no private licence.