

[Centro de Información de COVID \(CIC\): Charlas Científicas de Relámpago](#)

Transcripción de una Presentación por Judy Fox (Universidad de Virginia), enero 30 de 2024.



Título: [Expediciones: Investigación Colaborativa: Epidemiología Computacional Global Pervasiva](#)

[Perfil de la base de datos del CIC de Judy Fox](#)

Premio NSF #: [2151597](#)

[Grabación de YouTube con diapositivas](#)

[Información del seminario web del CIC del Invierno de 2024](#)

Traductor: Yonara Anastacio

Transcripción

Diapositiva 1

Gracias, Lauren, por tu amable introducción, y gracias a los anfitriones del Centro de Información de COVID del 2024, así como a nuestros compañeros ponentes, estudiantes y personal. Hoy voy a presentar nuestro proyecto *Expeditions* financiado por la NSF. Este es un proyecto de investigación colaborativa sobre epidemiología computacional global y ubicua. Soy Judy Fox y trabajo en la Escuela de Ciencias de Datos y en el Departamento de Ciencias de la Computación de la Universidad de Virginia.

Diapositiva 2

Un poco de contexto sobre este proyecto: es un proyecto multi institucional liderado por el Dr. Madhav Marathe, quien es el investigador principal coordinador. Contamos con colegas dentro de la Universidad de Virginia y del Instituto de Biocomplejidad. Se trata de una colaboración multiinstitucional, y tenemos colegas e investigadores maravillosos. Para mí, es una experiencia realmente inspiradora.

Diapositiva 3

Quería hablar sobre el futuro. Esa es la parte emocionante de mi presentación hoy. Imaginen el COVID en 2025. ¿Dónde estaremos? Queremos pasar de la intervención a la prevención, porque

las enfermedades infecciosas representan un problema social. Para 2050, se proyectan más de 10 millones de muertes anuales y un impacto económico de más de 100 billones de dólares. Hace apenas unos años, salimos de esta pandemia en la que, solo en los EE. UU., hubo más de 1.1 millones de personas fallecidas y más de 100 millones de casos. Esto equivale a casi un tercio de las familias que fueron infectadas. Gran parte de esto podría haberse prevenido con políticas gubernamentales mejor informadas. Sin embargo, el COVID-19 representa un problema de datos complejo. Primero, contamos con datos ruidosos que llegan constantemente. Es bastante difícil aprender y predecir las tendencias debido a la falta de datos o datos ruidosos. Hacer que las proyecciones de infecciones sean comprensibles o explicables también puede ayudar en la toma de decisiones. Esto puede contribuir a identificar las áreas geográficas y temporales clave, para que podamos alertar a los gobiernos y lograr una asignación más efectiva de recursos, previniendo así la propagación de la enfermedad.

Diapositiva 4

Quiero centrar el resto de la charla en la investigación realizada por mi grupo. Nuestro objetivo fue interpretar las infecciones de COVID-19 a nivel de condado en los Estados Unidos. Aplicamos el modelo de inteligencia artificial *Transformer*, que es un tipo de modelo de aprendizaje profundo utilizado en los modelos de lenguaje de gran tamaño.

Diapositiva 5

Una de las áreas en las que nos estamos enfocando es en plantear la pregunta: ¿por qué necesitamos predicciones? Las predicciones utilizando datos en tiempo real fueron destacadas en 2009 por el Dr. Harvey V. Fineberg y la Dra. Mary Elizabeth Wilson. Destacan el aspecto importante de por qué es relevante: queremos estudiar utilizando los datos más recientes, para estudiar el control de enfermedades e intentar observar y predecir. Las intervenciones serán las acciones que se tomen en el punto máximo, pero la intervención anticipada ayudará a aplanar la curva antes de que llegue a su pico.

Diapositiva 6

Usamos el modelo de aprendizaje profundo *Temporal Fusion Transformer* (TFT), el cual puede realizar pronósticos en tiempo real. En nuestros experimentos, utilizamos los últimos 13 días para predecir los próximos 15 días. La recopilación de datos proviene de diferentes modelos de predicción. Los categorizamos en conjuntos de datos estáticos co-variantes y dinámicos, como casos y muertes. También contamos con algunas entradas conocidas, como las próximas festividades de Navidad y Año Nuevo.

Diapositiva 7

Con un modelo así, nuestro objetivo es comprender cómo utilizar la IA interpretable para obtener el conocimiento e información sobre dónde y cuándo ocurrirá la infección. ¿Qué países están en mayor riesgo? ¿Cuáles son las comunidades vulnerables? Y tratamos de ayudar. El camino de este estudio abarca todo esto, con muchos obstáculos que debemos superar.

Diapositiva 8

El problema radica en la falta de precisión general. ¿Cómo puede nuestro modelo predecir de manera que tengamos plena confianza en la exactitud de nuestras predicciones? Además, una vez que obtenemos esta predicción, ¿cómo podemos explicar a los responsables de políticas cuáles son los factores clave que están impulsando el aumento actual de los casos? Para ello, necesitamos un análisis más detallado de los datos, a un nivel más granular, como las características específicas de cada condado. También es fundamental tomar decisiones en tiempo real para garantizar su relevancia.

Diapositiva 9

Es un estudio muy completo, ya que la disparidad en la salud no solo afecta a las personas, sino que también tiene un impacto socioeconómico. Hemos recopilado más de dos años y medio de datos de 3,142 condados en EE. UU. Clasificamos y reducimos las características del conjunto de datos de mortalidad de veinte a seis. Dos de ellas son estáticas: la disparidad en la salud y el grupo de edad de la población. Las demás son características observables, incluyendo la vacunación, la propagación de enfermedades, los casos transmisibles y la movilidad. Incorporamos estos factores en eventos conocidos y desconocidos para desarrollar un modelo de IA multimodal y articulado de gran complejidad.

Diapositiva 10

Permítanme cambiar de enfoque para mostrarles algunos de nuestros resultados de predicción. Comparamos el modelo TFT con el LSTM, un modelo básico de secuencia a secuencia. En el gráfico del lado izquierdo, podemos ver que el modelo TFT tiene el mejor desempeño. Su precisión es mayor y su error es menor, lo que se refleja en el mensaje de error.

Diapositiva 11

¿Qué sustenta la comprensión de la IA interpretable? Esto se basa en el mecanismo de atención dentro de una arquitectura codificador-decodificador. Dicho mecanismo es la base de los modelos de lenguaje actuales, incluido ChatGPT. Gracias al mecanismo de atención múltiple,

podemos capturar el contexto de la enfermedad a lo largo del tiempo, refinando así el espacio en el que analizamos las características y la relación entre causa y efecto. Esto nos permite priorizar los patrones espaciales y temporales en las áreas de mayor incidencia. A la derecha, se muestra la arquitectura de un modelo TFT. Se ingresan características del pasado para predecir eventos futuros. En este caso, se analizan los casos de infección, encapsulando y combinando las características, especialmente las estáticas, dentro de un modelo de secuencias para capturar patrones temporales dependientes. Estos patrones se propagan a través de un mecanismo de autoatención para filtrar la atención multijefe interpretable, permitiéndonos centrarnos en los patrones y regiones más significativas.

Diapositiva 12

Este modelo de predicción es capaz de identificar el patrón cíclico en la notificación de casos de COVID. También incorpora eventos temporales específicos, como días festivos y fines de semana, los cuales se pueden etiquetar claramente en este gráfico. En el lado derecho, podemos analizar retrospectivamente qué período de tiempo tiene la mayor influencia en la predicción futura dentro del intervalo de 0 a 13 días.

Diapositiva 13

Aquí está el gráfico de tendencias que podemos predecir. Al elegir los 100 condados más grandes, se puede ver que tenemos la predicción en comparación con la verdad de los hechos. Además, podemos comparar los condados con poblaciones más pequeñas; se puede observar que el resultado coincide mucho mejor, con menos picos y valores diferentes.

Diapositiva 14

¿Qué hay sobre la información de ubicación? Este gráfico muestra, en el lado izquierdo, la autoatención del modelo de IA, que captura la intensidad a nivel de condado. Esto no sería posible si nuestros datos estuvieran a nivel estatal. En el lado derecho, se muestra una representación de datos de los casos acumulativos según el CDC en más de 3,000 condados de EE. UU. Si comparamos estos dos resultados, podemos observar la correlación entre ambos conjuntos de datos y sus resultados.

Diapositiva 15

Medimos la correlación y concluimos que podemos interpretar el modelo de IA analizando los pesos de autoatención a nivel de condado. Hay una fuerte correlación entre el comportamiento del modelo y la predicción de los casos en comparación con los valores reales.

Diapositiva 16

R^2 proporciona la información que los responsables de políticas necesitan. Creemos que una pequeña reducción en la transmisión en los focos de infección puede llevar a una gran disminución en el número de contagios, especialmente en las etapas iniciales. La previsión en tiempo real y centrar nuestra atención en las regiones más importantes en términos de infecciones diarias es crucial. Contamos con un método detallado para capturar estas infecciones a nivel de condado, lo que reduciría significativamente el riesgo. Hay muchas líneas de trabajo futuras que podemos desarrollar a partir de este resultado. Podemos explorar numerosos impactos sociales y económicos, así como las disparidades, en estudios futuros.

Diapositiva 17

En la Universidad de Virginia (UVA), iniciamos el programa de IA para la ciencia y logramos atraer a más de 3,000 estudiantes de pregrado para participar. Seleccionamos a una docena de estudiantes para involucrarse en nuestro proyecto.

Diapositiva 18

Queremos estudiar más a fondo la sensibilidad de los grupos de edad de la población utilizando un estudio basado en el índice de Morris para determinar, dentro de cada grupo poblacional, quiénes son los más vulnerables a las infecciones por COVID.

Diapositiva 19

Elegimos un modelo de aprendizaje profundo para series temporales porque el mecanismo de tensión mencionado anteriormente puede proporcionar información y ayudar a comprender cómo se realiza la predicción del modelo.

Diapositiva 20

Hay varias lecciones que aprendimos de este camino. Las enfermedades infecciosas han sido un problema socioeconómico a nivel mundial que impacta la salud pública y la economía a gran escala. A partir de la pandemia de COVID-19, nuestra primera lección aprendida es que las pruebas son fundamentales para comprender el progreso de la pandemia. La segunda es que muchas infecciones varían significativamente en diferentes regiones más pequeñas. La situación es bastante dinámica. La mejor manera de abordar esto en las políticas es adaptándolas al nivel local. Existen muchas formas en las que podemos mejorar, incluyendo el desarrollo de herramientas como las que estudiamos para predecir con precisión la infección por COVID y futuras enfermedades infecciosas. Esto ayudará a los responsables de la toma de decisiones a

intervenir de manera basada en la ciencia. La intervención es el futuro. Queremos estar preparados y listos para futuras crisis, como una nueva pandemia.

Diapositiva 21

Antes de finalizar, realmente quiero compartir con todos que necesitamos construir confianza. Si hay alguna lección aprendida de la pandemia pasada, es que debemos mejorar la forma en que explicamos a la comunidad, a los expertos y a los responsables políticos lo que está sucediendo y cómo podemos aprovechar las políticas para generar impacto. Queremos desarrollar un método interpretable para modelar y evaluar cuantitativamente nuestras estrategias. También deseamos explicar el comportamiento de nuestro modelo y las predicciones basadas en inteligencia artificial a personas no expertas, incluidos el público en general y los estudiantes. Colectivamente, esperamos construir un futuro en el que estemos preparados para los eventos futuros.

Diapositiva 22

Con eso, quiero agradecerles a todos por participar en el taller de hoy. He compartido parte de nuestro trabajo aquí. No duden en contactarme. Espero con interés las futuras discusiones al final de este taller. ¡Gracias!