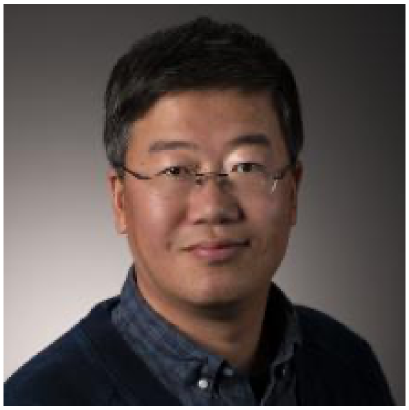
COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Xifeng Yan, (University of California Santa Barbara), November 13, 2020



Title: *Interventional COVID-19 Response Forecasting in Local Communities Using Neural Domain Adaptation Models*

Xifeng Yan CIC Database Profile

NSF Award #: 2029626

Youtube Recording with Slides

November 2020 CIC Webinar Information

Transcript Editor: Rhyley Vaughan

Transcript

Xifeng Yan:

Slide 1

Thank you very much for organizing this webinar. It is my great pleasure to report our recent progress. That work is funded by the NSF RAPID program. I'm Xifeng Yan from UCSB, and the goal of this project is to forecast the daily new cases, hospitalizations, and deaths across the United States for COVID -19 and disease. This work was collaborated with my Ph.D. student, Xiaoyong Jin and Professor Yu-Xiang Wang. Both of them are from UCSB.

Slide 2

As we can see, the number of daily new cases recently increased. It is concerning us.

Slide 3

Furthermore, the number of daily new deaths increased dramatically across the world. It sends us an alert. And our goal is to try to do the forecasting of those daily cases and tests. We believe this kind of forecasting would be very useful for our local decision makers, for hospitals, and for our governments.

Slide 4

There are many factors that are related to the number of COVID-19 cases and disease. There are so many different factors. For example, there are demographic factors, population density in a county, in a state, in a city, local business structures, sociocultural and psychological factors, and the interventional policies. All of those factors have a kind of influence on the contact ratio among people, then on the infection ratio among the people. These two numbers are going to impact the number of hospitalizations, the number of ICU beds used for COVID-19 patients, and in the end, they are going to influence the number of deaths. Furthermore, all of these numbers are influenced by the disease dynamics. The COVID-19 disease evolved dynamically. And all of these numbers can be represented as time series data, the infection ratio, hospitalization, ICU basis. And as we know, time series analysis has been studied for decades. There are many wonderful time series forecasting algorithms.

Slide 5

However, in this work we would like to do something different. We would like to leverage the newest deep-learning models and try to build a pure data-driven approach without assuming any kind of academic models. Our intuition is like [inaudible]. History repeats itself. Different regions share the COVID-19 and trending patterns as the spreading rate is usually determined by common factors such as social interactions, protections, and intervention policies. For example, if you want to forecast the situation of the United States, you can refer to Germany, France, Italy and then learn a lot. In order to forecast the cases in a certain region. We actually can refer to these other regions where the pandemic starts much earlier and just learn from their experience. This is the intuition behind our model.

Slide 6

Let me use one slide to very briefly introduce this intuition. Actually, this is called the attention mechanism in deep learning. Now there's a very successful model in natural language processing and modeling those sequence patterns. Suppose we would like to forecast new cases in California. What we can do is we try to refer to the similar situations in other states across the United States, and we regard these states as reference states.

Slide 7

Then, we can compare the current window in California and compare with historical windows in different states, trying to find those historical windows that share the similar trend. Then, we can use a follow-up trend to forecast the future cases in California. Of course you need to do a kind of weighting, so this is where the deep learning takes place. Then, we can just do this work again. That's it. You move the window and check the similarity among this window and weight different similarities. Then, use the follow-up chance to forecast the future cases in California. You can do it for the daily new cases. You can do it for hospitalizations. You can do it for the number of deaths. It turns out that this simple model purely built on the data actually can achieve top performance without using any kind of SIR/SEIR model. As you know, these models are leading epidemic models, which can help us to understand the disease spreading patterns. It is a little bit surprising to us. We can achieve quite good forecasting results, at least compared with those models, without understanding the underlying and disease spreading patterns.

Slide 8

This slide gives you a kind of summarization and results delivered by leading groups across the United States. These groups send those forecasting results to the CDC, and the CDC archives those results so that later we can compare the accuracy of those algorithms. As you can see, our algorithm, called ACTS, across time series forecasting, actually can achieve quite a good performance. These numbers actually measured using WAP as weighted absolute percentage errors. Using the same model, we can do the new cases forecasting, hospitalization forecasting, and test forecasting. We do the forecasting at the state level, but if there's a need, we actually can do it at the country level if the data is available. This is quite a surprising result, and currently, we are communicating with DC Action. Next Tuesday, there is a meeting with the CDC, and we are going to present this result to a group of experts working in this domain. We are quite new to this area, and we are not experts in time series analysis.

Our result is available in the following web link. We just submitted a paper in archive.org so that it can be shared with researchers much earlier, and we also publish our forecasting results on this website. The result has been sent to the CDC. This is it for our work, and thank you for your time.