

[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: Julie Meunier

Transcript

Lalitha Sankar:

Slide 1

Bonjour à tous, et merci Helen et Florence de me donner cette opportunité aujourd'hui. Donc, il s'agit d'une subvention de la NSF, une subvention RAPID financée par SATSI avec mes formidables collègues, Ni Trieu, Ming Zhao et Visar Brisha. Ni et Ming sont en informatique et Visar est dans le College of Health Solutions. Comme son nom l'indique, nous allons plonger profondément dans l'apprentissage automatique et l'utiliser pour une recherche plus approfondie sur la recherche des contacts.

Slide 2

Il y a en fait trois aspects à notre travail. Nous essayons d'améliorer les protocoles existants de recherche des contacts en utilisant à la fois le bluetooth et le GPS, et je vais y plonger dans un instant. L'idée principale est de voir si vous pouvez utiliser des enquêtes menées sur les campus aux États-Unis, en particulier à ASU. Nous avons une enquête de santé quotidienne à remplir. Pouvez-vous utiliser ce type de données d'enquête, l'utiliser pour construire un modèle de risque de base, puis le combiner avec la mobilité et même la phonation, qui est un biomarqueur ? Tout comme le test nasal, il s'avère que la phonation est un biomarqueur pour toute une série de problèmes respiratoires et même de santé mentale, et le combiner pour construire des modèles de points chauds, et ainsi de suite, et continuer à évaluer le risque et afficher un score de risque. Notre objectif principal dans ce travail est en fait de prendre ce système construit et de le déployer dans l'application mobile de ASU, et c'est un processus en cours.

Slide 3

Donc, très brièvement sur le traçage des contacts.

Slide 4

Les applications existantes de traçage des contacts utilisent essentiellement le bluetooth, il y a ces jetons qui sont utilisés, je n'ai pas le temps d'entrer dans les détails, mais ils sont vulnérables aux attaques de sécurité. Ils ne peuvent pas évaluer les emplacements de points chauds infectés sauf par le biais, disons, des tours cellulaires et ainsi de suite, car ils n'utilisent pas le GPS. Ainsi, ils n'exploitent pas vraiment beaucoup de données utilisateur spécifiques riches, et notre objectif dans ce projet-

Slide 5

- était en fait de combiner le bluetooth avec le GPS qui existe, mais nous avons pu construire ce protocole, le tester sur des appareils Android et fournir des garanties de confidentialité solides. Et l'autre effet secondaire est que nous pouvons également maintenant calculer un histogramme des points chauds en utilisant des techniques d'agrégation sécurisée sur plusieurs des serveurs à l'arrière, [non clair].

Slide 6

Très bien, voilà pour la partie traçage des contacts. Sur la partie appareil.

Slide 7

Notre objectif est en fait de créer un modèle de prédiction des risques et ce que nous allons faire, c'est que nous voulons avoir un risque de base basé sur vos symptômes de santé une fois, mais les symptômes quotidiens seront utilisés pour évaluer continuellement ce risque. Nous pourrions utiliser des modèles de mobilité, en particulier sur un campus universitaire, pour savoir s'il y a une épidémie dans une résidence ou à un endroit quelconque, comment pouvons-nous, vous savez, même déplacer le trafic, et en fin de compte, notre objectif est vraiment de créer des indices de risque basés sur la phonation et je n'aurai pas le temps d'en parler en détail, mais l'idée principale est de respirer dans, vous savez, utiliser notre application pour dire "ahh" pendant quelques secondes, et utiliser cela pour examiner à la fois les problèmes de santé respiratoire, les problèmes de santé respiratoire étant directement corrélés à la façon dont vous phonate et cela semble également corrélé à la confusion mentale maintenant. Il y a beaucoup de recherches ici, et Visar, mon collègue et moi travaillons sur cela. Bien.

Slide 8

Donc, je vais approfondir une seule chose dans l'intérêt du temps, c'est-à-dire comment utilisez-vous les enquêtes pour prédire le risque? Et nous avons eu beaucoup de chance grâce à cette subvention et à une subvention de Google AI for Social Good. Nous avons pu participer à une compétition organisée par Facebook et obtenir un ensemble de données de Facebook que CMU Delphi collecte pour eux à travers l'application Facebook. Fondamentalement, il s'agit d'une enquête, c'est une enquête quotidienne et ils ont eu 18 millions de réponses avec 53 000 participants et l'idée principale est de collecter un tas d'enquêtes basées sur les symptômes, les conditions de santé antérieures, la distanciation sociale, la santé mentale, les données démographiques, et de tout cela, il n'y a qu'environ 900 000 personnes qui ont passé le test. C'est donc la donnée que nous utilisons car c'est l'étiquette que nous utilisons pour prédire si, vous savez, en fonction de vos symptômes, vous pourriez être à risque de COVID.

Slide 9

Alors que faisons-nous, vous savez, je plonge un peu plus dans les types de modèles que nous utilisons. Nous allons utiliser XGBoost car c'est le type de jeu de données sur la santé. Il a un mélange de données discrètes et continues. Mais XGBoost est un algorithme fantastique, très bien connu et robuste. Ce que nous allons faire, en fait, c'est le rendre encore plus robuste en utilisant toute une classe de fonctions de perte que nous avons développées et je suis ravi de discuter de cela hors ligne.

Slide 10

En bref, ce que nous avons pu montrer, c'est que même en nous restreignant aux huit principaux symptômes, et c'est un ensemble de données très déséquilibré, nous avons 86 % de négatif et seulement 14 % d'étiquettes positives, des personnes qui ont passé le test. Nous pouvons améliorer, vous savez, XGBoost existant avec des choses encore meilleures en utilisant notre fonction de perte.

Slide 11

Encore plus intéressant, ce que nous avons pu faire, c'est que, en réalité, les données d'enquête sont très bruyantes. Nous avons fait beaucoup de prétraitement. Mais pouvez-vous réellement le faire continuellement alors que les données sont bruyantes ? Nous avons testé notre algorithme avec des étiquettes bruyantes, où nous avons inversé un tas d'étiquettes, nous l'avons gardé, nous avons fait deux expériences, stratifiée et non stratifiée, nous avons maintenu le déséquilibre et nous avons de très bons résultats.

Slide 12

Donc, dans l'intérêt du temps, encore une fois, je vais m'arrêter là et discuter de cela hors ligne. Notre saint Graal dans ce projet, si je puis dire, est de combiner ces trois choses, les mettre dans l'application ASU et pouvoir donner à l'utilisateur un risque d'enquête, et faire beaucoup de collecte de données en coulisses pour l'ASU.

Slide 13

Nous sommes un peu loin de cela, mais nous avons notre approbation du comité d'éthique de la recherche, donc nous allons mettre en place une application et faire des choses. Donc je m'arrête ici. Merci beaucoup.