COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Lalitha Sankar, (Arizona State University), April 2021

Title: FACT: Federated Analytics based Contact Tracing for COVID-19

Lalitha Sankar CIC Database Profile

NSF Award # : 2031799

YouTube Recording with Slides

April 2021 CIC Webinar Information

Transcript Editor: Lara Azar

Transcript

Lalitha Sankar:

Slide 1

Hi everybody, and thank you Helen and Florence, for giving me this opportunity today. So this is a NSF, grant, a RAPID Grant funded through SATSI with my awesome colleagues, Ni Trieu, Ming Zhao, and Visar Brisha. Ni and Ming are in computer science and Visar is in the College of Health Solutions. As the name suggests, we're going to be doing a lot of deep dive into machine learning, and use that to do a more holistic contact tracing.

Slide 2

So there's actually three aspects to our work, we're trying to enhance existing contact tracing protocols by using both bluetooth and GPS, and I'll dive into it in a moment. The big idea here is to say, can you ride or piggyback on surveys that are being done on campuses across the United States, particularly at ASU. We have a daily health survey we have to take. Can you use that kind of survey data, use it to both build a baseline risk model, and then combine it with mobility and even phonation, which is a biomarker? Just like the nasal thing, it turns out phonation is a biomarker for a whole bunch of respiratory and even mental health issues, and combine that to build hot spot models, and so on and so forth, and continue to evaluate risk and put up a risk score, and our overarching goal in this work is actually to, take this built system and deploy it into ASU's mobile app, and that's an ongoing process.

Slide 3

So very briefly on contact tracing.

Slide 4

Existing contact tracing apps basically use bluetooth, there are these tokens that are used, I don't have time to go into it, but they are vulnerable to security attacks. They also, because, do not, because they do not GPS, they cannot evaluate infected hotspot locations except through, say, cell towers and stuff like that. So they're not really exploiting a lot of rich user specific data, and our goal in this project-

Slide 5

-was to actually combine bluetooth with GPS which exists, but we've been able to build this protocol, test it on android devices, and provide strong privacy guarantees. And the other side effect is we can also now compute a histogram of hotspots using secure aggregation techniques at multiple of the severs the back, [unclear].

Slide 6

Okay so that's on the contact tracing part. On the device part.

Slide 7

Our goal is actually to come up with a risk prediction model and what are we going to do, we wanna have a baseline risk that's based on your health symptoms one time, but daily symptoms will be used to continually assess this risk. Could use mobility patterns, especially on a university campus to figure out if there's an outbreak at any dorm or any place, how can we, you know, even move traffic around, and ultimately our goal is really to do phonation based risk indices and I will not have the time to talk about this in great detail, but the whole point is to breathe into, you know, use our app to just, say "ahh" for, you know, a few seconds, and use that to look at both respiratory, respiratory health issues are directly correlated with how you phonate and its also correlated apparently with brain fog now. So there's a lot of research here, and Visar, my colleague and I are working on this. Okay.

Slide 8

So, I'm going to do a deep dive into only one thing in the interest of time which is, how do you do, how do you use surveys to predict risk ? And we've been very lucky because thanks to this grant and a grant from Google's AI for Social Good, we were able to get, participate in a competition that Facebook hosted and get a data set from Facebook that CMU Delphi has been collecting for them through the Facebook app and basically this is a survey, it's a daily survey and they've had 18 million responses with 53,000 participants and the whole idea is too collect a whole bunch of symptoms based survey, prior health conditions, social distancing, mental health demographics, and of all that there are only about 900,000 that are people who've taken the test. So that's the data we use because that's the label we use to predict how- whether somebody, or you know, based on your symptoms, whether you may be at risk for COVID.

Slide 9

So what are we doing you know, I'm diving a little bit deeper into what types of models we're using. We're gonna use XGBoost because this is the kinda health data set. It's got a mix of, you know,

discrete and continuous value data. But XGBoost is a fantastic, very well known robust algorithm. What we're going to do, in fact, is make it even more robust by using a whole class of loss functions that we've developed and I'm happy to take that offline.

Slide 10

In short, what we've been able to show is that even restricting ourselves to the top eight symptoms, and this is a very imbalanced data set, we have 86% negative only 14% positive labels, of the people who've taken the test. We can enhance, you know, existing XGBoost with even better things using our loss function.

Slide 11

Even more interesting, what we've been able to do is to, in reality, the survey data is very noisy. We did a lot of pre-processing. But can you actually, continually do this while the data is noisy? So we tested our algorithm against noisy labels, where we flipped a bunch of labels we kept it, we did two experiments, stratified and unstratified, we kept the imbalance and we have some very good results.

Slide 12

So in the interest of time, again, I'm gonna stop it there and take it offline, our holy, you know the grail of this project, if I may, is to combine all these three things, put in on the ASU app and be able to give the user a survey risk, and do a lot of back end data collection for ASU.

Slide 13

So we're a little bit away from that but we have our IRB approved so we're going to put an app in and do stuff. So I'm gonna stop right there. Thank you very much.