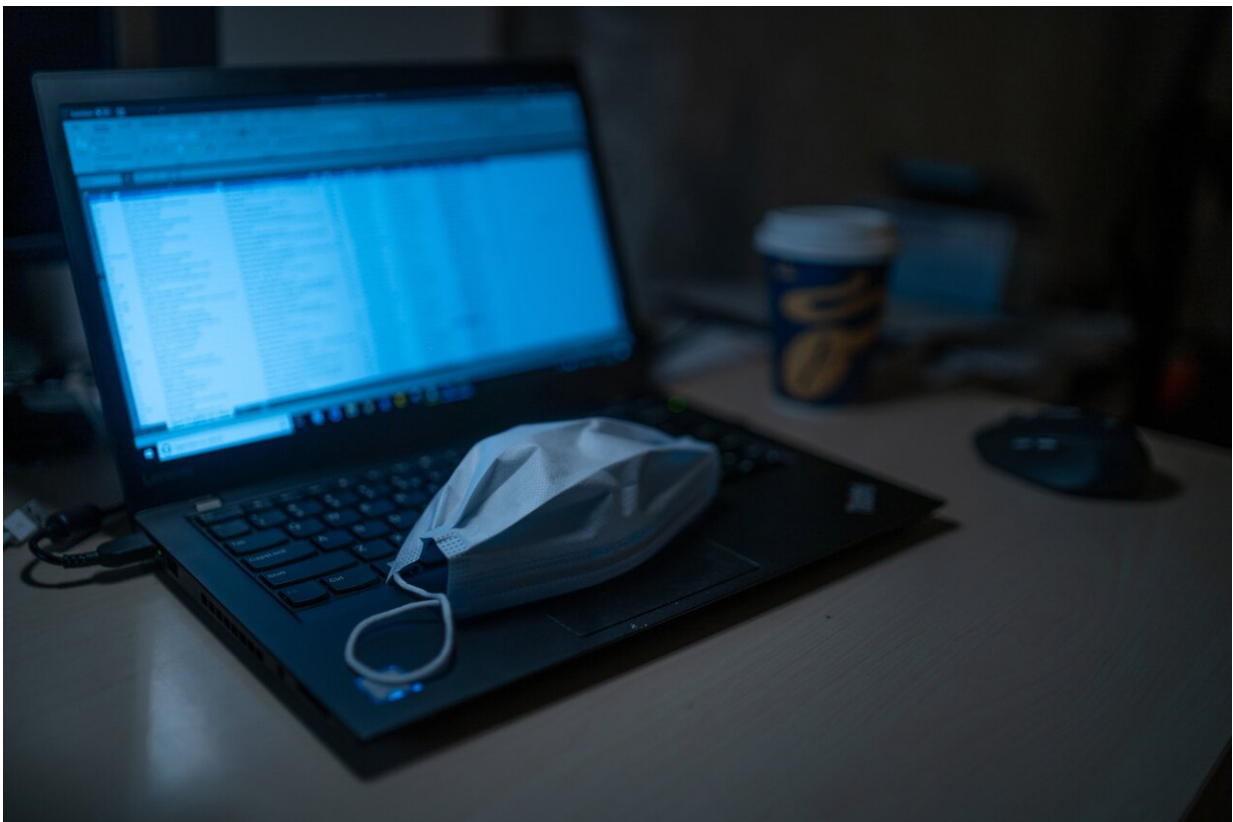


Q&A: Researchers answer questions about their COVID modeling tool and share advice for future forecasters

August 21 2020, by Ben Paul



Credit: Unsplash/CC0 Public Domain

Earlier this Summer we covered the work of Viktor Prasanna, the Charles Lee Powell Chair in the Ming Hsieh Department of Electrical

and Computer Engineering, and USC Viterbi senior research associate Ajitesh Srivastava who works in the Data Science Lab headed by Prasanna. The two, who have extensive experience in epidemic forecasting, had begun using some of their unique models to improve our understanding of how COVID-19 spreads. Now, two months later, they have added even more functions to their model and have begun working closely with the CDC on new strategies as well.

Below, Srivastava answers some questions about their work, the importance and challenges of forecasting, and their planned next steps for COVID-19 research.

You have been working on epidemic forecasting for years. Why is this such an important part of fighting an epidemic?

Forecasting is a vital aspect of resource management and policy making. When a pandemic spreads quickly, one of the biggest challenges is the stress it puts on unprepared healthcare systems. Death rates go up even more because hospitals are overwhelmed with patients and unable to provide proper care. This is what happened in New York earlier this year. The virus was more fatal there than it is now in California, even though California today has well over a half a million cases.

By looking at places hit earlier, scientists use forecasting tools to project the severity in other regions and prepare before the worst hits. And because decisions like lockdowns affect the economy, policy makers have to consider the effect on public health, the healthcare system, and the economy.

Your forecasting tool stands out for being especially fast and adaptable. Since its launch earlier this year,

what new parameters have been added to improve its ability even more?

We have added county-level forecasts which are now publicly available. Apart from providing forecasts of positive cases, we are also forecasting deaths since June. We have also made some adjustments to the algorithm itself, which allows us to make forecasts even faster. Now we can train and forecast for more than 3000 US counties in under 25 seconds. Fast forecasting is important to enable running the algorithms on different "what-if" scenarios. Like, "How many infections will we see in three months, if the current trend continues" or "What if we implement a lockdown for 1 month and reopen partially for the next 2 months."

And we don't even need any expensive or state-of-the-art equipment. I am running these forecasts on an old recycled desktop I have at home!

You are now sharing your forecasts of every US county and state with the CDC on a weekly basis. Have your forecasts directly resulted in changed policy or behavior?

We shouldn't claim the credit alone. We are a part of a "forecast hub" that includes models from 15-20 other teams. Based on aggregating the forecasts from all the teams, CDC generates reports describing where we may be heading in the near future and which states may be expected to show severe breakouts. In mid-June, forecasts were showing a rapid rise in cases as many states were partially open. In July, many states had gone back to more strict distancing rules. We can't say for sure that those decisions were made directly because of our forecasts, but there is a correlation!

You also attend regular meetings with CDC officials and other forecasters across the country. Based on those talks, can you share some insights into some next steps?

The meetings with CDC have definitely driven some of our forecasting decisions. We were reporting forecasts for every day in the future but switched to reporting weekly based on their suggestion. We realized it is difficult to evaluate and make sensible conclusions on day-level forecasts. Another influence from the CDC was to generate county-level forecasts. While we had planned on county-level forecasts for a while, we started generating them at the request of the CDC. You may have heard of some large-scale vaccine trials being planned. The county-level forecasts submitted by the forecast hub are directing where these trials should be performed.

We also learned that it will be helpful to have a more standardized system of data reporting across the country and, if possible, across the world. Non-standardized data reporting has made it hard to make forecasts and comparisons across different regions. As researchers, we need to help find ways to address this problem better.

You've been steadily adding new abilities to your forecasting tool for some time. What do you expect to add next?

Expanding on what-if scenarios is an update you will see soon. We can look into the past using our model to measure what lockdown and distancing measures have been the most effective. Of course, we can do the same to measure what regional strategies were the least effective as well. We can then [forecast](#) based on these best and worst-case scenarios.

We are already generating such scenarios in our public repository. You may see this feature on the interactive webpage later this month. Eventually, we will be adding more scenarios that the users can explore online. Forecasts on hospitalization are also under consideration.

We have also been working on generating forecasts for neighborhoods in Los Angeles, which we hope to make publicly available on our interactive page this month. Finally, we are looking into how different types of businesses may affect the spread of the virus.

What advice, technical and/or strategic, would you give to engineers working on future epidemic forecasting models?

Rely on your understanding of the nuances of the situation, but also keep it simple. The media likes to make big headlines about "AI solves" a problem or "AI failed" at something. The reality is that there may be a billion possible ways to solve a problem with AI/Machine Learning but perhaps only a handful of them are good. It's up to human scientists to make the right decision and end up with one of the "good" solutions.

When it comes to epidemic forecasting specifically, it is especially important to keep it simple.

Yes, by making models more and more complex by introducing more unknowns, we can fit any data. But these models don't perform well in long-term forecasting due to what we call "overfitting." In a recent paper we are sharing at the upcoming KDD conference, we showed how overly complex models can fit a lot of data but never actually learn true values and therefore never produce correct forecasts.

Simplicity is also critical in making the algorithm run faster. Of course,

forecasting needs to be fast for exploring many scenarios for thousands of counties. But speed is just as important for testing and debugging. Imagine if instead of 25 seconds, it took 25 hours to get the results. Then, if you realize the results don't make sense, you need to rerun everything!

Provided by University of Southern California

Citation: Q&A: Researchers answer questions about their COVID modeling tool and share advice for future forecasters (2020, August 21) retrieved 23 April 2024 from <https://medicalxpress.com/news/2020-08-qa-covid-tool-advice-future.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.