

[Centro de Información de COVID \(CIC\): Charlas científicas de relámpago](#)

Transcripción de una presentación de Amit Sheth (Universidad de Carolina del Sur) y Valerie Shalin (Wright State University), junio de 2022



Título: [Análisis semántico de redes sociales y nuevos Big Data para comprender el impacto de COVID-19 en la salud mental, la adicción y la violencia de género](#)

[Grabación de YouTube con diapositivas](#)

[Información del seminario web del CIC de junio 2022](#)

Editora de la Transcripción: Lauren Close

Editora de la Traducción: Isabella Graham Martínez

Transcripción

Amit Sheth:

Diapositiva 1

De acuerdo, de acuerdo. Así que vamos a discutir nuestro análisis de los problemas de salud pública. Cuando analizamos gran cantidad de contenido de redes sociales, el enfoque está en la salud mental y el análisis relacionado con la adicción y lo que llamamos calidad social del índice. El Instituto AI de la Universidad de Carolina del Sur y la Universidad Estatal de Georgia, Wright State University son participantes o los miembros del equipo pertenecen a estos.

Diapositiva 2

Así que todos sabemos del impacto masivo de esta pandemia.

Diapositiva 3

Y esto también ha resultado en un impacto significativo en la depresión, la ansiedad, otros problemas de salud mental, en las adicciones y el abuso de sustancias.

Diapositiva 4

Así que dijimos, bueno, podemos entender cómo varias decisiones políticas, políticas sociales, económicas y de salud pública, y las decisiones que el gobierno y los legisladores han tomado afectan el bienestar de la sociedad - los miembros de la sociedad. Y nuestra caja de herramientas es el conocimiento - análisis mejorado de las redes sociales.

Diapositiva 5

Por lo tanto, lo posicionamos como lo siguiente: capturamos los datos entre el período de, como ven aquí, marzo de 2020 y finales de julio a enero de 2021. Eso cubrió dos grandes ondas COVID-19. Teníamos 25 términos de rastreo. Recogimos 12 mil millones de tweets. Además de este big data, teníamos muchos otros

datos que usar. Y esa es una parte metódica muy importante del estudio porque no se trata solo de analizar las redes sociales. Los datos no son suficientes, se necesita mucha otra información relevante para poder analizar los datos, así que teníamos información relacionada con la ubicación, miramos subreddits específicos para entrenar nuestros modelos de lenguaje, y teníamos bases de conocimiento relacionadas con la salud mental, ontología del abuso de drogas DSM-V, DAO, DBpedia, Wikidata, UMLS. Además, tuvimos que examinar los acontecimientos sobre el terreno. Por ejemplo, ya sabes, alivio de préstamos, alivio de los ingresos de los hogares, política de pruebas, todas esas cosas que todas esas decisiones que varios líderes estatales hicieron o federales, ya sabes, los legisladores hicieron. Así que también tenemos que tener una comprensión de lo que sucedió cuando y donde. Y hay otros eventos que también tenemos que entender.

Diapositiva 6

Nos centramos en entender el contenido de las redes sociales relacionado con la salud mental. Así que puedes ver algunos ejemplos aquí.

Diapositiva 7

Y desarrollamos una medida empírica que llamamos Índice de Calidad Social. Agrega los componentes de salud mental y la adicción y los componentes del trastorno de consumo de sustancias, ¿verdad? Así que analizar los datos de redes sociales para estos componentes para llegar a un Índice de Calidad Social.

Diapositiva 8

Hay una infraestructura bastante completa ya que el análisis de este tipo es bastante desalentador, por lo que tienes datos de Twitter, pero tienes artículos de noticias. Voy a mencionar rápidamente por qué su artículo de noticias - su base de conocimientos diversos. Luego, hay que hacer toda una variedad de análisis, se puede ver allí en el segundo cuadro vertical y luego va a modelos de lenguaje y modelos de temas, asignaciones semánticas, una comprensión más profunda. Y, ya sabes, nuestro equipo trabaja en el aprendizaje infundido conocimiento, por lo que tenemos un particular, ya sabes, mejorado métodos de aprendizaje profundo que utiliza el conocimiento del dominio para entender mejor el idioma. A partir de ahí, para entender el contenido relacionado con la depresión, la ansiedad, la adicción, el uso de sustancias, todo en el contexto de COVID, que conduce a cálculos SQI, había componentes de entrenamiento de modelos de lenguaje, y todos los que no se muestran aquí en esta imagen.

Diapositiva 9

¿Cuáles son las innovaciones? Utilizamos las noticias para identificar continuamente nuevas entidades de interés en COVID. Y en este contexto, usamos extracciones de ubicación usando Geonames Ontology, Open Street Map API y otras cosas. Utilizamos múltiples vocabularios y gráficas de conocimiento para la extracción semántica, utilizamos un corpus subreddit especializado para el entrenamiento de modelos de lenguaje y modelos temáticos. Y entrenamos clasificadores para escalar el análisis de big data. Así que no pudimos hacer el, ya sabes, en nuestro análisis - para analizar tales grandes cantidades de datos, tenemos que crear estos clasificadores y luego entender la depresión, la adicción, la ansiedad y otros problemas. Con esos antecedentes, se lo voy a pasar a Valerie para explicarle lo que hemos encontrado.

Valerie Shalin:

Diapositiva 10

Gracias a ti, Amit. Permítanme darles una idea del tipo de análisis que se hacen posibles con el tipo de capacidades que Amit acaba de describir. Así que aquí hay algunos gráficos de estado de muestra de las tasas de COVID y SQI en el tiempo. Al igual que Brandon [Johnson], creemos que el tiempo es absolutamente el problema aquí. Así que las tasas COVID per cápita están a la izquierda, eje y. Y el SQI está en el eje derecho. Arriba es malo, abajo es bueno. El tiempo está en el eje x y hay una

discontinuidad marcada con las marcas hash. Y no espero que lea estos gráficos cuidadosamente, excepto que sin duda notará que hay muy poca correlación entre el Índice de Calidad Social que hemos identificado y la prevalencia de COVID. Hay un lugar - whoops, retroceder una diapositiva, solo por un segundo - hay un lugar que usted notará una correlación. Y eso es hacia el final, y creemos que eso probablemente refleja una variable latente. Esa es la temporada de vacaciones, esa es la temporada de vacaciones de invierno, y probablemente lo que está pasando es que las vacaciones de invierno están aumentando las tasas de COVID y ya sabemos que las vacaciones de invierno son muy difíciles para la salud mental. Así que eso tiene mucho sentido y sirve como una validación informal de la métrica SQL. Próxima diapositiva.

Amit:

No puedo avanzar, así que no sé qué - Lauren, no puedo...

Lauren Close:

Veamos, quizás pueda compartirlo en tu nombre.

Florence Hudson:

¿Has intentado hacer clic en la diapositiva o usar las flechas arriba o abajo? ¿Probaste algunas opciones diferentes? Bien, ahí lo tienes.

Valerie:

Diapositiva 11

Bien, así que uno de los puntos que queremos hacer es que las vacaciones, por supuesto, no son el único evento que ahora podemos examinar. Y este es un gráfico del estado de Washington. Y lo que les estamos mostrando son los cambios de política que hemos identificado en las noticias alineados con los diversos períodos que hemos marcado allí con las líneas verticales. Así que nuestro punto es que ahora es posible buscar patrones que relacionen este Índice de Calidad Social con la implementación de políticas por estado. Próxima diapositiva, Amit.

Diapositiva 12

Así que SQL es una medida estandarizada. Es independiente de la población estatal o de la cantidad de actividad de las redes sociales y debido a que está estandarizada de esa manera podemos comparar estados. Y hemos hecho - próxima diapositiva.

Diapositiva 13

Algunos están agrupados para ver si podemos discernir patrones. Y ven que Connecticut, Luisiana, Nueva Jersey, Nevada, etc. todos parecen seguir el mismo tipo de patrón. SQL es malo al principio, se pone mejor, y luego empeora. Esto es para un período de tiempo de alta resolución truncada. Y luego un grupo diferente: Florida, Georgia, Michigan, etc. tienen un SQL mejor - SQL peor patrón. Así que podemos revelar algunos de estos patrones con el tiempo. Por supuesto, tenemos que hacer más análisis longitudinales y series de tiempo para sacar las causas potenciales, pero nos parece que son los problemas económicos los que van a ser el factor principal aquí. Próxima diapositiva.

Diapositiva 14

Entonces, ¿qué es posible ahora con las herramientas que el equipo de Carolina del Sur ha desarrollado? Tenemos un análisis basado en el conocimiento de las redes sociales a nivel abstracto y proporciona un conjunto de datos de muy alta resolución, tanto en lo que respecta al tiempo y el espacio. Es rápido y proporciona medidas separables de la salud mental que no son factibles en las

métricas de encuesta. Podemos obtener muchos y muchos datos repartidos en grandes porciones de espacio y tiempo y ese es el tipo de datos que necesita para sacar las confusiones y comenzar a evaluar si sus decisiones políticas estaban teniendo un efecto en su población. Así que esta es una herramienta de investigación realmente importante que mejora nuestras capacidades en ciencia, pero más ampliamente permite el monitoreo en tiempo real y la preparación para la mitigación. Y, por supuesto, apoya a los responsables políticos en general. Y esa es nuestra presentación para usted.