Study shows how machine learning could improve COVID-19 predictive models

December 1 2021

Credit: CC0 Public Domain

Throughout the COVID-19 pandemic, case rates have ebbed and flowed in ways that have been hard for epidemiological models to predict. A

new study by mathematicians from Brown University uses an advanced machine learning technique to explore the strengths and weaknesses of commonly used models, and suggests ways of making them more predictive.

"There's an old saying in the modeling field that 'all models are wrong, but some are useful,'" said George Karniadakis, a professor of applied mathematics and engineering at Brown, and a senior author of the research published in *Nature Computational Science*. "What we show here is that the major COVID-19 models were wrong and also not very useful—at least in terms of predicting the course of the pandemic. There was a lot of Monday-morning quarterbacking, but not a lot of accurate predictions."

To find out why that was, the team looked at nine prominent COVID-19 models, all of which were some variation of the "susceptible-infectious-removed" or SIR model. These models divide a population into separate bins: Those who have not yet been infected (susceptible), those who are infected and could spread the virus to others (infectious), and those who have had the infection and can no longer spread it (removed). More complicated versions of the SIR model include additional bins that capture rates of quarantine, hospitalization, deaths and other quantities that could influence the spread of the virus.

There are a number of factors that affect the movement of individuals from one bin to another. Movement from "susceptible" to "infectious," for example, depends how efficiently the virus jumps from person to person along with how often people come in close contact with each other. Many of these factors can't be observed directly, and so the models must infer their values from available data. In modeling terms, these factors are known as parameters.

The study found that a major downfall of COVID-19 models was that

they treated key parameter values as being fixed over time, despite the fact that these factors shifted dramatically in the real world. For example, the community transmission rate of the virus varied widely depending upon mask use, business closings and re-openings, and other measures. Hospitalization rates changed over time as the availability of hospital beds shifted. And the death rate changed with new treatments. All of these evolving factors changed the trajectory of case rates and deaths, but prominent models held these parameters steady in time, which led to poor predictions, the researchers found.

The next question was whether there might be a way to capture these changing parameters in epidemiological models. To do that, the team used physics-informed neural networks (PINNs)—a machine learning technique developed at Brown by Karniadakis and his colleagues. PINNs are neural networks similar to those used to recognize images or transcribe speech to text. But unlike standard neural networks, PINNs are equipped with equations describing the physical laws that govern a system. Karniadakis and his team first used PINNs to discover velocities and pressures of fluid flows from images and videos. In those cases, PINNs were equipped with equations used in fluid dynamics. In this case, the team equipped the PINNs with equations used to calculate how pathogens spread.

"Considering the fact that pandemics evolve in time and there is continuous collection of data, PINNs can be retrained as new data is collected and update the models over time with inferred parameters," said Ehsan Kharazmi, a visiting scholar at Brown and study's co-lead author. "The computational time needed for re-training PINNs with new data is relatively short compared to the time-scale of pandemic evolution."

The team fed the PINN-equipped models real-world data—taken from New York City, the states of Rhode Island and Michigan, and national

data from Italy—and allowed the PINNs to infer values for key parameters over time. The PINNs were also able to quantify their uncertainty about the inferred parameters. Then the team used the PINN-informed models to make predictions about the future. In January 2021, the team made predictions for the next six months based on the time-adjusted parameters. Then, in comparing actual case rates to what they predicted, they found that the actual case rates from January through June 2021 fell within the uncertainty window predicted by the models. That was true for each of the four datasets used in the study.

The findings suggest that while no model can accurately capture all the dynamics that play out during an extended pandemic, models with the ability to adjust key parameters on the fly could make for more useful predictions.

"The inferred models using PINNs can be used to assess possible future trajectories by tweaking the model parameters," Kharazmi said. "This can provide some insights for making or adjusting policies."

Provided by Brown University