ISSN 2350-109X
www.indianscholar.co.in

Indian Scholar

An International Multidisciplinary Research e-Journal

DATA SCIENCE - REMODELLING PANDEMIC ADMINISTRATION

Yaksh Dinesh Dharod
Utpal Shanghvi Global School
yakshdharod@gmail.com

Abstract

Pandemic diseases are highly contagious. Great efforts will be needed to successfully combat this disease. In this paper, we evaluate the role of data science. Data science, combined with statistical analysis, computer science and computational biology, is helping in numerous ways with applications including epidemiology, drug discovery, and molecular design for diagnostic and therapeutic purposes. A number of data driven models, mathematical models, correlations and predictive models have been established. Challenges faced by the data scientists at present have been highlighted. Finally, open source datasets sources are mentioned that can be potentially used in curing and evaluating the health policies.

Keywords:- *Data Science Technology,Medical Science,Data Analytics, Artificial Intelligence*

INTRODUCTION

Infectious diseases represent a main challenge for health systems worldwide, either in private or public sectors. With the increase in cases related to these problems, combined with the recent global pandemic of COVID-19, the need to study schemes to cure these health disturbs is even more latent. Data Science has been addressed in this situation with the chance of predicting, mapping, tracking, monitoring, and raising alertness about these epidemics and pandemics. Thus, the main motive of this study is to find how it can help in cases of pandemics and epidemics. Data science technologies today are supporting medical science in reaching new milestones in medical imaging, genetics and genomics, drug discovery, patient– customer assistance, and predictive medicine. COVID-19 has put this in a spotlight. Data analytics has been effectively used to monitor real time disease outburst, forecasting, and spotting real time trends for governments, health organizations, and society in general. Data science field can be categorized into data management, data visualization and statistical machine learning. Each field has methods that can be used for organization, sorting, processing and enabling real time data analysis. At present, chemical engineers can aid in managing COVID-19 response with suitable data analytical techniques. These techniques are generally used to exploit the correlations in the datasets due to the mass transfer, energy transfer and basic thermodynamics and can be implemented for process modelling, diagnostics and predictions. In comparison to the earlier outbreaks, open source datasets for countries and cities have been made widely available for

COVID-19. Combining these with the socio-economic factors, investigators have been engaged in mathematical modelling and use of artificial intelligence. Applications in the field of risk assessment, diagnostics, modelling, contact tracing, economic and logistic planning, understanding effect of government policies and social interventions have helped us in gaining insights into the pandemic. The current research to detect the anomalies of pandemic, besides the diagnostics and therapeutics, is around detection of virus using CT scans data mining to collate information from social media and patient records, diagnostics of lungs and respiratory sound analysis in addition to speech and sound processing. The use of modelling techniques has become particularly beneficial during the COVID-19 pandemic for forecasting of trends and apply it for anticipating resource requirements, informed policy making, and ensuring adequate non-pharmaceutical interventions. This article aims to review the various ways in which modelling and artificial intelligence are being used to handle the pandemic situation efficiently and effectively.

Theory

1. Numerical analysis of pandemic trends there has always been a nearby association between biostatistics and epidemiology for health planning and policy making. Statistics can be grouped into descriptive statistics and inferential statistics. Utilization of statistical analysis for describing the diagnostic test accuracy has been a benefit in the present COVID-19 scenario. Datasets are accessible publicly to enable population-based study taking individuals age and chronic medical conditions to scrutinize the consequence on death rate and approximation of recovery time. The most popular use of statistical correlation is evident in analyzing the consequence of climate conditions as well as the socio-economic factors on spread of disease. However, there are a few statistical explorations that claim that the spread of pandemic depends on population and its density, and that it is impervious to the weather conditions. Clustering analysis has been used extensively to visualize the likenesses in the countries based on the active cases per population and per area network analysis to estimate the incubation period of virus, infection rates, and fatality rates and spatial statistical analysis for analyzing the spread of the disease along with the grouping of the similar behaving countries. To normalize the irregularities in the datasets, fractal interpolation and dimensions has been confirmed to be well-organized and effective. The isolation strategy, a probable option for lockdown, has been established to assist health planning based on the predictive analysis of the number of tests.
2. Mathematical modelling: Due to the nuances involved in modelling of contagious diseases, mathematical models including the Susceptible Exposed-Infectious-Recovered model have been developed specifically for epidemiology. Some variations of the model consist of simulations of the spread of the disease based on constraints. Researchers have aimed to alleviate limitations of the model by accounting for health care capacity, underreporting of data, ICU beds accessible and precision of modelling the number of fatalities.
3. Use of artificial intelligence in modelling of pandemic: The flexibility of artificial intelligence has surged up the momentum to implement the modus operandi for medical and societal adversity. AI cover a varied biomedical area. However, only few

implementations are successful. The application of multilayer perceptron neural network in predicting the incident rates, recurrent neural network in combination with long short-term memory models for forecasting the infected cases and convolutional neural network for detecting the virus by analyzing the patients X-ray have offered useful insights about the spread of disease. In addition, algorithms like support vector regression for predicting the spread and analyzing the growth/transmission rate, random forest machine learning model for anticipating compound growth rate with respect to social distancing stringency and as a discrimination tool for early screening have contributed towards gaining an improved understanding of the potential risk factors. For early stage diagnostics, scope of implementation of methods like random forest classification, lasso based generalized linear models and supervised machine learning algorithms are being explored. However, for predicting the trends, genetic programming and models based on ensemble empirical mode decomposition with artificial neural network have proven to be effective.

4. Managing pandemic using data science and technology: Data technology has helped in assessing the associated risk factors, curing and surveillance. Furthermore, data analysis assists in realizing the effectiveness of non-pharmaceutical interventions & development of vaccines. Technology start-ups in partnership with the clinicians, academicians and government officials, are helping in managing the pandemic. The Bluecoat artificial intelligence, a Canadian start-up, was among the first one to identify, track and forecast outbreak in Hubei province and predicted the first 8 cities that would import the virus. Their real time data intelligence helps with monitoring of the pandemic movements. AI powered diagnostic systems developed by Alibaba has 96% efficiency for diagnosing coronavirus in few seconds based on the CT scans of patients. Google Deep mind AI is helping to identify the structure of protein associated with the respiratory diseases like coronavirus whereas Benevolent helps in determination of potential drug candidates. Terra drones are being used to deliver the medical supplies to minimize the human to human contact whereas robots are being used for cleaning and sterilization. In the present pandemic situation, data technology and artificial intelligence are playing a critical role in effective management of the outbreak.

Experimental

Before policymakers reopen their economies, they must be certain that the resulting new COVID-19 cases will not force indigenous healthcare systems to resort to crisis standards of care. Doing so needs not just prevention and suppression of the virus, but ongoing measurement of virus activity, assessment of the efficacy of suppression measures, and forecasting of near-term demand on local health systems. This demand is extremely variable given community demographics, the prevalence of pre-existing conditions, and population density.

Data science can already offer ongoing, precise estimates of health system demand, which is a constraint in almost all reopening plans. We need to go beyond that to a dynamic approach of data collection, analysis, and forecasting to inform policy decisions in real time and iteratively optimize public health recommendations for re-opening. While most reopening plans propose extensive testing, contact tracing, and monitoring of population mobility, almost none consider setting up such a dynamic feedback loop. Having such response could determine what level of virus activity can be tolerated in a zone, given regional health system capacity.

By using current technology, it is likely to set up that feedback loop, which would maintain healthcare demand under the threshold of what is available in an area. Just as the maker community stepped up to cover for the failures of the government to provide adequate protective gear to health workers, this is a chance for the data and tech community to partner with healthcare professionals and provide a measure of public health planning.

For the data science effort to work successfully, we need to fix delays in data collection and access introduced by existing reporting processes. At this time, many sectors of public health are collecting and reporting metrics that are not helpful, and are reporting them with 48 hour delays, and often with mistakes. Although there are examples of regional excellence in such reporting, by and large, the recommendations from the health IT community around precise and fast public health reporting remain overlooked. For instance, consider the amount of COVID-19 hospitalizations, which is the greatest indicator of the disease's burden on the regional health system. Currently, as a result of time lags in confirming and reporting cases and a failure to distinguish between current and cumulative hospitalizations, even regions that report hospitalization data often provide only an unclear picture of the burden on the regional health system. Regions should ideally account both suspected and confirmed hospital cases and specify the date of admission, in addition to the date of report or authorization.

Even with perfect reporting, there are fundamental delays in what such data can tell us. For example, new admissions to a hospital today reflect virus activity as of 9 to 13 days ago. Not factoring in such considerations have led to significant over-estimation of hospitalization needs nationwide. We therefore need to measure virus activity via proxy measures that are indicative early in the lifecycle of the virus. We must benchmark these against the number of new and total COVID-19 hospitalizations as well as ideally the number of new infections, assuming it is accurately measured through large scale testing. Available proxy measures include test positivity rates in health systems, case counts, deaths and perhaps zero positivity rates. Ongoing symptom tracking via smartphone apps, daily web or phone surveys, or cough sounds can identify potential hotspots where virus transmission rates are high. Contact tracing, which currently requires significant human effort, can also help tracking of potential cases if it can be scaled using technology under development by major American tech companies.

With reliable tracking and benchmarking in place, we can calculate infection prevalence as well as daily growth and transmission rates, which is essential for determining if policies are working. This is a problem not only of data collection but also data analysis. Issues of sensitivity, daily variability, time lags, and confounding need to be studied before such data can be used reliably. For instance, symptom tracking is nonspecific and may have difficulty tracking virus activity at low prevalence. Other emerging data sources such as wastewater and smart thermometer data hold similar promise but will have to grapple with these same issues.

Several efforts have quantified the impact of mobility on virus transmission and a few have suggested "safe" sorts of mobility. While there are many potential ways to quantify population mobility — like via traffic patterns, internet bandwidth usage by address, and site of MasterCard swipes — the foremost scalable mechanism to live mobility appears to be via tracking of smartphones. Groups like the COVID-19 Mobility Data Network offer such data daily in anonymized, aggregated reports.

Once the power to project from mobility to transmission to health system burden is made, we will "close the loop" by predicting what proportion mobility we will afford given measured virus activity and anticipated health system resources within the next fortnight. Researchers have already attempted to calculate "tolerable transmission" within the type of maximum contagion prevalence during a given geography which may not overload health systems. Coupling such tolerable spread estimates with daily assessments of a legitimate sample of the population would allow nursing of changes in transmission which may alert us to the necessity to intervene, like by reducing mobility. As new measures like contact tracing cut transmission rates, these same monitoring systems can tell us that it's harmless to extend mobility further. Uninterruptedly analyzing current mobility also as virus action and projected health system capacity can allow us to line up "keep the distance" alerts that trade off tolerable broadcast against allowed mobility. Concretely, then, the crucial "data science" task is to search the counterfactual function linking last week's population mobility and today's transmission rates to project hospital demand fortnight later. It's blurry what proportion days of knowledge of every proxy measurement what scientific form this function might take, and the way we do that correctly with the observational data accessible and avoid the trap of mere function-fitting. However, this is often the info science problem that must be undertaken as a priority so that the situation comes back to normal.

RESULT

Implementing such technology and data science to stay anticipated healthcare needs under the edge of availability during a region requires numerous privacy trade-offs, which can require thoughtful legislation in order that the solutions invented for enduring the current pandemic do not lead to loss of privacy in perpetuity. However, given the immense economic also as hidden medical toll of the shutdown, we urgently got to construct an early warning system that tells us to reinforce suppression measures if the next COVID-19 outbreak peak might overwhelm our regional healthcare system. It is imperative that we focus our attention on using data science to anticipate, and manage, regional health system resource needs supported local measurements of virus activity and effects of population distancing.

CONCLUSION

Contagious and infectious diseases represent a serious test for health systems, both public and personal, round the world. As it may be a recent epidemic, little has been originate within the literature on the utilization of knowledge within the control and monitoring of pandemic. Thus, methods and advancements used in other epidemics and pandemics have been mapped to recount them to the COVID-19 pandemic, seeking to highlight possible strategies that may help in this context.

Therefore, in order to classify how Data Science can help in the fight against epidemics and pandemics, we mapped the types of data and sources used for analysis and creation of techniques that support the fight against epidemics and pandemics. Equally, the techniques wont to treat these data were also mapped, thus showing their correlation with the present COVID-19 pandemic.

The changes Data Science can transport back the healthcare sector promise to be much bigger than many, governments, companies and organizations, can realize. With the emergence of a series of smart devices capable of collecting, storing, analyzing, and sharing user data during a mist, will make many data available to many people. This situation will change, almost completely, the way the health science outcome is attained.

The limitations of this study is that the incontrovertible fact that the review used only articles from journals, excluding studies published in other sources and websites to some extent.
We have realized that the Health Area is a fruitful field for Data Science, in which not both can broadly benefit from future studies, but the whole society as well. From these studies, results, dashboards for monitoring information, and even insights on data from patients who contracted the disease, future outbreaks could be early identified and stop illnesses to spread. Lastly with the data science we can eradicate the pandemic quickly.

REFERENCES

Nikita Saxena, Priyanka Gupta, Ruchir Raman & Anurag S. Rathore. Role of data science in managing COVID-19 pandemic. Full article: Role of data science in managing COVID-19 pandemic. https://www.tandfonline.com/doi/full/10.1080/00194506.2020.1855085 (accessed April. 21, 2021)

Eric Luellen. How is data science helping with the COVID-19 pandemic? How is data science helping with the COVID-19 pandemic? / By Eric Luellen. https://towardsdatascience.com/how-is-data-science-helping-with-the-covid-19-pandemic-ecb524df6a19 (accessed April. 21, 2021)

Abe Dearmer. How Has COVID-19 Impacted Data Science? COVID-19 and Its Impact on Data Science / Xplenty. https://www.xplenty.com/blog/covid-19-and-data-science/ (accessed April. 21, 2021)
Alana Corsi, Fabiane Florencio de Souza, Regina Negri Pagani & João Luiz Kovaleski Journal of Ambient Intelligence and Humanized Computing (2020). Big data analytics as a tool for fighting pandemics: a systematic review of literature. https://link.springer.com/article/10.1007/s12652-020-02617-4 (accessed April. 21, 2021)

Castro R., Luz P.M., Wakimoto M.D., Veloso V.G., Grinsztejn B. and Perazzo H., "COVID-19: A Meta-Analysis of Diagnostic Test Accuracy of Commercial Assays Registered in Brazil", Brazilian Journal of Infectious Diseases, 24(2), pp. 180–187 (2021).

Banerjee A., Pasea L., Harris S., Gonzalez-Izquierdo A., Torralbo A., Shallcross L., Noursadeghi M., Pillay D., Sebire N., Holmes C., Pagel C., Wong W.K., Langenberg C., Williams B., Denaxas S. and Hemingway H., "Estimating Excess 1-Year Mortality Associated with the COVID-19 Pandemic According to Underlying Conditions and Age: A Population-Based Cohort Study", The Lancet, 395(10238), pp. 1715–1725 (2021).

Barman M.P., Rahman T., Bora K. and Borgohain C., "COVID-19 Pandemic and its Recovery Time of Patients in India: A Pilot Study", Diabetes and Metabolic Syndrome: Clinical Research and Reviews, 14(5), pp. 1205–1211 (2021).

Menebo M.M., "Temperature and Precipitation Associate with COVID-19 New Daily Cases: A Correlation Study Between Weather and COVID-19 Pandemic in Oslo, Norway", Sci. Total Environ., 737, pp. 139659 (2021).

Livadiotis G., "Statistical Analysis of the Impact of Environmental Temperature on the Exponential Growth Rate of Cases Infected by COVID-19", PLoS One, 15(5), pp. e0233875 (2021).

Kumar A., Rani P., Kumar R., Sharma V. and Purohit S.R., "Data-driven Modelling and Prediction of COVID-19 Infection in India and Correlation Analysis of the Virus Transmission with Socio-Economic Factors", Diabetes Metab Syndr Clin Res Rev, 14(5), pp. 1231–1240 (2021)

Chatterjee A., Gerdes M.W. and Martinez S.G., "Statistical Explorations and Univariate Timeseries Analysis on COVID-19 Datasets to Understand the Trend of Disease Spreading and Death", Sensors (Switzerland), 20(11), pp. 3089 (2021)

Meraj G., Farooq M., Singh S.K., Romshoo S.A., Sudhanshu N.M.S. and Kanga S., "Coronavirus Pandemic Versus Temperature in the Context of Indian Subcontinent: A Preliminary Statistical Analysis", Environ Dev Sustain (WHO), pp. 1–11 (2021)

Zarikas V., Poulopoulos S.G., Gareiou Z. and Zervas E., "Clustering Analysis of Countries Using the COVID-19 Cases Dataset", Data Br, 31, pp. 105787 (2021).

Wang P., Lu J., Jin Y., Zhu M., Wang L. and Chen S., "Statistical and Network Analysis of 1212 COVID-19 Patients in Henan, China", Int. J. Infect. Dis., 95, pp. 391–398 (2021).

Neil M., Fenton N., Osman M. and McLachlan S., "Bayesian Network Analysis of COVID-19 Data Reveals Higher Infection Prevalence Rates and Lower Fatality Rates Than Widely Reported", J. Risk. Res., 0(0), pp. 1–14 (2021).

Huang R., Liu M. and Ding Y., "Spatial-Temporal Distribution of COVID-19 in China and its Prediction: A Data-Driven Modeling Analysis", J. Infect. Dev. Ctries., 14(3), pp. 246–253 (2021).

Franch-Pardo I., Napoletano B.M., Rosete-Verges F. and Billa L., "Spatial Analysis and GIS in the Study of COVID-19. A Review", Sci. Total Environ., 739, pp. 140033 (2021).

Melin P., Monica J.C., Sanchez D. and Castillo O., "Analysis of Spatial Spread Relationships of Coronavirus (COVID-19) Pandemic in the World Using Self Organizing Maps", Chaos Solitons Fractals, 138, pp. 109917 (2021).

Păcurar C.M. and Necula B.R., "An Analysis of COVID-19 Spread Based on Fractal Interpolation and Fractal Dimension", Chaos Solitons Fractals, 139, pp. 110073 (2021).

Djilali S. and Ghanbari B., "Coronavirus Pandemic: A Predictive Analysis of the Peak Outbreak Epidemic in South Africa, Turkey, and Brazil", Chaos Solitons Fractals, 138, pp. 109971 (2021)

Maheshwari H., Yadav D., Chandra U. and Rai D.S., "Forecasting Epidemic Spread of COVID-19 in India Using Arima Model and Effectiveness of Lockdown", Adv Math Sci J, 9(6), pp. 3419–3432 (2021).

Ren H., Zhao L., Zhang A., Song L., Liao Y. and Lu W., "Early Forecasting of the Potential Risk Zones of COVID-19 in China's Megacities", Sci. Total Environ., 729, pp. 138995 (2021)

Mohamadou Y., Halidou A. and Kapen P.T., "A Review of Mathematical Modeling, Artificial Intelligence and Datasets Used in the Study, Prediction and Management of COVID-19", Applied Intelligence, 50, pp. 3913–3925 (2021)

Annas S., Isbar Pratama M.I., Rifandi M., Sanusi W. and Side S., "Stability Analysis and Numerical Simulation of SEIR Model for Pandemic COVID-19 Spread in Indonesia", Chaos, Solitons Fractals, 139, pp. 110072 (2021).

Karako K., Song P., Chen Y. and Tang W., "Analysis of COVID-19 Infection Spread in Japan Based on Stochastic Transition Model", Biosci. Trends., 14(2), pp. 134–138 (2021).

Fox G.J., Trauer J.M. and Mcbryde E., "Modelling the Impact of COVID-19 on Intensive Care Services in New South Wales", Med. J. Aust., 212(10), pp. 468–469 (2021).

Hu Z., Qianqian C., Junmei H., Xia W., Sha W.E.I. and Zhidong T., "Evaluation and Prediction of the COVID-19 Variations at Different Input Population and Quarantine Strategies, a Case Study in Guangdong Province, China", Int. J. Infect. Dis., 95, pp. 231–240 (2021).