COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Ajitesh Srivastava (University of Southern California) and Viktor K Prasanna (University of Southern California), September 2020

Viktor K. Prasanna CIC Database Profile

Title: *Accurate Predictions and Resource Allocation for COVID-19 Epidemic Response*

NSF Award #: 2027524

YouTube Recording with Slides

September 2020 CIC Webinar Information

Transcript Editor: Brian Buckley

Transcript

Katie Naum:
That brings us to our final speaker Ajitesh Srivastava of the University of California. Ajitesh, you can share whenever you're ready.

Ajitesh Srivastava:

Slide 1

Alright, thank you. I hope you can see my screen. Okay. My name is Ajitesh Srivastava. I am the Co-PI of this Rapid award called *Recover: Accurate Fine-grained Predictions and Resource Allocation for COVID-19 Epidemic Response*.

Slide 2

The goals of this project are the following: accurate COVID-19 forecasts at various levels, including for various countries, for state level, country level and fine-grained forecasts like neighborhood level. We'd like to incorporate data-driven identification of a number of unreported cases which are not really observed, but they significantly affect long-term forecasts, and finally, how to utilize these forecasts effectively for resource management during the Pandemic.

This is what we have achieved so far. We have an accurate COVID-19 forecasts. In a few seconds, we can perform training and forecasting for 3,000 counties in less than 30 seconds. We provide reliability guarantees on unobserved factors. For example, how much underreporting is going on. In some cases, we can identify that number and we can mathematically guarantee that that number is close to the truth.

We have a publicly available web-interface and Github repo and we also provide weekly evaluations comparing our approach against models currently being used by the CDC. We are providing our case and death forecasts to the CDC, which includes weekly U.S. national level, state level and county level. These are to be used to inform the public and to be used for vaccination trial site selections.

We are also providing our forecasts to develop ensemble models to UMass forecasting hub and to KIT, where we are providing Germany national and state level forecasts. Here are the publications: two of them are pre-prints and one of them has been accepted at the KDD conference.

Slide 3

So, the central idea that we claim here is that a model is not enough. Just having a model that is complex is not enough. Forecasting requires making decisions regarding what learning strategy you use, what kind of data pre-processing you use, what you choose as parameters, what you decide to have as hyper-parameters. All these decisions can significantly affect your forecast outcomes.

With that in mind, we would like to have a model that captures various complexities and yet, we prefer simple learning approaches to avoid overfitting so that we know what we are learning is actually close to the truth, and that and we want to mathematically check learnability - whether what we are claiming we have – did that just fit the past data? Or did it actually learn something? And we would like to have fast execution to enable scenario analysis. You can read this paper to understand how we start from a complex model and simplify it into a mathematically equivalent model which is easy to learn.

Slide 4

So one central aspect of our work is to address learnability. Despite simplifications, one equation in our model is still non-linear. So we can always fit your data to any model, but does it reflect the truth? Because there may be multiple solutions. For example, here for under-reporting factor, you see four curves. All these four models in a way, they fit the past data very well, but they lead to significantly different outcomes. So how do you know which of these four curves reflects the truth? So, we prove that this under-reporting factor can be learned reliably from data only under certain conditions and the details of this are in the paper.

Slide 5

We have an online visualization where you can interactively perform forecasts and you can also look at various scenarios of where we are going and what would happen if we put our best effort vs. what would happen if we put our worst effort where we might be in a few weeks from now.

Slide 6

On our page, we also provide comparisons of our approach against other approaches by the CDC and we have consistently been in the top two, top three among about 25 models that are being used currently.

Slide 7

Other relevant work we are considering is reliably learning the effects of various policies. We are also interested in performing, solving some resource allocation problems now that we have accurate forecasts. So, how do we allocate PPEs or how do we decide where the testing should happen across the map and where we should have vaccination sites?

And with that, I'll end my talk. Thank you.