Predictive Monitoring of COVID-19

Jianxi Luo
Data-Driven Innovation Lab
Singapore University of Technology and Design

Working draft, updated on May 2, 2020

Abstract

During the current COVID-19 pandemic, there have been various efforts to forecast the infection cases, deaths and medical or economic indicators. Various models and methods have been adopted or developed for different prediction contexts and purposes. Some of the forecasting projects have influenced the policies in some countries. However, the prediction of future is uncertain by nature. No model or data can accurately represent the complex, dynamic and heterogenous realities of the pandemic in different countries. In this case, we do not aim to make perfect predictions about the future or test how accurate a prediction will be. Instead, to address the uncertainty of predictions in dynamic real-world scenarios, we explore the potentials of using predictive monitoring, or namely continually monitoring predictions, for not just stable predictions but also changes of predictions, to derive implications regarding what are happening now in the real world and also make the planning, behaviour and mentality now be more "future-informed".

Introduction

Since the outbreak of COVID-19 in January 2020, researchers around the world have adopted classic or latest data science and AI techniques and applied them to the data available to forecast the developments, trends and key dates of transitions or ending of COVID-19 in different countries or regions. The noticeable efforts include the publicly available and continually updated forecasts by the Institute of Health Metrics and Evaluation (IHME) at University of Washington [1], the MRC Centre for Global Infectious Disease Analysis at the Imperial College London [2], the University of Texas COVID-19 Modelling Consortium [3], and MIT IDSS Isolate [4]. Some methods focus on forecasting deaths and hospitality needs [5,6,7] and infection cases and peaks [8,9,10], while others focus on the impact of social distancing as a key operational control measure [7,11].

Some published studies have rigorously tested the precisions of certain prediction methods [8,10,12] mostly based on recent data from China. However, even the most noticeable forecasting method from the IHME has been found with model issues and high errors with tested with later coming actual "future" data [5,7,12]. Researchers are learning and improving the methods and tools on the go to in order to make more accurate predictions

on COVID-19 pandemic [5,12]. Despite the limitation and uncertainty of predictions, some forecasting methods, tools and studies have already influenced policies or informed policy makers to some extents and in certain ways, despite the intrinsic uncertain nature of forecasting or prediction, especially in the current complex and dynamic COVID019 pandemic environment across the global.

Given the value of predictions but also the difficulty to do it well in practice, we aim to explore the values and potentials of predictive monitoring to deal with the uncertainty of predictions and make use and make sense of prediction excises for suitable good. Predictive monitoring means the continual monitoring of predicted likely future events, such as the turning and ending of the ongoing pandemic, using the latest data generated daily, more for the monitoring purposes to derive understanding about what are happening now than the accurate prediction of the future. Predictive monitoring differs from the common prediction practice that makes a one-shot prediction and then hope to see the prediction come true later. It also differs from the common monitoring practice that reports actual historical cases of infection, recovery and death every day.

The Predictive Monitoring Experiment

- Theory

Here we aim to experiment predictive monitoring in the realistic context of the on-going COVID-19, in order to explore its potentials and develop specific guidelines and strategies for the right use of it. To run the experiment, the first is to choose a model and data source for data-driven prediction, before we can update and monitor the updated predictions with more and more data coming in over time. The propagation of infectious diseases often follows a life cycle pattern, from the outbreak to the acceleration phase, inflection point, deacceleration phase and eventual ending. Such a life cycle is the result of the infection process, property of the virus, the nature of a population and the adaptive and countering behaviours of agents including individuals (avoiding physical contact) and governments (locking down cities) in the population. However, the pandemic life cycles vary by countries, and different countries might be in different phases of the life cycles at a same point in time.

For instance, on April 21, in Singapore, Prime Minister Hsien-Loong Lee announced the extension of circuit breaker to June 1 in response to the spikes of COVID-19 cases, on the same day when Prime Minister Giuseppe Conte announced Italy’s plan to reopen from May 4. Ideally speaking, such decisions and planning can be rationalized by well knowing where our own country (together with other countries and the world as a whole) is in its own pandemic life cycle, when the turning point is coming if it has yet come, and when the pandemic will end. Adjustments may be made according to the changes in the estimations

and predictions on these fronts. The basis for such actionable estimation is the pandemic's life cycle.

- Model

The pandemic life cycle pattern is expected to appear as a S-shape curve when one plots the accumulative count of infection cases over time or equivalently as a "bell-shape" curve of the daily counts over time (see examples in Figure 1). Note that the bell here is not expected to be symmetrical with no expectation of a normal distribution, but a long tail to the right. Such patterns as well as the underlying dynamics have been well studied in various domains including population growth, diffusion of new technologies and infectious diseases, and have theoretically established mathematical models, including the logistic model that describes a general life cycle phenomenon (such as population growth) and the SIR (susceptible-infected-recovered) model [13,14,15] that describes the dynamic process of the spread of infectious diseases.

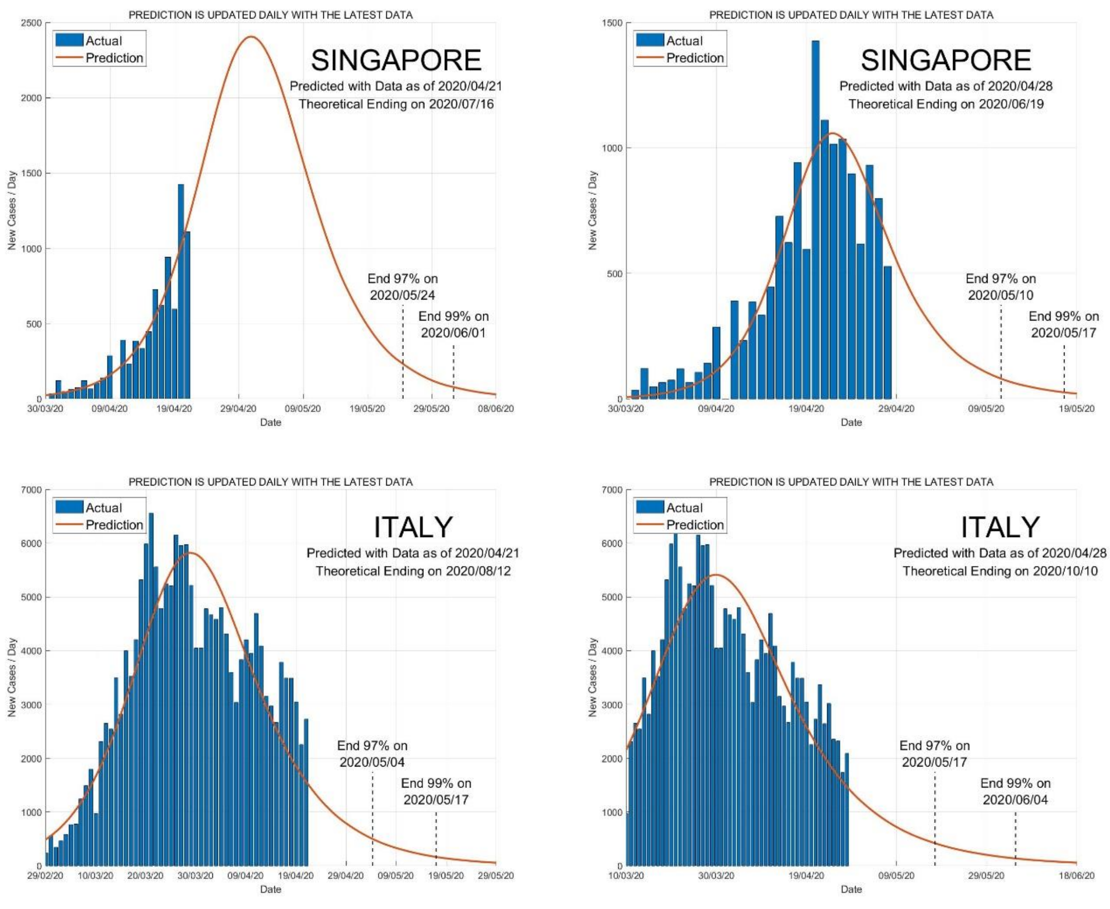

Figure 1. Continuous Data-Driven Estimations of COVID-19 Life Cycle, Turning and Ending Dates for Singapore and Italy as of April 21 versus April 28, 2020

The SIR model is employed in this experiment for a few reasons. One, it is context-specific and models the dynamic process of inflections in a population over time. Second, it requires

simple data inputs that are publicly available. Third, there are open source computer codes available for quick adoption. Here we will not repeat the details of the SIR model in this paper, which can be easily found in many mathematics textbooks. Essentially, the SIR model incorporates two main parameters, beta and gamma, for computing three key variables including S (number of susceptible people), I (number of infected people) and R (number of people that have either recovered, died or immured from the disease). Beta is the number of days one is contagious and a property of the virus. Gamma is the average number of people infected by a previously infected person and results from not only the interaction patterns of people in the society (which social distancing can influence) but also the infection process property of the virus.

- Implementation

The values of these two parameters fundamentally determine the shape of an infectious disease's specific life cycle curve for a population. In particular, the model (with three equations for S, I and R in its original form) can be reduced to one function about the total infection count, or equivalently the daily new infection counts. This key variable is the sum of the variables I and R and has publicly available data reported by official channels every day. Please refer to this paper [16] by Milan Batista for the model reduction. Therefore, only the data of the total infection account (which can be also used to derive the daily new case count) is required to regress the two key parameters and other constants and thus train a model that estimates the dynamic pandemic curve.

Batista also developed open-source computer codes to implement the regression using the reduced function [17]. In our experiments, we applied the codes of Batisa to the COVID-19 accumulative infection data for each country from "Our World in Data" [18] to regress the parameters and constants of the basic SIR model. Note that, more sophisticated versions of the SIR model, such as the SEIR model, have also been used to predict the turning and ending points of COVID-19 in China, but they require more sophisticated data inputs which we do not have. Regressions are run for individual countries and updated daily with the newest accumulative and daily infection count data becoming available daily. Not the data for all countries can produce statistically meaningful regression results. Only the countries with satisfactory goodness-of-fit between model and data as measured by R^2 greater than 0.8 are accepted, analysed and reported. For these countries, the regressed model for each of them is used to estimate the full pandemic life cycle and plot the life cycle curve.

Makes Sense of the Predictive Monitoring

As shown in the examples in Figure 1, the initial segment of the curve is fitted with the data to date and the remaining segment of the curve is predicted. With the estimated full life cycle curve, one can easily observe which phase of the pandemic life cycle a specific country

is in (with actual data plotted together), when the inflection point (the peak in the bell-shape curve) is coming (for the interests of the countries still in the accelerating phase), and when the pandemic will end (for the interests of all countries). The inflection point is specific and appears as the peak in the bell-shape curve. However, estimating the “ending date” is not straight-forward and may be done differently for different considerations.

Most theoretically, one can define the pandemic’s end date as the day with the last infection case of the pandemic, and thus operationalize the estimation of the end date as the day with the last predicted infection at the right most end of the estimated pandemic life cycle curve. However, practically, estimation of the theoretical ending might not be useful to provide guidance for the planning of activities of governments, companies and individuals. One might consider an early date when predominately most predicted infections (indicated by the regressed pandemic life cycle curve) have been actualized and only a small portion of the total predicted epidemic population is left. The total predicted infection population size is the area under the curve. In our experiments, we have monitored three alternative estimates of end dates in the order of conservativeness.

- The date to reach the last expected case;
- The date to reach 99% of the total expected cases;
- The date to reach 97% of the total expected cases.

In any case, specifying an end date is arbitrary in nature. With uncertainty and flexibility in mind, one may simply just exploit the estimated life cycle curve, especially its right most tail segment, to screen and sense when the pandemic gradually vanishes to which extent.

It is noteworthy that the bell-shape curve is chosen to visualize the life cycle because it allows easy detection of the inflection point as the peak of the curve to distinguish countries in acceleration and deacceleration phases. For instance, Figure 1A visually reveals on April 21 Singapore was still in its acceleration phase, whereas Figure 1C shows Italy has passed its inflection point. At the time, the estimated “future” turning date for Singapore would be May 1. However, as shown in Figure 1B, on April 28, Singapore has already past its inflection point, earlier than the turning date predicted on 21 April (in Figure 1A). In contrast, from April 21 (Figure 1C) to April 28 (Figure 1D), the curves of Italy are slightly lifted, and the later predictions for Italy suggest consistently later 97%, 99% and 100% ending dates.

These changes are discovered through <u>predictive monitoring</u>, in which we continually monitor the predictions, not really hoping the previous predictions to be tested true later when the real “future” comes, but for detecting in the changes of the predictions over time. From a traditional perspective, the difference between a future prediction and a previous one on the same variable would be considered a bad thing and a proof of failure of the prediction model. However, here we tend to make sense of such changes from the earlier to later predictions for the signals as to what are happening in the dynamically changing real-

world scenarios, with the assumptions that predictions made over time should be different when the real-world scenarios are changing.

In other words, our default expectation in predictive monitoring is that predictions will change, especially when the real-world scenarios, such as government policies and human behaviours, are also rapidly changing. In such cases, we should not expect the model trained with data as of today to be true for a different scenario later. When considering the dynamics of human behaviours and government policies and other real-world scenarios that the mathematical model and training data cannot accurate represent, predictive monitoring would be a more valuable exercise, than making a prediction now to it is a hit or miss in future. Therefore, the changes in the predicted theoretical events (such as ending) may indicate the uncertainty in the environment. So, we also report the standard deviations in N latest and connectively predicted theoretical end dates as an indicator of uncertainty. If the standard deviation of the connective predicted ending dates is small (regardless of their accuracy), it indicates the real-world scenarios are not changing. If it is high, it might imply changes are happening in the real-world scenarios.

For example, the changed predictions of the theoretical pandemic end dates of Singapore over time may reveal the effects of the recently strengthened measures of the Singapore government and more cautions of the local citizen from PM Lee's announcement of circuit breaker extension on 21 April. The changed predictions of pandemic end dates for Italy may result from the slightly relaxed government control measures and human behaviours in Italy in the past week. The pandemic curves of Singapore and Italy have shifted over time, as the real-world scenarios have dynamically changed. It would be wrong to expect the curve estimated with data from the previous scenario to represent the curve for a later scenario. Instead, the curves should be continually re-estimated with the latest data, the predictions based on these curves should be continually monitored, and the changes in the predictions may indicate changes in real-world scenario changes over time. Monitoring and detecting such changes in the predictions provides the main value of predictive monitoring.

In sum, these examples here suggest the importance of predictive monitoring or continually monitoring predictions to address uncertainty, detect and evaluate changes (such as human behavers and government control measures) made in the dynamic real-world scenarios in real time. It also allows the estimation of the uncertainty or stability of the predictions themselves as a result of the underlying real-world scenarios. Predictive monitoring differs from making a one-shot prediction for it to become true in the future and differs from the monitoring of actual cases every day.

Caution

Predictive monitoring for each country should be read together with what are happening in the real world and government policy changes. For instance, Singapore government's strengthened restrictions in April may have bended its curve earlier than the previously predicted ones, and the early relaxation of social distancing and lockdown in Italy and Germany might increase infection rates and thus delay the pandemic ending as predicted now. Also, the predictive monitoring of a country should not be read in isolation, but together with the predictions and real time situations of other countries. No country is in isolation in the world today. The monitoring and control of one country must be coupled with the monitoring and control of other countries.

For example, while the predictive monitoring shows the pandemic has "theoretically" ended in China, South Korea and Australia (despite a small number of domestic cases reported daily), it also shows the world will still suffer till the end of 2020 if we remain in our present trajectories of government policies and individual behaviours and without medical cures and vaccines for COVID-19. Therefore, the governments of China, South Korea and Australia may not want to open their international ports so soon and lift the domestic restrictions so quickly, until the pandemic nears its end in the world as whole. Although it is the time for all of us to isolate and distance physically from each other, it is also the time that needs more sharing of data, information and knowledge and more close coordination.

Because of the complex, dynamic and heterogenous realities in different countries, the curve, inflection and end dates must be continuously re-estimated with the newest data from official channels every day. That is, the predictions themselves are also needed to be monitored over time, in addition to monitoring the actual cases [19]. Especially, for countries that are still early in their own pandemic life cycles, the prediction of the rest of the curve, inflection point and ending dates will be more teasing, but also inherent less relevant to the "real future" to come given that the actual data only cover a smaller and early portion of the total life cycle and many real-world scenarios that the model cannot describe are expected to change. By contrast, for countries that have passed their inflection points and been approaching ending phases, prediction is expected to be more accurate because it is based on data covering more different phases of the life cycle, but also less useful when uncertainty is low. When uncertainty is low, it is more likely that we can derive and approve a highly predictive model. However, in such cases, the trained model is more about explaining the history and less about predicting the future. For those countries, a new epidemic wave might come if the governments and individuals lift controls and disciplines too early, especially when the pandemic is still prevalent in other countries.

Summary

Data-driven predictive monitoring may completement the traditional historical case monitoring practice and the traditional accuracy-expected prediction practices to deliver

additional insights. The value of continuous predictive monitoring might be greater when the real-world scenarios that the models cannot describe are inherently dynamic and more uncertain. We will continually monitor the estimated pandemic life cycle curves and end dates and explore valuable insights from the monitored prediction changes, as an experiment to explore the potentials of as well as develop guidelines and strategies for valuable predictive monitoring practices.

In the meantime, readers must take any prediction, regardless of the model and data, with caution. Over-optimism based on some predicted end dates is dangerous because it may loosen our disciplines and controls and cause the turnaround of the virus and infection. Although prediction based on science and data is aimed to be objective, it is uncertain by nature. One thing that is certain is that the model, data and prediction are inaccurate and insufficient to fully represent the complex, evolving, and heterogeneous realities of our world. The model we use in the experiment is only theoretically suitable for one stage or wave of the epidemic evolution, and relatively more meaningful when applied to data for each single stage if the country has experienced multiple stages (such as Singapore). The prediction is also conditioned by the quality of the data. The data publicly available today is based on tests, which are done differently in different countries and over time periods. They do not necessarily represent the total infection account which is the theoretical input of the model. One should expect changes in the continually monitored predictions, instead of fixed expectations.

Future is always uncertain. We must keep this in mind when doing and reading any prediction. No one predicted the COVID-19 outbreak beforehand. With acknowledging the uncertain nature of the ongoing COVID-19 pandemic and our growing inter-connected and complex world, what are eventually and fundamentally needed are the flexibility, robustness and resilience of people, organizations and governments, as well as sharing and coordination, to deal with unpredictable and unwanted future events.

References

[1] Institute of Health Metrics and Evaluation (IHME) at University of Washington https://covid19.healthdata.org/united-states-of-america

[2] the MRC Centre for Global Infectious Disease Analysis at the Imperial College. https://mrc-ide.github.io/covid19-short-term-forecasts/index.html

[3] University of Texas COVID-19 Modelling Consortium https://covid-19.tacc.utexas.edu/

[4] MIT Institute of Data, System and Society (IDSS) COVID-19 Collaboration (Isolat) https://idss.mit.edu/research/idss-covid-19-collaboration-isolat/

[5] IHME COVID-19 health service utilization forecasting team, Christopher JL Murray. Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries. MedRxiv 2020.04.21.20074732; doi: https://doi.org/10.1101/2020.04.21.20074732

[6] IHME, COVID-19 health service utilization forecasting team, and Christopher JL Murray. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months, 2020. https://www.medrxiv.org/content/10.1101/2020.03.27.20043752v1.

[7] Spencer Woody, Mauricio Garcia Tec, Maytal Dahan, Kelly Gaither, Michael Lachmann, Spencer Fox, Lauren Ancel Meyers, James G Scott. Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones. medRxiv 2020.04.16.20068163; doi: https://doi.org/10.1101/2020.04.16.20068163

[8] Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, Liu P, Cao X, Gao Z, Mai Z, Liang J, Liu X, Li S, Li Y, Ye F, Guan W, Yang Y, Li F, Luo S, Xie Y, Liu B, Wang Z, Zhang S, Wang Y, Zhong N, He J. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis 2020;12(3):165-174. doi: 10.21037/jtd.2020.02.64

[9] Verity R, Okell L, Dorigatti I, Winskill P, Whittaker C, et al. 2020. Estimates of the severity of coronavirus disease 2019: A model-based analysis. Lancet Infectious Diseases

[10] Petropoulos F, Makridakis S (2020) Forecasting the novel coronavirus COVID-19. PLoS ONE 15(3): e0231236. https://doi.org/10.1371/ journal.pone.0231236

[11] STEPHEN M. KISSLER, CHRISTINE TEDIJANTO, EDWARD GOLDSTEIN, YONATAN H. GRAD, MARC LIPSITCH. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science, PUBLISHED ONLINE, 14 APR, 2020 DOI: 10.1126/science.abb5793

[12] Roman Marchant, Noelle I. Samia, Ori Rosen, Martin A. Tanner, and Sally Cripps. Learning as we go: An examination of the statistical accuracy of covid19 daily death count predictions, 2020. doi: https://doi.org/10.1101/2020.04.11.20062257

[13] Hethcote, H. W., The mathematics of infectious diseases, SIAM Review 42, 599–653 (2000).

[14] David Smith and Lang Moore, "The SIR Model for Spread of Disease - The Differential Equation Model," Convergence (December 2004)

[15] The structure and function of complex networks, M. E. J. Newman, SIAM Review 45, 167-256 (2003)

[16] Milan Batista (2020) Estimation of the final size of the COVID-19 epidemic. Medarxiv https://www.medrxiv.org/content/10.1101/2020.02.16.20023606v5. doi: https://doi.org/10.1101/2020.02.16.20023606

[17] Milan Batista (2020). fitVirusCOVID19 (https://www.mathworks.com/matlabcentral/fileexchange/74658-fitviruscovid19), MATLAB Central File Exchange. Retrieved May 1, 2020.

[18] Our World in Data https://ourworldindata.org/coronavirus-source-data

[19] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis; published online Feb 19. https://coronavirus.jhu.edu/us-map