Infection Transmission Rate Forecast Simulation of COVID-19

Website: https://cmpsc465.000webhostapp.com/

Shivran Muralidhar, Vivek Gopalan, Harsh Maniar, Devkumar Ojha, Ameya Hampihallikar, Piyush Nagasubramaniam

Description

COVID-19 is a very contagious respiratory illness that spreads at an ultra-fast rate. Although people of all ages can be infected by COVID-19, patients of all ages and ethnicities have been affected by the disease giving COVID 19 a fairly high mortality rate. To slow the spread of the virus and prevent citizens from passing away, governments are restricting travel, quarantining residents, and locking down entire cities all across the globe.

This project builds a sophisticated computational model that forecasts the spread and transmission of COVID-19 on a given city or state based on data established by the CDC, WHO, and other organizations. This project will allow different bodies of government to better understand the spread and flow of COVID-19 which enables them to make wise decisions when estimating the time periods on lockdowns.

Simulation Method

This can be used as a decision support tool in order for the general public to use it as a guideline. In order to construct our probabilistic model, we utilized datasets in the raw input in order to find the mean and standard deviations of the number of cases in respective states and counties. We experimented with the raw data set by hosting it on a web server that we use for our website to ease the computational strain on our computers. PHP and MYSQL admin was used to manage and connect with the data from our python code which will be constructed on PyCharm Environment. Using the raw data, we can develop a Poisson model to predict the spread and demise of COVID at certain intervals of time.

We define lambda as our rate of infection, and the Poisson model simulates how the disease can spread across populations such as counties and states. This follows the classical data analytics and the probabilistic computation model widely used in academic research. With our model, our hope is to identify COVID 19 hotspots/epicenters and create some public policy recommendations in order to combat the spread of this vicious disease. Basically, our simulation will be based on trusted sources that will be executed using Pandas, NumPy, and Statistics library sets on python. We will use this quantified characteristic to apply a predictive analytic tool provided using APIs provided by Google and other API providers. These predictive analytic API tools will use data-driven unsupervised machine learning algorithms to predict clusters of similarities in data sets and forecast our required needs. With our simulated data predictions, we can implement clustering algorithms to predict hotspots and develop public policy precautions. The Poisson process equation used in our simulation is $P(x = \mathrm{N(s+t) - N(t)}) = (e^{-\lambda t}(\lambda t)^x)/x!$. With this equation, we can predict how COVID cases evolve in a region over periods of time. In a Poisson process, our rate/mean would be equal to u=λt, and x is the number of people infected within delta t (i.e 7 people develop COVID between two distinct periods of time). The discrete Poisson process gives us the probability that x number of people get infected within a certain change in time (delta t) which then enables us to predict hotspots.

Update

At first, using the Java programming language, we read in all the zip codes in the borough of Brooklyn. After the zipcodes are initialized, we develop an array and use keys to get the number of COVID cases in each zip code. A discrete-time Poisson process is implemented to get the probability that x will be infected within time delta t. With t being fixed, we can get a rate/lambda to get a good mathematical estimation of how COVID is spreading within all the zip codes in Brooklyn. X cannot be a continuous random variable as 7.5 people cannot be infected with COVID. People are represented by whole numbers hence the discrete-time process to estimate the probability of COVID spreading within-population/zip codes in Brooklyn. For example, we take a look at a hypothetical zip code in Brooklyn, and we wanted to find the probability that 5 people get COVID in a four day period. If the rate or lambda were 2, the probability would be $(e^{-2*4}(2*4)^5/5!)$ = 0.09160366159. This essentially implies that there's a 9.16% chance that 5 people can develop COVID in that specific area of Brooklyn within a span of 4 days. With public policy implementations , we can use these probabilities to forecast outcomes in each Brooklyn zip code.

We first read in a Brooklyn zip code map. Then we create a zeroed matrix that has the same dimensions of the image map that we read. Next, We pull individual data points from the .xlsx file of calculated COVID cases from an excel sheet. The data point is segmented by x-coordinate, y-coordinate, and COVID cases of the corresponding zip code to our image dimensions. The COVID cases value is updated to the zeroed matrix with the respective x-coordinate and y-coordinate locations of the matrix. The image is run through a Gaussian filter to distort the values of the matrix. Then the values are converted to a density matrix. The density matrix is passed onto a heatmap_overaly that overlays our matrix onto the Brooklyn zip code map then converts the density matrix to a heatmap using the heatmap toolbox. Finally, at the end of the iterations, a colorbar is added to the end of the figure to show the relative differences between the zipcodes.

Reflections

After the combined herculean efforts of our group to utilize programming languages such as Matlab, HTML, Python, and Java and statistical techniques such as stochastic modeling, our group believes that 5 percent extra credit is deserved, given the over hundred hours dissipated to learn and translate the newly acquired skills to help mitigate the global COVID epidemic.

Reference Data Sets

County Data

https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/

State Data

https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html

New Cases, Active Cases, Total Deaths, New Deaths and Recovery Data for each state

https://www.worldometers.info/coronavirus/country/us/

Birth and Death Data

https://en.wikipedia.org/wiki/Birth%E2%80%93death_process