

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

Transcripción de una presentación de Hong Qin (University of Tennessee, Chattanooga), 26 de julio de 2023.



Título: [Desarrollar y evaluar Marcos Computacionales para Predecir y Prevenir Futuras Pandemias de Coronavirus.](#)

[Perfil de la base de datos CIC de Hong Qin](#)

Premio de NSF#: [2200138](#)

[Grabación de YouTube con diapositivas](#)

[Verano 2023 CIC Información de webinario](#)

Transcripción Editada: [Karem Coca](#)

Transcripción:

Hoja 1

Hola, me llamo Hong Qin y soy profesor en la Universidad de Tennessee, Chattanooga. Esta charla trata de un proyecto en el que he tenido la suerte de trabajar con un gran equipo de colegas de North Carolina A&T, Spelman College y Catholic University of America. Estos son los colaboradores sobre el papel, pero en realidad son muchos más. En nuestra colaboración hemos tenido probablemente más de una docena de miembros del equipo que han estado trabajando en esto.

Se trata de un proyecto financiado por la National Science Foundation. Se llama PIPP Fase Uno. Mi objetivo es desarrollar un marco basado en IA para predecir y prevenir futuras pandemias de coronavirus. Aunque digamos que se trata de coronavirus, el modelo tiene ahora la posibilidad de generalizarse a presumiblemente muchas otras pandemias víricas

Hoja 2

Esta es la... vale, se está reenviando automáticamente... mis disculpas. Bien, este es el resumen que propusimos al principio, aunque ahora hemos hecho muchas modificaciones. La idea sigue siendo la misma. Proponemos cómo predecir el virus - una nueva pandemia de virus? Nuestra idea es que primero generemos todos los posibles SARS-CoV-X a partir de los que son predecibles basándonos en patrones de recombinación o cambio de hábitat. A continuación, predecimos esas posibles reacciones del SARS-CoV-X. Este es el reto: es fácil generar esas secuencias, pero ¿cómo sabemos cuál de ellas se volverá virulenta o se transmitirá en una población humana? Esa es realmente la clave. Así que recurrimos a la IA. Cuando digo que la IA es una caja negra, bueno, en este caso eso es probablemente también una bendición porque

realmente no sabemos cómo un virus salta de un huésped, de un animal, a un humano para volverse virulento. Así que la IA es una caja negra, pero probablemente sea una gran herramienta que podamos utilizar aunque sea una caja negra. Ese es mi argumento.

Y mis disculpas, tuve que tachar ciertas cosas porque mi Universidad está solicitando una patente basada en la función de esto. Así que el objetivo es aquí tenemos secuencias de entrada - esto es sólo secuencias de nucleótidos primarios. A partir de esta secuencia, vamos a hacer ingeniería de características para generar algunas características útiles. A continuación, utilizamos la característica de secuencia para alimentar nuestro modelo de IA. Basándonos en nuestro modelo de IA predeciremos el final. Dado el reciente avance de la IA, también podemos utilizar la simulación dinámica [inaudible] o incluso métodos experimentales para verificar algunos genes candidatos como posibles impulsores del cambio de virulencia. Dado el modelo de IA, también generalizamos esto - básicamente usamos el aprendizaje por transferencia de los coronavirus, presumiblemente a otros virus como la gripe o el VIH, el VPH. Hay muchos virus. En realidad no soy experto en biología, sólo en informática.

Y así los desafíos que realmente estamos tratando con - mis disculpas, es automáticamente reenviar de nuevo. El reto es realmente predecir la patogenicidad de las secuencias. También podemos encontrar una regla potencial y probar la capacidad general mediante la transformación de predecir el - futuros virus y predecir el resultado. Será AI habilitado, el sistema de alerta temprana. Queremos utilizar la IA para predecir el desarrollo de vacunas. Dado que sabemos cómo cambia la varianza de virulencia. Así que esa es la gran imagen de esto.

Hoja 3

Así que aquí están algunos resultados que hemos desarrollado - que hemos logrado. Esta es una estimación de la aptitud viral basada en la información - los datos que tenemos. Esto es en realidad para la sub-variante Omicron y podemos comparar la variante Omicron utilizando un método con el desarrollo de comparación por pares. Entonces podemos estimar la diferencia relativa entre esas variantes.

Hoja 4

Basándonos en estos datos podemos generar el llamado paisaje de fitness viral. Para los conocedores de la teoría de la evolución - en la evolución existe el llamado paisaje de fitness y realmente hay muchos argumentos al respecto, pero también hay mucha teoría basada en el paisaje de la evolución. Con el paisaje de la aptitud de un virus potencial, podríamos en teoría predecir su trayectoria evolutiva. Esto - podemos - tenemos un método para hacer esto. Este es el resultado estimado para las subvariantes alfa beta delta omicron en EEUU. También estamos en el proceso de aplicar esto a la gripe o la viruela. No tenemos suficientes datos para la viruela u otros virus, pero la gripe parece prometedora. Se trata de un hallazgo importante de nuestro proyecto actual.

Hoja 5

He aquí algunos resultados del modelo. Sin divulgar demasiados detalles sobre el modelo de IA que hemos desarrollado. Hay realmente un montón de parámetros, pero la precisión del modelo actual puede alcanzar una precisión bastante alta. Por supuesto, esto se basa en lo que sabemos. El reto es realmente desconocido. Aunque estemos prediciendo lo desconocido, los datos ya son conocidos. En el aprendizaje automático, los entrenamos con entradas de entrenamiento y prueba. El reto de predecir lo desconocido sigue siendo una cuestión abierta.

Hoja 6

Algunas personas pueden argumentar, bueno, ¿por qué tenemos que hacer aprendizaje profundo? Hay muchos otros métodos estadísticos, genéticos, genómicos, de tratamiento cuantitativo, ¿por qué tenemos que recurrir al aprendizaje profundo? Aquí tenemos algunas pruebas que demuestran que el aprendizaje profundo, como mínimo, detecta firmas diferentes de los estudios estándar de asociación de todo el genoma. A la izquierda, un resultado de nuestro aprendizaje profundo. Esto es en realidad sobre la base de la OMS etiquetados varianzas Alfa, Beta, Delta, Omicron, y otros. Queremos ver qué firma en el genoma del SARS-CoV-2 son importantes para contribuir a este aumento de la virulencia. Esas señales altas que puedes ver al final del genoma del SARS-CoV-2. Y alrededor de esta región que estoy resaltando, entre 20.000 y 25.000, se encuentra el gen spike que la OMS [ha identificado] y ante el que reaccionará la mayor parte de la respuesta inmunitaria. Según los conocimientos convencionales, así es como la OMS clasificó las variantes del SARS-CoV-2. Sorprendentemente, los modelos de IA captan esas señales, pero no son las más fuertes, sino muchas otras. Algunas de las señales más fuertes no están ahí. El uso de métodos genómicos estadísticos convencionales, como el estudio de asociación del genoma completo, sigue eligiendo el gen de la espiga como la señal más fuerte. Hay algunos otros, pero no muy fuertes. Así que en este caso, la IA y la selección del método estadístico convencional, o al menos, dan un peso diferente a esas señales. Esto es sorprendente y también tranquilizador, en cierto modo. Así que elegimos la sabiduría convencional, el aumento, pero también elegimos algunas otras señales que pueden o no ser verificados por el método experimental. Pero ¿cómo predecimos y cómo verificamos esto, verdad? Así que usamos la IA para predecir muchas cosas y sabemos que la IA puede generar muchas predicciones falsas. En este caso, ¿cómo podemos verificarlo? Eso es bastante desafiante.

Hoja 7

Estamos intentando, en general, dos maneras diferentes. Una es generar un sistema experimental modelo para verificar los hallazgos. Nuestro equipo está utilizando un modelo biológico este es el [inaudible] también tenemos un sistema de línea celular para convertir esos genes en el ciclo de vida y medir sus actividades relativas.

Hoja 8

Entonces también tenemos una persona computacional para realizar la dinámica molecular para simular cómo esas mutaciones afectan [inaudible] de su actividad. En este caso, ACE2 y RBD, pero también tenemos otro [inaudible] en el que predecir en diferentes regiones cómo reaccionan con los genes humanos - potencial dirigido a los sistemas inmunológicos humanos.

Hoja 9

También organizamos un taller y paneles de expertos para discutir cómo diseñar IA confiable para promover la confianza social en la IA - IA equitativa. La pandemia COVID ha demostrado que hay una gran disparidad en nuestro sistema actual, por lo que si utilizamos la IA para predecir las pandemias futuras, puede amplificar fácilmente las disparidades ocultas en nuestro sistema actual, probablemente también en nuestros datos actuales. Si reconocemos esto y organizamos un panel bastante diverso, e incluyendo gente de Sudáfrica, Kenia, hay una persona de Europa y a través de la - con diferentes antecedentes como abogados, médicos, políticos, gubernamentales - NIST - agencias gubernamentales para tener todo tipo de perspectiva sobre cómo promover la IA y promover la confianza en la IA en las comunidades históricamente marginadas.

Hoja 10

También enfatizamos mucho el desarrollo de la fuerza laboral dada nuestras colaboraciones con las Instituciones de Servicio a las Minorías, incluyendo Spelman College, North Carolina A&T y Catholic University of America, que contienen muchos estudiantes hispanos, y mi institución de origen, la Universidad de Tennessee, Chattanooga. También colaboramos con un sistema hospitalario Global South.

Hoja 11

Por último, quiero dar las gracias a NSF, UTC y la Iniciativa AI Tennessee por su apoyo. Y por último, yo no puse en las diapositivas, NSF ha lanzado llamada para la fase 2 de solicitud para un Centro Nacional, a estar en una escala probablemente 10 veces mayor, si no 100 veces más grande, así que estoy buscando colaboradores, especialmente con el compromiso de la comunidad, He oído que alguien trabaja en política pública. Así que deseo conectar después de esto.

Así que me detendré aquí, gracias.