



Transcripción de una presentación de Lalitha Sankar, (Arizona State University), abril de 2021

Título: Seguimiento de contactos basado en Federated Analytics para COVID-19

[Perfil de Lalitha Sankar en la base de datos de CIC](#)

Subvención de La Fundación Nacional de Ciencias (NSF, por sus siglas en inglés) #: [2031799](#)

[Grabación de YouTube con diapositivas](#)

[Información del seminario web del CIC de Abril 2021](#)

Editora de la Transcripción: Lara Azar

Editora de la Traducción: Isabella Graham Martínez

Transcript

Lalitha Sankar:

Diapositiva 1

Hola a todos, y gracias a Helen y Florence, por darme esta oportunidad hoy. Así que esta es una subvención de la NSF, una subvención de RAPID financiada a través de SATSI con mis increíbles colegas, Ni Trieu, Ming Zhao y Visar Brisha. Ni y Ming están en ciencias de la computación y Visar está en el Colegio de Soluciones de Salud. Como su nombre indica, vamos a hacer una inmersión profunda en el aprendizaje automático, y usaremos eso para hacer un seguimiento de contacto más holístico.

Diapositiva 2

Así que en realidad hay tres aspectos de nuestro trabajo, estamos tratando de mejorar los protocolos de rastreo de contactos existentes mediante el uso de bluetooth y GPS, y voy a bucear en un momento. La gran idea aquí es decir, se puede montar o ir a cuestras en las encuestas que se están haciendo en los campus en los Estados Unidos, en particular en ASU. Tenemos que hacer una encuesta de salud diaria. ¿Se puede utilizar ese tipo de datos de la encuesta, utilizarlo para construir un modelo de riesgo de línea de base, y luego combinarlo con la movilidad e incluso la fonación, que es un biomarcador? Al igual que la cosa nasal, resulta que la fonación es un biomarcador para un montón de problemas respiratorios e incluso mentales, y combinar eso para construir modelos de puntos calientes, y así sucesivamente, y seguir evaluando el riesgo y poner una puntuación de riesgo, y nuestro objetivo general en este trabajo es, tomar este sistema integrado e implementarlo en la aplicación móvil de ASU, y ese es un proceso continuo.

Diapositiva 3

Muy brevemente sobre el rastreo de contactos.

Diapositiva 4

Las aplicaciones de rastreo de contactos existentes básicamente usan bluetooth, hay estos tokens que se usan, no tengo tiempo para analizarlos, pero son vulnerables a los ataques de seguridad. También, porque no, porque no tienen GPS, no pueden evaluar ubicaciones de puntos de acceso infectados excepto a través de, por ejemplo, torres de telefonía celular y cosas por el estilo. Así que no están explotando una gran cantidad de datos específicos de usuarios ricos, y nuestro objetivo en este proyecto-

Diapositiva 5

-era en realidad combinar bluetooth con GPS que existe, pero hemos sido capaces de construir este protocolo, probarlo en dispositivos Android, y proporcionar fuertes garantías de privacidad. Y el otro efecto secundario es que ahora también podemos calcular un histograma de puntos de acceso utilizando técnicas de agregación segura en múltiples de los cortes de la espalda, [confuso].

Diapositiva 6

Bien, eso está en la parte de rastreo de contactos. En la parte del dispositivo.

Diapositiva 7

Nuestro objetivo es realmente llegar a un modelo de predicción de riesgo y lo que vamos a hacer, queremos tener un riesgo de línea base que se basa en sus síntomas de salud una vez, pero los síntomas diarios se utilizarán para evaluar continuamente este riesgo. Podría usar patrones de movilidad, especialmente en un campus universitario para averiguar si hay un brote en cualquier dormitorio o en cualquier lugar, ¿cómo podemos, ya sabes, incluso mover el tráfico, y, en última instancia, nuestro objetivo es realmente hacer índices de riesgo basados en la fonación y no tendré tiempo para hablar de esto con gran detalle, pero el punto es respirar, ya sabes, usar nuestra aplicación para solo, decir "ah" por, ya sabes, unos segundos, y usar eso para mirar ambos problemas respiratorios, salud respiratoria se correlacionan directamente con la forma en que se fonética y su también se correlaciona al parecer con la niebla cerebral ahora. Así que hay mucha investigación aquí, y Visar, mi colega y yo estamos trabajando en esto. Vale.

Diapositiva 8

Por lo tanto, voy a enfocarme en solo una cosa en el interés del tiempo que es, ¿cómo lo haces, cómo se utilizan las encuestas para predecir el riesgo? Y hemos tenido mucha suerte porque gracias a esta subvención y una subvención de la IA de Google para el Bien Social, Fuimos capaces de conseguir, participar en un concurso que Facebook alojado y obtener un conjunto de datos de Facebook que CMU Delphi ha estado recogiendo para ellos a través de la aplicación de Facebook y, básicamente, esta es una encuesta, es una encuesta diaria y han tenido 18 millones de respuestas con 53,000 participantes y toda la idea es demasiado recoger un montón de encuesta basada en síntomas, condiciones de salud anteriores, el distanciamiento social, demografía de salud mental, y de todo lo que hay solo unos 900.000 que son las personas que han tomado la prueba. Esos son los

datos que usamos porque esa es la etiqueta que usamos para predecir cómo- si alguien, o usted sabe, basado en sus síntomas, si usted puede estar en riesgo de COVID.

Diapositiva 9

Así que lo que estamos haciendo ya sabes, estoy buceando un poco más profundo en lo que los tipos de modelos que estamos utilizando. Vamos a usar XGBoost porque este es el conjunto de datos de salud. Tiene una mezcla de datos de valor discretos y continuos. Pero XGBoost es un algoritmo fantástico y robusto muy conocido. Lo que vamos a hacer, de hecho, es hacerlo aún más robusto mediante el uso de toda una clase de funciones de pérdida que hemos desarrollado y estoy feliz de tomar que fuera de línea.

Diapositiva 10

En resumen, lo que hemos podido demostrar es que incluso restringiéndonos a los ocho síntomas principales, y este es un conjunto de datos muy desequilibrado, tenemos un 86% negativo, solo un 14% positivo, de las personas que han hecho la prueba. Podemos mejorar, ya sabes, XGBoost existente con cosas aún mejores usando nuestra función de pérdida.

Diapositiva 11

Aún más interesante, lo que hemos sido capaces de hacer es, en realidad, los datos de la encuesta son muy ruidosos. Hicimos un montón de pre-procesamiento. Pero, ¿puedes hacer esto continuamente mientras los datos son ruidosos? Así que probamos nuestro algoritmo contra etiquetas ruidosas, donde volteamos un montón de etiquetas que guardamos, hicimos dos experimentos, estratificados y no estratificados, mantuvimos el desequilibrio y tenemos algunos resultados muy buenos.

Diapositiva 12

Así que en el interés del tiempo, de nuevo, voy a detenerlo allí y sacarlo de línea, nuestro santo, ya sabes el grial de este proyecto, si puedo, es combinar todas estas tres cosas, poner en la aplicación ASU y ser capaz de dar al usuario un riesgo de encuesta, y hacer una gran cantidad de recogida de datos de back-end para ASU.

Diapositiva 13

Así que estamos un poco lejos de eso, pero tenemos nuestro IRB aprobado, así que vamos a poner una aplicación y hacer cosas. Así que voy a parar ahí. Muchas gracias a todos.