Influenza Based COVID-19 Forecasting

Anonymous Authors[1]

1. Abstract

Currently at the time of writing, the COVID-19 pandemic has taken over 230 thousand lives in the United States and with our current efforts, shows very few signs of slowing down this pace. Though some measures of management have improved since the initial outbreaks of the pandemic, one of the biggest issues governments still face is understanding when and where a major outbreak will occur in order to implement precautionary measures. As this is a fairly new disease, many subjects relating to COVID-19 are still being researched. Given the nature of this disease, and more common diseases within it's family, most modelling methods mirror those of the Influenza virus. In attempting to model and predict COVID-19 spread, we will be using data from (for Disease Control & Prevention, 2020), and attempting to implement improved influenza modelling methods from (Zimmer & Yaesoubi, 2020). As will be seen later, we were able to gain some preliminary knowledge of the successes and failures of (Zimmer & Yaesoubi, 2020)'s model when applied to COVID-19 data. With respect to the biggest influences on this implementation, one major issue was a lack of COVID-19 data that we could use for training our model given how long the pandemic has been tracked. In looking forward to future related work, one major suggestion would be to solve the issue of a lack of data before moving on to improving model calculations and processes. As suggested later, one potential solution to this issue may be to use a Generetive Adverserial Network styled architecture to generate more data for COVID-19 models to use.

2. Review and Critique of (Zimmer & Yaesoubi, 2020)

2.1. Summary of (Zimmer & Yaesoubi, 2020)

In the paper (Zimmer & Yaesoubi, 2020), the authors take a look into the seasonal influenza epidemic(s). More specifically, they ponder the challenge of modelling this seasonal epidemic, as posed by the Center for Disease Control (CDC). Currently, one of the biggest impediments in real-time forecasts for seasonal influenza is the delay in releases of data sets from the CDC. More specifically, the CDC tends to have a 1-3 week delay in releasing their data, which allows for models to miss valuable information, such as major spikes and declines in cases which can greatly affect how forecasting models may perform.

In reviewing this, (Zimmer & Yaesoubi, 2020) breaks down two schools of approaches, one of which attempts to find cheap and easily accessible data for forecasting, and the other of which attempts to improve upon models used in forecasting. In contributing to the ongoing topic of disease forecasting, this paper first proposes a Gaussian Process (GP) which optimizes parameters by using the variance, as well as the squared diagonal matrix of eigenvalues.

Next, (Zimmer & Yaesoubi, 2020) proposes a method of how to implement this GP, in three simple steps. Firstly, (Zimmer & Yaesoubi, 2020) suggests narrowing down the weeks where a given epidemic is considered active, and breaking this data into weeks. Next, (Zimmer & Yaesoubi, 2020) proposes to use data from years prior for training, and some of the current epidemic's data as testing. Finally, this model is used to map a prediction of a given week(s) from a prior year to the current epidemic.

Finally, (Zimmer & Yaesoubi, 2020) introduces a more mathematical consideration for this modelling. One of the primary focuses of this is to introduce the proof that this method can tighten the range of predictions from a given model. Of the ideas presented in this section, one of the most important to this is reinforcing the selection of data where an epidemic is active. Put simply, the data outside of these active weeks are expected to be noisy and forces the model to fit points that are much less important for consideration.

2.2. Critical Review of (Zimmer & Yaesoubi, 2020)

Though (Zimmer & Yaesoubi, 2020) takes a somewhat unique approach to influenza forecasting, there are some assumptions and further information to be considered. One of the first assumptions to be mentioned is related to that of the GP. Namely, it is referenced that data of prior years is used to improve the model for a current year. Though

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

this is a fairly useful method, the paper fails to mention how it would approach the issue of time-invariance (e.g if an epidemic is much shorter/longer than prior years). One of the other large assumptions taken in (Zimmer & Yaesoubi, 2020), that becomes increasingly relevant in this and coming years is that of overlapping epidemics. Though at the time of this paper being written, it can be safe to assume an overlapping epidemic was not expected, it is still worth considering multiple factors for modelling. Given our current situation, we have to wonder how this model may react to predicting influenza, given active precautions due to our current pandemic.

3. Review and Critique of (Lubars & Tan, 2019)

3.1. Summary of (Lubars & Tan, 2019)

Moving on to (Lubars & Tan, 2019), the authors of this paper step back from the more analytical nature of AI and approaches the larger moral question of where the line is of what AI should and should not do. One of the growing concerns in the field of AI is not what can be accomplished, nor what can be improved, but rather what should be acceptable to do. Throughout the conception of artificial intelligence humans have always wondered what the outer limits of artificial intelligence is and its potential for mishaps. Referring prior work, (Lubars & Tan, 2019) breaks this topic of AI into three groups, all of which share the same theme of human opinion and preferences, which (Lubars & Tan, 2019) uses to dive deeper on the subject.

In first analyzing what humans take into consideration when delegating AI, (Lubars & Tan, 2019) suggests human motivation to da a task, task difficulty, risk of a task, and the general trust for AI in a situation. Firstly, it is assumed that with lesser human motivation, a task is less likely to be accomplished, and with a lesser human motivation brings a greater preference for AI to take over a task. On a similar note, with respect to difficulty, it is simply assumed that humans prefer low effort and expertise tasks. In other words, many tasks delegated to AI are those that humans can find repetitive, or far too complex. Moving on to risk, the main subfactors considered are accountability, uncertainty, and impact. With respect to all three of these (Lubars & Tan, 2019) considers who is to blame for AI mistakes, whether mistakes are likely, and how influential mistakes may be. Finally, with trust, (Lubars & Tan, 2019) states that humans tend to look at model accuracy, explainable AI (as in why decisions were made), and whether the model's decisions agree with that of the user.

In analyzing these positions through surveys there is one major conclusion made, which somewhat reframes how this process can be approached. With persons surveyed, most strongly ignore the concept of delegation (in the sense of doing an entire topic without supervision), and consistently opt to allow for AI assistance at most. Though interesting, this conclusion provides an interesting thought, to be considered in the critical review section.

3.2. Critical Review of (Lubars & Tan, 2019)

As hinted at in the end of the (Lubars & Tan, 2019) summary, the topic of AI delegation versus AI assistance is an interesting idea, which largely ignores the idea of how AI perspectives change with time. Expanding upon this, it may be interesting to consider what people deem acceptable with respect to AI, however this largely ignores potential advancements within AI, and how this affects human perception. As an example, if a certain task that AI currently cannot accomplish sees a breakthrough, with more success than humans counterparts, this may affect what people deem as an AI acceptable task. In terms of the future, this study does fail to consider how perspectives my change with advances in the field of AI. However, with respect to current times, (Lubars & Tan, 2019) does prove to represent human opinions on AI fairly well, but in order to improve and generalize, (Lubars & Tan, 2019) should look deeper at what root thoughts it's four main factors are tied to with respect to human decision making.

4. Review and Critique of (Graßhoff et al., 2020)

4.1. Summary of (Graßhoff et al., 2020)

In (Graßhoff et al., 2020), the authors take a higher level perspective on Gaussian Processes (GP), a topic critical to the improved modelling of influenza, however this approach mainly focuses on what can be done to reduce the time and space complexities associated with these GPs. Put simply, a GP is a process that attempts to fit, or regress, a Gaussian distribution among many sets of features. Currently, GP algorithms are very inefficient in terms of their computational complexities. With respect to time complexity, GPs are typically near cubic, and with space complexity, GPs can perform fairly quadratic.

In adding to this topic, (Graßhoff et al., 2020) proposes a series of scenarios in reducing computational complexity of a given GP. Firstly, with respect to kernels whose points show inconsistent distancing, it is suggested that interpolation can be used between points to adjust matrices and better approximate points. In reducing computational complexity, this allows for a more generalized approach to warped kernels. Moving on to one-dimensional functions, these can use exhaustive computation that can further be generalized with the above warped kernel scenario, assuming the kernel has a period of 2 * pi. Both of the above conversions result in

computationally linear operations, significant improvements from prior kernels with cubic complexities.

Using various biomedical datasets to test this, (Graßhoff et al., 2020) demonstrates success in improving GP kernel runtime, with each set used significantly outperforming it's unoptimized counterpart. It is worth noting that in accounting for noise in the data, noise injection was used on the original dataset.

4.2. Critical Review of (Graßhoff et al., 2020)

Though a generally complex operation, (Graßhoff et al., 2020) shows to notably improve Gaussian Process time complexities in practice. One interesting note to be made in the experimentation of the improved Gaussian Process was the change in space complexity. Discussed earlier in the paper was the conversion from quadratic space complexity to near linear complexity, however there is no mention of space complexity in experimentation. Though the very positive improvement (Graßhoff et al., 2020) provides to time complexity, it is worth considering what affect, if any the improvements pose on space complexity, and if the potential trade-off of time versus space is worth using.

5. Description of (Zimmer & Yaesoubi, 2020) Implementation

With respect to (Zimmer & Yaesoubi, 2020), the next few subsections will aim to implement the ideas presented in the paper. We will be looking at our results, generally, with a goal of an accurate disease forecaster that could, in theory, aid governments in responding to disease outbreak other than Influenza (in our case COVID-19).

5.1. (Zimmer & Yaesoubi, 2020) Method Breakdown

In implementing the ideas presented in (Zimmer & Yaesoubi, 2020), I wanted to first narrow down what I hoped to test and what I hoped to learn from this paper. In narrowing this down, I focused in on two main ideas that attempted to improve upon current and previous disease modelling practices.

Firstly, in creating a model for this forecast, the authors of (Zimmer & Yaesoubi, 2020) suggest a pseudo code routine. In this routine, there is an outer loop that iterates for each week in a forecasting horizon T. I felt that this forecasting horizon was not clearly defined, so I would like to state that I interpreted this to be the amount of time we want to predict ahead of time (i.e. if we are in week 1 of a pandemic, forecasting horizon allows us to predict accurately up to week 1+T of the pandemic). Next, within this forecasting loop was another loop that iterates for each feature set N. In both our case and (Zimmer & Yaesoubi, 2020) a feature set corresponds to a given year of influenza data.

Once within these loops, we begin to start our forecasting by firstly assembling T training data inputs for both out X and Y variables (both of which will be described later in detail). Once assembled, we then move to train a gaussian process based on the aforementioned split data. Once this gaussian process is trained, we make a forecast, which results in a mean and standard deviation. This concludes the innermost for loop which iterates for each feature set N. After this, we build an ensemble forecast as a result of these N iterations.

Secondly, and somewhat unconventionally, (Zimmer & Yaesoubi, 2020) proposes that in analyzing data, we order it not as one whole set, but as an ordered set by year and then by week. This poses a very unique solution method, wherein we try to solve for a given pandemic year as opposed to generalizing all pandemic years into one model. Though I understand one of the main benefits of doing this (that being having a better model for a specific year), it is somewhat of a unique and unusual method that I feel may create more unique, yet useless models for forecasting.

In analyzing this data, we want to set a few goals of what we hope to gain or learn from implementing (Zimmer & Yaesoubi, 2020)'s methods. Firstly, we want to make sure that we can implement (Zimmer & Yaesoubi, 2020) with the data it was designed for. For this, we will have somewhat of a baseline model produced to predict Influenza data of recent years. Before moving onward, we hope to have this model performing at a fairly satisfactory level, namely with a coefficient of determination at or near 0.9.

Once we can confirm our original model method works, we then want to move on to testing COVID-19 forecasting. With this, we first want to try to fit a model to the present COVID-19 data from (for Disease Control & Prevention, 2020). Through doing this, there will be one major problem that could affect our model fitting, that problem being the lack of COVID-19 data present to train our model with. Because of this, our model may perform very inconsistently, and may even just perform poorly, as it is unable to train properly. In a very rough attempt to alleviate this issue, we will add one more exercise, where we will concatenate the COVID-19 and Influenza data together.

In doing all of this, we have one main goal, that being an accurate model to predict COVID-19 mortalities across the United States. As we will have three models trained and tested, we will want one final exercise to see how each of our three models may perform in predicting COVID-19 data. In hypothesizing the results of this exercise, I expect the Influenza model to perform the worst, as it is fit to a completely different set of data. I then expect the model created from combined COVID-19 and Influenza data to perform the best, as it has COVID-19 data with 'testing

data' though it is worth noting that 'testing data' will not be too similar to the COVID-19 data. Finally, I expect the model trained on COVID-19 data only to be somewhere in between these two models, as it is trained with a proper dataset, though not necessarily trained enough.

5.2. Data Used

In implementing this pseudo code, it is important to first touch on what data will be used. For this task, we will be using (for Disease Control & Prevention, 2020), a dataset provided from the Center for Disease Control and Prevention. With this data set are 11 features as follows: Year, Week, Percent of Deaths Due to Pneumonia and Influenza, Percent of Deaths Due to Pneumonia, Influenza or COVID-19, Expected, Threshold, All Deaths, Pneumonia Deaths, Influenza Deaths, COVID-19 Deaths, and Pneumonia, Influenza or COVID-19 Deaths.

When analyzing this set, I observed that the Pneumonia Deaths, Influenza Deaths, and COVID-19 Deaths would prove to be very useful. As a result of this, we will use this data on disease death as a primary indicator of disease presence, and attempt to model it. Though death may not necessarily indicate the infectiousness of a disease, the rate of deaths from a given disease can still pose useful information to governments trying to manage the disease.

With this data, there is one underlying assumption, that being the relation of pneumonia deaths. As pneumonia is commonly caused by coronavirus strains (e.g. Influenza and COVID-19) it is difficult to determine which disease may have caused the pneumonia. On top of this, it appears that (for Disease Control & Prevention, 2020) separates influenza deaths pneumonia deaths whenever possible. Given the current public attention of COVID, we will make an assumption with the data, in that (for Disease Control & Prevention, 2020) will mark COVID-19 deaths wherever possible, and that the pneumonia deaths were purely results of influenza complications.

5.3. Code Implementation

Moving on to the code implementation of (Zimmer & Yaesoubi, 2020), we can break our code down into eight simple sections that implement the paper's proposed algorithms. Firstly, we implement some python libraries to be used later, such as numpy, matplotlib.pyplot, pandas, and sklearn.

Our next two sections are related to importing and formatting the data from (for Disease Control & Prevention, 2020). With the first section, we read in the data, and create a dictionary structured with a given year being an index, and the data within that index corresponds to a list of pandemic week, and its associated influenza deaths.

With respect to the COVID-19 data, there is only one year provided, so indexing by year is not necessary. Instead, we format this data similarly to the Influenza data, with a corresponding pandemic week and COVID-19 related deaths list.

Moving on to the fourth section, we approach more of the analytical side of our code, with the design of our fitting and forecasting functions. As the COVID-19 data and the Influenza data were formatted slightly differently as well as the fact that we had far less COVID-19 data to train from, we needed to create two fairly similar functions to accomplish fitting and forecasting. With respect to the influenza data, I created a function that does the following.

Firstly, we initialize our forecast horizon, T, to be an arbitrarily chosen 8 week period. Next we initialize our feature set, N, variable to be the length of the inputted dictionary (i.e. how many years are in the set). Moving on, we create a kernel object based upon the ExpSineSquared kernel from the sklearn library with a length_scale of 1 and a periodicity of 2. The reasoning for the length_scale and periodicity variables being the values they are was purely a result of trial and error to understand which parameter values performed best.

Before going forward, it's very important to talk about the reasoning behind using ExpSineSquared as a kernel for our Gaussian Process. When the authors of (Zimmer & Yaesoubi, 2020) propose their methodology, they mention using a kernel of a covariance function. Given that this is not a fairly prevalent kernel within the sklearn library that we are using, the next best implementation would be to design a custom kernel for the Gaussian Process. When attempting to implement this however, I ran into many issues relating to the formatting of the data (which typically would be shaped differently for other problems). Given my limited understanding of kernels, I decided that in creating a custom covariance kernel I may cause more issues to the model than improvements. Given that the model used Gaussian Processes, and needed a kernel, I chose to research other available kernels on sklearn, and in my inquiry, I found the ExpSineSquared function. Given the somewhat inconsisted peaks and periodic nature of our data, I found this to be a fairly good replacement for the covariance kernel. It is worth noting that this ExpSineSquared should be used only to make shorter-termed forcasts, and it's longer-termed forcasts may be too inaccurate given that our model and process weren't exactly designed for its use.

Once our variables have been initialized, we move into our two for loops, one for T and one for N. Within the body of these, we break the inputted data into training and testing sets of roughly size T elements. After splitting, we then reshape each of the training and testing datasets for future use in our model. Once the data is properly formatted, we then initialize a Gaussian Process with a kernel as described

above, and we fit this Gaussian Process with our training data. Once trained, we add our model into a dictionary, where the stored index is the model's score on the testing data (by using the coefficient of determination), and within that index, we store the model and the data used for testing. After iterations, we then return this dictionary for future use.

As with the Influenza fitting function, most processes of the COVID-19 fitting function are the same, however one major difference is the feature set, N, variable. Because there is only one provided feature set, the original models of COVID-19 performed fairly poorly. Because of this, a slight modification of the N value, namely increasing it from 1 to 10 was implemented to improve COVID-19 model performance.

Once each of these models have been compiled, we then test three implementations of them. Firstly, we test influenza data with the influenza model to generate a baseline test to ensure the model from (Zimmer & Yaesoubi, 2020) performs well on its suggested data. Then we use our COVID-19 data from (for Disease Control & Prevention, 2020) and try to fit our COVID-19 model to this data. Finally, we combine our COVID-19 data with our influenza data, and attempt to fit using our influenza-based model. For each of these returned models, we print out the best performing model's score, and then plot the testing data vs the predicted data for each of the three models. Finally, we display each of these three models with their projections on just the COVID-19 testing data to see each model's performance with this dataset.

5.4. Evaluation Methods

As mentioned above, we will be using a coefficient of determination to evaluate each of our models. This has been chosen as the best possible evaluation score as, by definition, it illustrates the variance of predicted points from their actual values. As with a coefficient of determination, we desire a value as close to 1 as possible for the best performing model.

With some of the results discovered, it is important to note the significance of the 'bounds' of the coefficient of determination, those being [0,1]. In plain English, the coefficient of determination translates to what percentage of our model fits the present data. As mentioned above, we desire a value close to 1, which would signify a model that fits 100% of the data. Conversely, a coefficient near 0 signifies a model that fits 0% of the data.

When considering the bounds of this coefficient, it is actually possible for a coefficient to be less than zero. As some of the calculation methods suggest in finding the coefficient of determination, we actually are calculating the fit of a model as compared to a horizontal line of the average of the data. Because of this, it is more than possible that this coefficient to be less than zero, illustrating that the model is a worse fit than just using the horizontal average of the data.

5.5. Discussion of Results

Shown below are the associated figures for the Influenza, COVID-19, and Combined models. The coefficient of determination of each is 0.87, -0.92, and 0.87 respectively.

One of the first things that we can notice in this set of information would be the repeated occurrences of 0.87 as a coefficient of determination. Given the unusual splitting of the data, and the method of training models year by year, this may make a little sense. Because all of the Influenza data is contained within the Combined data, it is very likely that the pseudo-random training and testing splits were the same for each of these data sets.

With respect to all three sets, we were able to still see two fairly well performing models, however one big surprise is the failure of the COVID-19 model. One of the main contributing factors in this failure may be related to the small presence of training data points for this model. Because our data set contained only a small fraction of points as compared to our influenza model, the COVID-19 model was not able to make many changes to its initial state, which caused our Gaussian Process Regressor to return a model fairly close to a constant value, as opposed to a somewhat periodic fit.

Finally, with respect to figure 4, we can see the projections of each model onto the COVID-19 data. With each of these models, we can see a very poor fit on the dataset, which may suggest that our current methodologies need some modifications. In terms of things that we may want to look into for a future iteration of this project would be potentially implementing a method to generate realistic COVID-19 data similar to that of a generator in a Generative Adversarial Network. Though creating some code to generate realistic COVID-19 information may be difficult to train, this would still be a very valuable process, as more COVID-19 data points would allow for our model to iterate and improve for much more time.

5.6. Results

Shown below are the results of our implemented methods from (Zimmer & Yaesoubi, 2020). In the order that the figures appear, we see the model trained with only Influenza data, the model compiled with only COVID-19 data, and the model compiled with both of those two datasets. For each of the figures 1-3, we show the testing data, and the predictions with that testing data. Below, in figure 4 we see each of these figures mapped onto the COVID-19 data, with all of the true data of the COVID-19 model shown, as well

as predictions from each of the three models when using the COVID-19 model's testing data for predictions.

Figure 1. Influenza Model

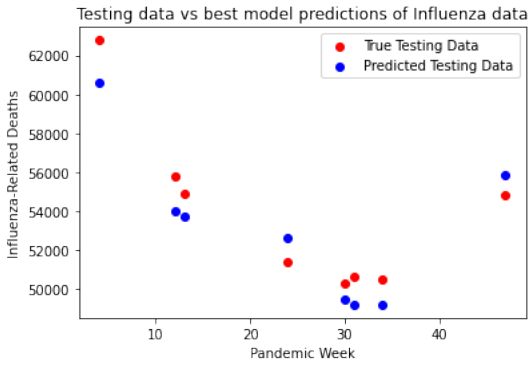


Figure 2. COVID-19 Model

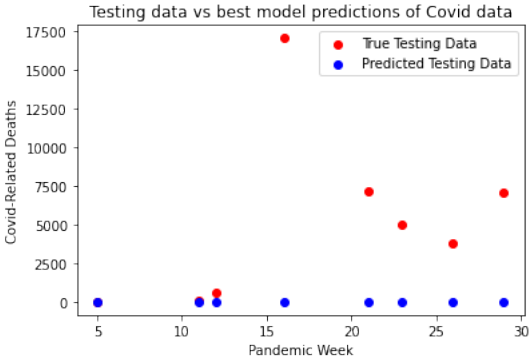


Figure 3. Combined Model

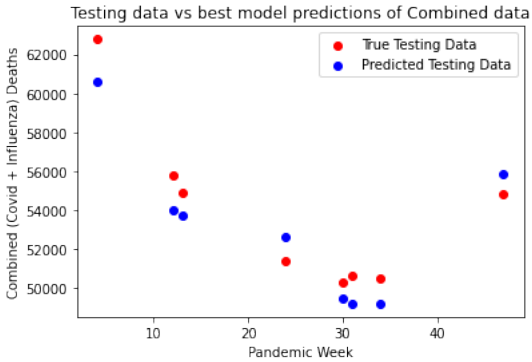
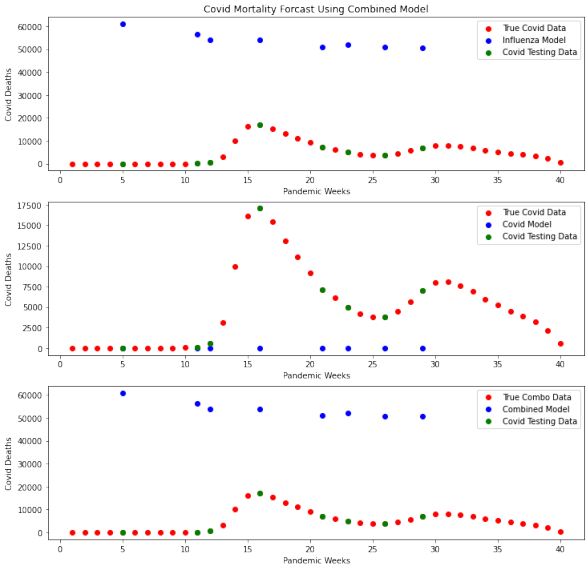


Figure 4. Models Transposed on COVID-19 Data



References

for Disease Control, C. and Prevention. Pneumonia, influenza, and covid-19 mortality from the national center for health statistics mortality surveillance system. In *Weekly U.S. Influenza Surveillance Report*, 2020. URL https://www.cdc.gov/flu/weekly/weeklyarchives2020-2021/data/NCHSData40.csv.

Graßhoff, J., Jankowski, A., and Rostalski, P. Scalable gaussian process separation for kernels with a non-stationary phase. In *International Conference on Machine Learning 2020 (ICML-2020)*, 2020. URL https://proceedings.icml.cc/static/paper_files/icml/2020/2702-Paper.pdf.

Lubars, B. and Tan, C. Ask not what ai can do, but what ai should do: Towards a framework of task delegatability. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, 2019. URL https://papers.nips.cc/paper/8301-ask-not-what-ai-can-do-but-what-ai-should-do-towards-a-framework-of-task-delegability.pdf.

Zimmer, C. and Yaesoubi, R. Influenza forcasting framework based on gaussian processes. In *International Conference on Machine Learning 2020 (ICML-2020)*, 2020. URL https://proceedings.icml.cc/static/paper_files/icml/2020/1239-Paper.pdf.