COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Judy Fox (University of Virginia) January 30,

Title: Expeditions: Collaborative Research: Global Pervasive Computational Epidemiology

Judy Fox CIC Database Profile

NSF Award #: 2151597

YouTube Recording with Slides

Winter 2024 CIC Webinar Information

Transcript Editor: Lauren Close

Transcript

Slide 1

Thank you, Lauren, for your kind introduction, and thanks to the 2024 COVID Information Commons hosts and our fellow speakers, students, and staff. I'm going today to report on our NSF funded Expeditions project. This is a collaborative research for global pervasive computational epidemiology. I'm Judy Fox. I'm working in the data science school and in computer science at the University of Virginia.

Slide 2

A little bit of background of this project - it is a multi-institutional project led by Dr. Madhav Marathe and he is the coordinator PI. We have colleagues within the University of Virginia and Biocomplexity Institute. This is a multi-institutional collaboration and we have wonderful colleagues and researchers. It is quite an inspiring experience for me.

Slide 3

I wanted to talk about the future. That is the exciting part about my presentation today. Imagine COVID in 2025. What will we be? We want to move from intervention to prevention because infectious disease is such a societal problem. By 2050, there will be projected deaths every year of over 10 million and the economic impact will be over a hundred trillion dollars. Just a few years ago, we came out of this pandemic, in which the U.S. alone had over 1.1 million people dead and there are over 100 million cases. This is almost one-third of the families who were infected. Much of this could have been prevented through better informed government policy. However, COVID-19 is a complex data problem. First, we have non-stationary data coming in. It's quite difficult to learn and predict the trends with a lack of data, noisy data. Making the infection forecasting understandable or explainable can also help the decision making. They can help to identify the important geographical and temporal focus areas so that we can signal the governments to make a more effective allocation of resources to prevent the disease's spread.

Slide 4

I want to focus the rest of the talk about the research conducted in my group. We wanted to interpret county level COVID-19 infections in the U.S. We applied the Transformer AI model, which is a type of deep learning model used by the large language models.

Slide 5

One of the areas we're focusing on is to ask the question: why do we need prediction? Predictions using real time data have been highlighted in 2009 by Dr. Harvey V. Fineberg and Dr. Mary Elizabeth Wilson. They draw out the important aspect of why it's important - we want to study using the latest data - to study disease control and try to observe and predict. Interventions will be the action operate at the peak point, but intervention will flatten the curve ahead of the time.

Slide 6

We use the deep learning model the Temporal Fusion Transformer (TFT) and this model can forecast in real time. In our experiments, we use the past 13 days to predict the future 15 days. Data collection comes from different prediction models. We categorize them into static co-variant and dynamic datasets, like cases and deaths. We also have some known inputs such as upcoming Christmas and New Year holidays.

Slide 7

With such a model, our goal is trying to understand how to use Interpretable AI to get the knowledge and information to where and when the infection will happen. Which countries are at the most risk? Who are the vulnerable communities? And we try to help. The journey of this study is all of this with a lot of road blocks we need to overcome.

Slide 8

The problem comes from that the general issue with the accuracy. How can our model predict so we are confident our prediction is accurate? Also, when we have this prediction, how do we explain to policy makers what are the important factors for the current rise of the cases? That requires us to have a deeper understanding of the data itself at a much finer grade level, like the county level characteristics. We also want to make a decision in real time so that its relevance is guaranteed.

Slide 9

It's a very rich study because health disparity is not related just to people but also a socio-economic impact. We've collected over two and a half years of data for 3,142 U.S. counties. We categorize and reduce the death set features from twenty to six features. Two are static features: the health disparity and the population age group. There are observable features, including the vaccination, the disease spread, transmissible cases, and mobility. We incorporate into known and unknown events so that we have a very complex multimodal Articulated AI model.

Slide 10

Let me switch gears to show you some of our prediction results. We compared the TFT model with the LSTM, a basic sequence to sequence model. We can show that in the chart on the left hand side the TFT model performs the best. It gives the error message and the accuracy is higher and the error is lower.

Slide 11

What is underpinning the understanding of Interpretable AI? This comes from the tension mechanism using an encoder decoder architecture and the tension mechanism is the underpin of the current large language model, including ChatGPT. Because we have this multi-attention mechanism we can capture the context of the disease with the times going by and so that will refine the space where we look at the feature - cause and effect. So we can focus the importance of the spatial and the temporal patterns with the hotspot areas. On the right hand side, you can see the architecture of a TFT model. We input the past features and try to predict the future happening. In this case, it is the cases of infection which encapsulate embedding the features, especially the static features, into using sequences model to capture the dependable time patterns. We propagate all these patterns to self-attention to try to mask out interpretable multi-head attention so that we can focus on the important patterns and areas.

Slide 12

This forecasting model is able to provide us with the cyclic pattern which is reporting of the COVID cases. It also accounts for special time events, like holidays, like weekends. We can clearly label them in this chart. On the right hand side we can even look backwards to which period of time has the most impact of the future prediction in the time frame from 0 to 13 days.

Slide 13

Here is the chart of trends we can predict. Choosing the top 100 largest counties, you can see that we have the prediction compared with the ground truth. Further, we can compare the smaller population counties - you can see the result matches much better and has less spikes and the different values.

Slide 14

How about the location information? This chart shows on the left hand side the Self-attention of the AI model, and captures the intensity at the county level. This is not possible if our data is at the state level. On the right hand side is a data representation of cumulative cases from the CDC of over 3,000 U.S. counties. If you look at these two results, you can see the correlation between these two datasets and results.

Slide 15

We measured the correlation and concluded that we can interpret the AI model by capturing the self-attention weights at county level. There is a strong correlation with the model behavior versus the prediction of the cases compared to the ground truth.

Slide 16

R2 provides the information that policy makers need. We believe a small reduction in the hotspot transmission can lead to a large reduction in infections, especially in the early stage. Real time forecasting and focusing our attention on the most important regions in daily infection is crucial. We have a fine-grained method to capture these infections at the county level that would significantly reduce the risk. There are many future works we can do coupled with this existing result. We can explore many social and economic impacts and disparities for future work.

Slide 17

At UVA, we started the AI for science program and attracted over 3,000 undergraduates to participate. We selected a dozen students involved in our project.

Slide 18

We want to study further the population age group sensitivity using Morris based index study to know from each population group, who are most vulnerable to COVID infections?

Slide 19

We chose a time-series deep learning model because the tension mechanism as aforementioned can give insight and understand how the model is predicted.

Slide 20

There are several lessons we learned from this journey. Infectious disease has been a socio-economic globe-wide problem that impacts public health and the economy at such a large scale. From the COVID-19 pandemic, our first lesson learned is that testing is vital in understanding the pandemic's progress. The second is that a lot of infections are very different in different smaller regions. The situation is quite dynamic. The better way to address this in policy is adapted to the local level. There are many ways we can improve, including that we built the tools like we studied to accurately predict the COVID infection and the future infectious disease. That will help the policy maker to intervene in a science-based manner. Intervention is the future. We want to be prepared and ready for future crises, such as the pandemic.

Slide 21

Before I wrap it up, I really want to share with everyone that we need to build trust. If there are any lessons learned from the past pandemic, we want to improve the explanation to the public, to experts, to the policy makers what is happening and how we can leverage policy to impact. We want to have an interpretable method to quantifiably model and evaluate our methods. We also want to explain our model's behavior and AI-based predictions to non-experts including all the public and students. Collectively, we hope to build a future so that we're ready for the future events.

Slide 22

With that, I want to thank you, everyone, for participating in today's workshop. I provide some of our work here. Please feel free to contact me. I look forward to future discussions at the end of this workshop. Thank you!