Centro de Información de COVID (CIC): Charlas científicas de relámpago

Transcripción de una presentación de Murat Kantarcioglu (Universidad de Texas en Dallas), 16 de octubre de 2020

Título: RAPID: Colaboración: Un marco de evaluación del riesgo de privacidad para el intercambio de datos a nivel de persona durante pandemias

Perfil de Murat Kantarcioglu en la base de datos de CIC

Subvención de La Fundación Nacional de Ciencias (NSF, por sus siglas en inglés) #: 2029661

Grabación de YouTube con diapositivas

Información del seminario web del CIC de octubre 2020

Editora de la Transcripción: Macy Moujabber

Editora de la Translación: Isabella Graham Martínez

Transcripción

Diapositiva 1

Creo que ahora es una buena tarde para algunos. Me llamo Murat Kantarcioglu. Soy profesor de informática en la Universidad de Texas en Dallas. Este es nuestro trabajo RÁPIDO sobre la evaluación de riesgos de privacidad para compartir datos durante pandemias con el profesor Brad Malin de la Facultad de Medicina de Vanderbilt.

Diapositiva 2

Así que como muchos de ustedes siguen las noticias, a medida que el COVID-19 progresa hay nuevas preguntas que vienen con respecto a si está relacionado con la raza, si fumar, por ejemplo, aumenta su riesgo con COVID-19, malos resultados, etc. Por lo tanto, con el fin de responder a este tipo de preguntas si es relación de la raza y la mortalidad COVID-19 u otros factores como fumar alta, la presión arterial. Claramente, necesitamos recopilar más datos y, por supuesto, estos datos necesitan ser analizados.

Diapositiva 3

Pero desafortunadamente, el- o afortunadamente en algún sentido, el secreto y los requisitos de privacidad pueden retrasar este intento Si nos fijamos en este artículo de Ciencia publicado en julio, veríamos que California no quería compartir- Los funcionarios estatales de California no querían compartir la COVID-19 detalles de datos con algunos investigadores, obstaculizando cierta investigación.

Diapositiva 4

Por lo tanto, en un sentido tienen razón porque se sabe que los datos compartidos se pueden volver a identificar fácilmente. Se ha demostrado en los últimos años que si combinas diferentes conjuntos de datos, por ejemplo, si tienes un COVID-19 que informa código postal, edad y género, hay un mayor riesgo de vincular esto con las listas de registro de agua e identificar a las personas que están en el COVID-19 datos de pacientes. Y, por supuesto, esto se conoce desde hace tiempo, por lo que no es un problema nuevo y la comprensión de estos riesgos ha sido una investigación activa en los últimos veinte años. Pero aún así, COVID-19 cambió muchas cosas.

Diapositiva 5

Una cosa es que ahora tenemos, especialmente con las redes sociales, tenemos diferentes fuentes de re-identificación. Este es un ejemplo de The New York Times, donde se hizo un perfil de muchos pacientes que fallecieron debido a COVID-19 en sus historias de vida.

Diapositiva 6

Por lo tanto, si vuelve a este escenario de ataque, ahora puede tener una fuente adicional donde sabe que algunas personas han fallecido y su información y puede usarla para eliminar los objetivos y volver a ejecutar los ataques de identificación.

Diapositiva 7

Por lo tanto, este RAPID realmente está tratando de entender los desafíos únicos debido al intercambio de datos pandémicos, especialmente los riesgos de re-identificación. Por lo tanto, una cosa que nos damos cuenta es que a diferencia del trabajo pasado, los atributos que se necesitan para ser compartidos, como fumar o la raza, pueden estar cambiando con el tiempo. Por lo tanto, tenemos que entender realmente eso. Además, los números de los casos están cambiando, lo que significa que algunos riesgos de privacidad pueden estar cambiando con el tiempo, por lo que realmente necesitamos actualizar los riesgos de privacidad con el tiempo a medida que se comparten nuevas fuentes de datos con otros. Por lo tanto, este comportamiento cambiante e incierto del intercambio de datos pandémicos afecta los marcos de evaluación de riesgos, y por esa razón ahora estamos desarrollando y actualizando los modelos existentes.

Diapositiva 8

Por lo tanto, nuestro objetivo es realmente buscar en lo que es la mejor utilidad que podemos lograr dados los objetivos de privacidad de datos, que en este caso para asegurarse de que el riesgo de re-identificación es baja en otras palabras, es muy difícil volver a identificar a las personas. Y también, ¿cuál es la mejor privacidad posible que podemos proporcionar dados los objetivos de utilidad de datos? Y de nuevo, nuestro objetivo es dar la información más precisa posible. La segunda opción entra en juego porque existe el argumento de que deberíamos centrarnos realmente en combatir la enfermedad para que la privacidad de los datos sea secundaria. Aún así, nuestra investigación pasada mostró que, dados sus objetivos con respecto al intercambio de datos, diferentes alternativas pueden tener diferentes resultados de privacidad. Entonces, entender eso sería importante. Por supuesto, la dirección principal sería: cuál es la utilidad de datos y cómo definirla y esa es nuestra investigación en curso.

Diapositiva 9

Para concluir, construimos nuestro propio trabajo pasado en este análisis del riesgo de identificación y actualizamos y desarrollamos técnicas basadas en este desafío único de la información cambiante y cambiante debido a la pandemia. Y pronto, con suerte, como dentro de un mes más o menos, tendríamos un documento sobre políticas generales de intercambio de datos para guiar a los funcionarios de salud sobre qué datos se pueden compartir con bajo riesgo, cuáles son las opciones de mayor riesgo y su impacto. Así que por favor síganos si tiene alguna preocupación sobre el intercambio de datos y la privacidad de los datos pandémicos. Así que, me detendré aquí supongo que un poco con el tiempo. Gracias.