LITERATURE REVIEW

Predicting SARS-CoV-2 Transmission Risk for Specified Establishments

Daiwik Pal – December 22nd, 2020

Massachusetts Academy of Math and Science

Table of Contents

Introduction

It is vital to predict the effects of a pandemic, so that people and countries are better suited to handle the situation. Pandemics create detriments that ruin the lives of people personally, socially, and economically. To prevent these problems from exacerbating, epidemiologists create scientific models that aim to predict factors like economic loss and exposure risk. They factor in various indictors ranging from the total population of a community to seemingly unrelated indicators such as the economic fluctuation. Some indicators that are hard to measure can be estimated through computer learning systems. For example, a deep learning object detection system can be used to analyze parking lots and provide the model with an estimation of the population density at specific buildings or events. With a combination of these two methods, a more accurate model can be created that can predict transmission risk at specific establishments or venues.

Modeling Viruses

Introduction to Scientific Modeling

Modeling real-world situations help researchers solve complex problems by analyzing various factors that relate to these problems. Scientific models help explain different aspects of the real world. They are used to make predictions of certain phenomena and are usually tested against data from the real world. These models use complex mathematics to be built and investigated, so computers are a useful tool. They are extremely useful once they have been proven to be reliable because now scientists have a cost-effective and easy way to predict results without having to conduct extensive research and experiments ("Science Uses Models to Explain Aspects of the Real World.," 2012). On top of that, they can be used in fields where conducting experiments may be impossible, such as space or deep ocean research, so, in these cases, the only option is to use a model. In general, models are built for scientific investigation and help researchers gain a better understanding of the phenomena in question.

In this paper, a model for predicting potential economic losses due to viral disease spread and a model that predicts transmission risk for COVID-19 will be discussed.

Economic Model

Scientific Modeling helps researchers make predictions and decisions more effectively. For example, when nations are faced with challenges that impede their economic processes and strain their infrastructure, they have to expend money in an unplanned fashion. When the COVID-19 pandemic hit the United States, the country lost 701,000 jobs in the first month (Feritas, 2020). Also, by the end of November 2020, hospitals in hot spots, such as the southwestern US, were projected to hit 90% capacity (Rio & Bogel-Burroughs, 2020). Moreover, as COVID-19 cases began to rise in October 2020 the stock market took a turn for the worse; specifically, the Dow Jones Industrial Average stock market index decreased by 650 points (Oh, 2020). With the implementation of widespread quarantine, many establishments like schools, offices, and public areas shut down, which led to indirect economic loss. For similar reasons, this model aimed to predict the economic loss that countries and people can incur during a pandemic when parts of the country are shut down.

<u>Functionality</u>

When trying to evaluate which type of closure (school or public transportation closures) would lead to the least amount of economic loss, researcher Jérôme Adda (2016) decided to model the economic effects of three diseases: influenza, chickenpox, and gastroenteritis. Because the country had been recording data regarding the three diseases since 1983, he chose France as the nation to make and test the model on. Moreover, having a large data set to analyze made the model more accurate (Adda, 2016). Even though France never had a formal lockdown for these viruses, it did have events where people refrained from going to public establishments, such as holiday seasons and train transportation strikes. These events emulate quarantining and a complete shutdown of transportation systems to prevent disease spread, allowing for a more accurate model as well.

To create the model, Adda (2016) first analyzed his data and tried to find correlations between the economy, school closures (vacations), and incidence rates for the three viruses. What he found was that cases for influenza and gastroenteritis were lower in communities when children are out of school. This was interesting since the incidence rates were not only lower for children as they went out of school but also for adults and elders. Second, he found that when the economy was doing well, there were more cases for all three diseases. His third finding was that, when the new rail lines connected cities to places other than Paris opened, there was a jump in cases for these diseases in all of those cities. However, in later years when these rail lines went on strike, there was a decrease in cases (Adda, 2016). The findings gave him enough proof that public transportation, school closures, and economic activity were all correlated to incidence rates for the three diseases. Equipped with this data, he then created an Ordinary Least Squares Regression mathematical model.

After numerous iterations and testing, he arrived at this model:

$$\begin{aligned} I_{rt} = {} & I_{rt-\tau} S_{rt-\tau} \sum_{k=1}^{K} \alpha^{k}_{within} W^{k}_{rt-\tau} \\ & + \sum_{c \in R \setminus r} I_{ct-\tau} S_{rt-\tau} \sum_{k=1}^{\tilde{K}} \alpha^{k}_{between} \tilde{W}^{k}_{rct-\tau} + X_{rt}\delta + \eta_{rt}. \end{aligned}$$

This model predicts incidence rates (I_{rt}) based on variables such as public building closures, weekly average temperatures, the disease's incubation period, population density, and measure of economic activity. After creating the model, he investigates the effectiveness of closing down schools versus closing down public transportation and uses the model to predict the decrease in the prevalence of flu-like illnesses and acute diarrhea. When closing schools for one week every month, his model predicted, on average, that incidence would lower about 92 percent during the winter months, and close to 100 during the summer. Next, when closing public transportation, the model predicted that incidence would lower by about 95 percent in the winter months and close to 100 percent in the summer months for both diseases. After that, he performs a cost-benefit analysis for each closure for each age group - children, adults, and elders - because the economic effects for each age group are different. For children, he factors in the cost of medication for each disease and how less schooling (closure #1) can affect their salaries in the future; he concluded that the country would lose about 350 euros per capita on children. For adults, the main reason they would lose money is by staying home (fewer work hours) when they are sick, and he calculated they would lose 150 euros per capita. Finally, for elders, the main costs would be due to medications and large death tolls, however, the country would only lose about 19 euros per capita. Next, he analyzed cases for public transportation shutdowns and found that it contributes to only 1 percent of cases annually, so shutting down would not have a big impact. Also, he reasoned that transportation shutdowns would be a loss for companies since workers would not be able to get to their offices on time. He concluded that for viral disease with above-average death tolls the school closure is the best method, as it saves more money and reduces cases more effectively than a public transportation closure (Adda, 2016).

This model is a prime example of how indirect models can help researchers find solutions because the model that he used did not directly predict economic loss. Instead, it predicted the incidence rate, and then he used those predictions to estimate economic loss for the school closure policy and public transportation policy.

Probabilistic Models

Other than predicting economic loss, models, specifically probabilistic models, help researchers predict exposure risk for viruses.

In hopes of finding a good balance between a full quarantine and reopening, the United States has had a record increase in COVID-19 cases (Shumaker, 2020). Due to the great detrimental health, social, and economic effects of contracting COVID-19, Americans have to make difficult decisions when deciding to visit public establishments, such as grocery stores, schools, malls, and theaters. It is vital that a system/tool exists which will aid in making these risky decisions. If the public is given an estimation of risk, they will have the necessary information to make responsible decisions, helping prevent unnecessary cases.

Functionality

One probabilistic model, by Sun, Di, Sprigg, Tong, and Casal (2020), predicts the probability that someone actively has the SARS-CoV-2 virus in an establishment. They took an interesting approach at the model, basing it on the popular birthday paradox problem: if *n* people are in a room what is the probability that *r* people have the same birthday. Similarly, in terms of COVID-19, the problem is what is the probability *Pr* that one person has COVID - 19 in a room of *n* people. The researchers created this model based on the problem:

$$Pr_{(p,n,a)} = \begin{cases} 1 - \frac{(p - a_i + a_{i-d})^n_!}{p^n_!}, \; if\ p \neq 0\ and\ n \neq 0\ and\ a_i \neq 0 \\ 0, \; if\ p = 0\ or\ n = 0\ or\ a_i = 0 \end{cases}$$

The model predicts Pr, the probability that at least one person has the virus base on factors such as p, the total community population; a, the total number of COVID-19 cases in that population on day i and day $i - d$, where d is the number of days it takes someone to recover or die; and n, the number of people in the establishment. With a risk of less than 25% considered low, 25 - 50 % meaning medium, 50 -75% as high, and greater than 75% meaning extremely high risk of encountering someone with SARS-COV-2, the model outputs a number between 0 and 1, which is interpreted as the percentage risk (Sun et al., 2020).

This model is much simpler than the incidence rate model for economic prediction because, unlike influenza and chickenpox, there is not a great amount of information available for COVID-19. The researchers who created this model intended it to be used for all counties in the United States, so they were restrained to only use general information. Data regarding the relation of temperature, transmissive properties of the virus, and ecology were not known. If these factors were used, not only could it result in an erroneous model, but also a model not compatible with all counties in the US (Sun et al. 2020)

<u>Presentation of the Model</u>

Along with creating the model, it was also very important how the researchers presented the model and made it as easy as possible for users to gain their information. Firstly, almost all researchers presented the model in the form of a web app. These web apps can be utilized by anyone with an internet connection and a compatible device, allowing them to spread information efficiently to a larger audience.

Another feature that these websites had was informative and eye-catching visuals. For example, the model presented in the article by Chande, Lee, and Harris (2020) had a map which updated instantly as the user inputted their values. With a darker shade of orange meaning higher risk, the map was intuitive and operated with no lag, giving users a great experience on the risk assessment tool.

An aspect that made these tools easier to use was concise inputs. For example, the model presented in the article by Sun et al (2020) simplified its three inputs to only a zip code. The tool then used this zip code in conjunction with a data set from Johns Hopkins University to locate the three specific inputs for the model (Sun et al., 2020). Although a simple feature, it increased the usability of the tool and made it more effective.

<u>Computer Learning Systems</u>

Occasionally, models need to acquire data that cannot be easily obtained. In that case, models estimate the indicator using computer learning systems. An example of a difficult indicator to acquire is real-time population density at specific establishments or venues. For this, a model could use object detection systems that use computer learning to analyze parking lots and obtain population density.

Machine Learning vs Deep Learning:

Computer learning systems allow scientists and researchers to utilize computers to their fullest potential and predict outcomes through the analysis of various factors. There are two types of computer learning systems: machine learning and deep learning.

Generally, machine learning uses algorithms to analyze data, learn from the data, and apply the learning to make a prediction. Instead of computer scientists "hard coding" algorithms for a computer, the computer is trained on a large data set and learns on its own. However, some human intervention is still required to provide the machine with the right classifiers - template algorithms that computers build off of - to perform the task (Copel, 2016). In contrast, deep learning, a form of machine learning that is inspired by the biology of human brains, uses artificial neural networks. Neural networks use complex mathematical equations, activation functions, and weights to create "connections" between the simulated neurons. Deep learning algorithms are different from machine learning algorithms in the sense that they do not need predetermined classifiers, giving them an advantage over traditional machine learning. However, training neural networks is computationally intensive, which made them popular only recently with advancements to hardware, specifically, the GPU and RAM (Copel, 2016).

The machine learning systems use a variety of classifiers to predict outputs from their respective datasets. In the area of machine learning, for example, if one wishes to simply predict a value from a fairly linear data set of values, they could use a Simple Linear Regression Classification Model. On the other hand, if one wanted to predict a value from a highly complex data set (*Figure 1*) with unconventional correlation, they could utilize a Kernel Support Vector Machine (SVM) Classification. Specifically, they could use the SVM with the Gaussian RBF Kernel, to plot the data points in a 3D space (*Figure 2*) and classify them using that information (Eremenko & Ponteves, 2020). In other words, the Simple Linear Regression Classification and the SVM Classification algorithm are two prime examples that portray the immense range in complexity that machine learning models can have depending on the problem they are presented with.

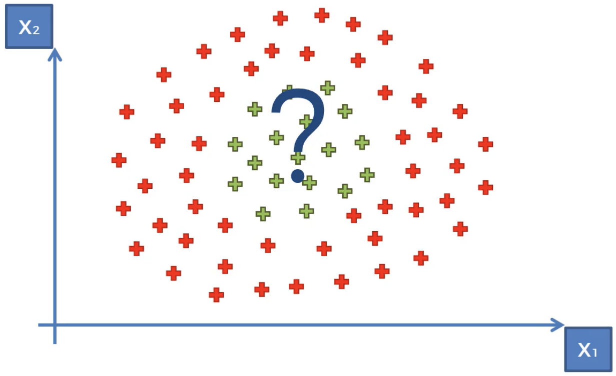

Figure 1. *Data Set with Unconventional Correlated Points: This data set has points that are unconventionally correlated, making a simple model impractical to use.*

Source: Eremenko, K., Ponteves, H. D. (Producers). (2020). Kernel SVM Intuition [Video file]. Retrieved December 21, 2020, from https://udemy.com/course/machinelearning/learn/lecture/6113144#overview

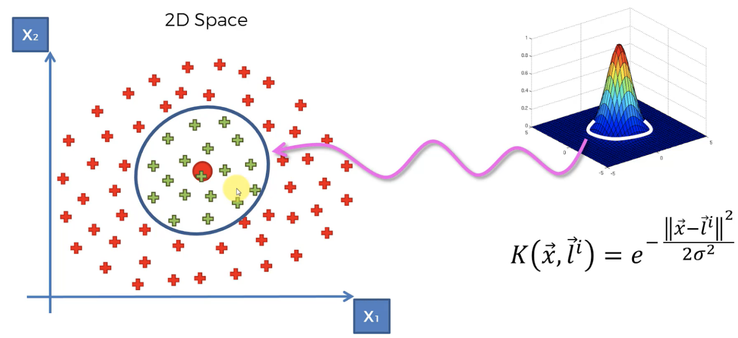

Figure 2. *Visualization of RBF Kernel Categorizing Data: The Gaussian RBF Kernel (bottom right) applies two vectors to plot the data set in a 3D plane, which then allows the model to differentiate the data points in the center from the data points in the outer area.*

Source: Eremenko, K., Ponteves, H. D. (Producers). (2020). Kernel SVM Intuition [Video file]. Retrieved December 21, 2020, from https://udemy.com/course/machinelearning/learn/lecture/6113144#overview

Object Detection Systems

Object detection, also known as computer vision, refers to algorithms that allow computers to identify, track, and label subjects in an image or video. There are two types of object detection systems: ones that employ deep learning and ones that utilize machine learning. Machine learning algorithms analyze a multitude of features in a picture, such as the color histogram, pixels which may be part of a larger object, and edges of objects (Object Detection Guide | Fritz Ai, n.d.). Those features are then fed into a machine learning regression model that can track and label various objects in an image. On the other hand, deep learning systems use convolutional neural networks (CNN), which learn to extract various features on their own to identify objects in an image. In this paper, the deep learning algorithm YOLO and a real-world application will be discussed.

YOLO Algorithm (You Only Look Once)

To perform object detection, there are many algorithms, and the You Only Look Once algorithm is one of them. YOLO is a highly accurate real-time deep learning object detection algorithm, and it is efficient since it requires only "one look" through the neural networks to make predictions. The outputs of this algorithm are bounding boxes around various objects in the image. YOLO has three main benefits: it is very fast; it analyzes an entire image when it is training, allowing it to find a correlation between various objects; and YOLO can identify generalizable objects, such as patterns in artwork (“Overview of the YOLO Object Detection Algorithm,” 2018). Compared to others, this algorithm is suited to run on slower hardware because of its efficiency and easy implementation.

Example: Parking Lot Vacancy Detection

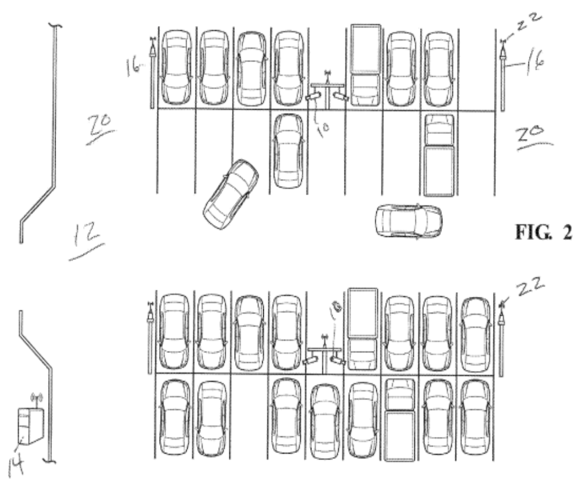

Figure 3. Diagram of Parking Lot
Source: Osment, M. (2015). U.S. Patent No. 20150310745A1. Washington, DC: U.S. Patent and Trademark Office.

Labels:
10 - cameras

14 - operative receiver to receive the output of the cameras in a wired or wireless manner

16- signal indicators visible to drivers indicate whether there are parking spaces available or not

18 - elevated mounts strategically placed in the parking lot so all of the parking spaces are captured

One real-world example of object detection in action is in parking lot vacancy detection. In these systems, CCTV cameras (#10) are mounted on elevated supports (#18) throughout the parking lot and are aimed towards the parking spaces themselves (Figure 3) These cameras use object detection algorithms, similar to YOLO, to detect vacancies in parking spots. The cameras' outputs are transferred to a server (#14) to store, the information and disperse it when necessary. The information is dispersed to drivers through a light (#16) at the end of each parking row (Figure 3). The indicators either turn green if there are vacant spaces or red if there are none (Osment, 2015).

This system is less expensive when compared to other vacancy detection systems because of the minimal materials needed; other systems need to invest in sensors for every single parking space while the object detection system requires a few strategically placed cameras. Second, this system can work in difficult situations, like snow or heavy rain, where major reference points, such as parking lines, are obscured. In these cases, instead of looking for vacancies in parking spaces, the object detection system can adapt and detect large enough gaps for cars to park in (Osment, 2015). The adaptability also allows this system to be used in scenarios where traditional parking lots are not present, such as cities with parallel parking or parking area lots with unmarked spaces. This adaptability allows this system to be easily used for other cases such as estimating the population at the parking lot's establishment or event.

Conclusion

Machine learning and deep learning serve as great tools for training computers to analyze and predict various information. Specifically, with easy to implement algorithms like YOLO, parking lot vacancy tools can be created and utilized. Because they use cameras, these systems are highly adaptable and with slight modification can be used to estimate the population inside an establishment. Finally, through scientific modeling, researchers can create valuable tools that aid various users like governments, businesses, and the public to predict important information such as economic loss or potential risks when visiting public establishments. In addition, both of these technologies can be used in conjunction with each other, making a more accurate model that can factor in population density.

References

Adda, J. (2016). Economic activity and the spread of viral diseases: Evidence from high frequency data. *The Quarterly Journal of Economics*, *131*(2), 891–941.

Amato, G., Carrara, F., Falchi, F., Gennaro, C., Meghini, C., & Vairo, C. (2017). Deep learning for decentralized parking lot occupancy detection. Expert Systems with Applications, 72, 327–334. https://doi.org/10.1016/j.eswa.2016.10.055

Chande, A., Lee, S., Harris, M., Nguyen, Q., Beckett, S. J., Hilley, T., Andris, C., & Weitz, J. S. (2020). Real-time, interactive website for US-county-level COVID-19 event risk assessment. Nature Human Behaviour, 4(12), 1313–1319. https://doi.org/10.1038/s41562-020-01000-9

Copel, M. (2016, July 29). *The difference between ai, machine learning, and deep learning? | nvidia blog*. The Official NVIDIA Blog.

Eremenko, K., Ponteves, H. D. (Producers). (2020). Kernel SVM Intuition [Video file]. Retrieved December 21, 2020, from https://udemy.com/course/machinelearning/learn/lecture/6113144#overview

Freitas, I. (2020, April 3). Economic damage rises as coronavirus hit worsens – here's what to watch. CNBC. https://www.cnbc.com/2020/04/03/us-economy-loses-701000-jobs-in-march-heres-what-to-watch-now.html.

Object detection guide | fritz ai. (n.d.). Retrieved December 21, 2020, from https://www.fritz.ai/object-detection/

Osment, M. (2015). U.S. Patent No. 20150310745A1. Washington, DC: U.S. Patent and Trademark Office.

Overview of the YOLO Object Detection Algorithm. (2018, September 25). Medium.

Oh, S., Watts, W. (2020, October 26). *Dow ends 650 points lower as rising COVID-19 cases, stalled stimulus efforts highlight recovery jitters*. MarketWatch. Retrieved November 4, 2020, from https://www.marketwatch.com/story/dow-futures-fall-260-points-as-covid-19-cases-surge-stimulus-remains-stalled-11603712231

Rio, G. M. N. del, & Bogel-Burroughs, N. (2020, October 23). 'At capacity': Covid-19 patients push u. S. Hospitals to brink. The New York Times. https://www.nytimes.com/2020/10/23/us/covid-hospitalizations.html

Science uses models to explain aspects of the real world. (2012, January 17). Science or Not? https://scienceornot.net/2012/01/17/science-uses-models-to-explain-aspects-of-the-real-world/

Shumaker, L. (2020, November 5). U.S. sets new record for increase in COVID-19 cases day after election. Reuters. https://www.reuters.com/article/us-health-coronavirus-usa-idUSKBN27L07L

Sun, Z., Di, L., Sprigg, W., Tong, D., & Casal, M. (2020). Community venue exposure risk estimator for the COVID-19 pandemic. Health & Place, 66, 102450.