La inteligencia artificial aprende sobre lenguaje a través de los ojos y oídos de un niño

Un nuevo modelo de aprendizaje automático, nutrido con videos y audios grabados a un bebé desde los seis meses hasta su segundo cumpleaños, ha aportado nuevos conocimientos sobre la adquisición del habla en la infancia. Los resultados de este trabajo de investigadores de la Universidad de Nueva York ofrece información sobre cómo aprendemos palabras y conceptos y servirá para desarrollar sistemas de IA que usen el lenguaje de forma más parecida a la humana.

Ana Hernando 1/2/2024 20:00 CEST

Foto de un bebé de 18 meses con una cámara en la cabeza. / Wai Keen Vong

Entre los seis y los nueve meses de edad, los niños aprenden sus **primeras palabras** y empiezan a relacionarlas con objetos y conceptos del mundo real. Cuando tienen entre 1,5 y dos años, la mayoría puede comprender una media de 300 palabras. Sin embargo, no se sabe bien cómo las adquieren y las relacionan con sus equivalentes visuales.

Comprender mejor este proceso podría servir de base a los sistemas de

inteligencia artificial (IA) de nueva generación que desarrollan vínculos entre las palabras y las representaciones visuales.

Los actuales sistemas de IA, como **Chat GPT-4**, ya pueden aprender y utilizar el lenguaje humano, pero lo hacen a partir de **cantidades astronómicas de datos** lingüísticos, mucho más de lo que reciben los niños cuando aprenden a entender y hablar. Los mejores sistemas de IA se entrenan con textos que contienen billones de palabras, mientras que los niños solo reciben millones al año.

Debido a esta enorme laguna de datos, los investigadores se han mostrado escépticos ante la posibilidad de que los recientes avances de la IA puedan decirnos mucho sobre el **aprendizaje y el desarrollo del lenguaje humano**.

Para avanzar en este ámbito, un equipo de la Universidad de Nueva York (NYU, por sus siglas en inglés) decidió desarrollar un nuevo modelo de aprendizaje automático, no a partir de datos masivos, sino tomando como ejemplo la experiencia de cómo aprende a hablar un único niño, al que llamaron bebé S. Los resultados del estudio se publican ahora en *Science.*

El equipo ha desarrollado un modelo de IA, no a partir de datos masivos, sino tomando como ejemplo la experiencia de cómo aprende a hablar un único niño (bebé S)

Los autores diseñaron un experimento que consistió en entrenar un sistema de IA multimodal a través de los ojos y los oídos de bebé S. Para ello utilizaron grabaciones de vídeo de una cámara frontal que recogieron desde que tenía seis meses hasta su segundo cumpleaños. Y examinaron si el modelo podía aprender palabras y conceptos presentes en la experiencia cotidiana de un niño.

Wai Keen Vong, investigador de la universidad estadounidense y primer

firmante del estudio, explica a SINC que en su ensayo utilizaron el conjunto de datos SAYCam, "un recurso muy rico e interesante que consiste en vídeos capturados con cámaras montadas en la cabeza en niños en desarrollo".

"Nos centramos en un solo niño (**bebé S**) porque era el que tenía la mayor cantidad de datos del habla transcritos y esto nos facilitaba la tarea de modelarlo. Todo ser humano necesita aprender a hablar a partir de su propia información –y no de la de otros–, por lo que explorar si es posible adquirir aspectos del lenguaje con un modelo computacional, a partir de la información sensorial de un solo niño, es una forma única de abordar esta cuestión", subraya este científico de datos y experto en IA.

Las conclusiones del estudio demuestran que el modelo, o red neuronal, puede aprender un número considerable de palabras y conceptos utilizando fragmentos limitados de la experiencia del niño. El coautor aclara que los vídeos solo captaron alrededor del 1 % de las horas de vigilia de bebé S, pero fue suficiente para nutrir su modelo.

> *Es el primer trabajo que aborda realmente este tipo de aprendizaje con datos reales y naturales, y de una forma que refleja con mayor exactitud lo que los bebés ven y oyen*
>
> Wai Keen Vong, primer autor (Universidad de Nueva York)

Las palabras y sus equivalentes visuales

"En nuestra investigación estamos interesados en un sistema que aprenda las relaciones entre las palabras y sus equivalentes visuales –por ejemplo, cómo saber que la palabra 'pelota' se refiere a imágenes de cosas redondas que rebotan–. Aunque ya se han propuesto muchos modelos computacionales de aprendizaje de palabras, a menudo se han entrenado con entradas simplificadas o con ciertos supuestos incorporados, que en realidad no funcionan muy bien (¡o no funcionan en absoluto!) cuando se aplican a imágenes reales o al lenguaje natural", dice Vong.

El experto destaca que este "es el primer trabajo que aborda realmente este tipo de aprendizaje con datos reales y naturales, y de una forma que refleja con mayor exactitud lo que los bebés ven y oyen". En su

opinión, los resultados del estudio “demuestran cómo los recientes avances algorítmicos, emparejados con la experiencia de un único niño, tienen el potencial de remodelar nuestra comprensión de la **adquisición temprana del lenguaje y de los conceptos**".

El equipo analizó el proceso de aprendizaje de bebé S en sesiones de vídeo semanales desde los seis a los 25 meses, utilizando más de 60 horas de grabación. Los audios contenían aproximadamente un cuarto de millón de palabras comunicadas. Muchas de ellas estaban repetidas y vinculadas con fotogramas de vídeo de lo que veía cuando las pronunciaba al realizar distintas actividades a lo largo del desarrollo, como comer, la lectura de libros y los juegos.

A continuación, los investigadores de la NYU entrenaron una **red neuronal multimodal** con dos módulos separados: uno tomaba fotogramas de vídeo individuales (el codificador de visión) y otro incluía el habla transcrita del niño (el codificador de lenguaje).

Algoritmo de aprendizaje contrastivo

Los dos codificadores se combinaron y entrenaron mediante un algoritmo llamado aprendizaje contrastivo, cuyo objetivo es aprender características de entrada útiles y sus **asociaciones intermodales**. Por ejemplo, cuando uno de los padres dice algo a la vista de los hijos, es probable que algunas de las palabras utilizadas se refieran a algo que el

niño pueda ver, lo que significa que la comprensión se inculca vinculando las señales visuales y lingüísticas.

Después de entrenar el modelo, los investigadores lo probaron utilizando el mismo tipo de evaluaciones que se emplean para medir el aprendizaje de palabras en los bebés: presentándole la palabra objetivo y una serie de cuatro imágenes diferentes y pidiéndole que seleccionara la imagen que correspondía a la palabra objetivo.

Los resultados mostraron que el modelo de IA intermodal era capaz de aprender un número considerable de palabras y conceptos presentes en la experiencia cotidiana del niño y luego generalizar

Los resultados mostraron que el modelo de IA intermodal era capaz de aprender un número considerable de palabras y conceptos presentes en la experiencia cotidiana del niño. Además, para algunos de los conceptos que aprendía, el sistema podía generalizarlos a instancias visuales muy distintas de las observadas durante el entrenamiento, algo que también ocurre con niños cuando se les somete a pruebas en laboratorio.

"Estos hallazgos sugieren que este aspecto del aprendizaje de palabras es factible a partir del tipo de datos reales que reciben los niños, mientras utilizan mecanismos de aprendizaje relativamente genéricos como los que se encuentran en las redes neuronales", observa **Brenden Lake**, también de la NYU y autor principal del estudio.

Mejorar el lenguaje de la IA

Respecto a las implicaciones que el trabajo puede tener en la mejora del lenguaje de los sistemas de IA, Vong señala: “Los niños son extraordinariamente hábiles en el aprendizaje del lenguaje. Con dos años, ya tienen mejores resultados que nuestro modelo. Por otro lado, de adultos hablamos utilizando solo cientos de millones de palabras,

comparado con los últimos sistemas de IA, que necesitan miles de millones, o incluso billones, para adquirir fluidez. Creo que un estudio más profundo de la adquisición del lenguaje humano podría arrojar luz sobre cómo podemos **aprender de forma tan eficiente a partir de datos limitados**, y es de esperar que estos conocimientos se trasladen también a la inteligencia artificial", destaca.

> *Un estudio más profundo de la adquisición del lenguaje humano podría arrojar luz sobre cómo podemos aprender de forma tan eficiente a partir de datos limitados*
>
> Wai Keen Vong

El investigador también comenta a SINC que en el estudio se tuvieron muy en cuenta los **aspectos éticos**, ya que se usaban grabaciones de vídeo en primera persona de un niño.

"Debido a la naturaleza sensible de la información, para llevar a cabo la investigación con este conjunto de datos –alojados en la web de la NYU Databrary– se requirió la **aprobación ética** de antemano de la universidad. Es algo que siempre estuvo en mi mente al hacer la investigación. La otra consideración importante era preservar la **privacidad de los padres y del niño**, pero puedo compartir que su nombre es Sam, ahora tiene 11 años, le va muy bien y estudia 6º curso", cuenta Vong.

Referencia:

Wai Keen Vong et al. "Grounded language acquisition through the eyes and ears of a single child". *Science* (2024)

TAGS | INTELIGENCIA ARTIFICIAL | LENGUAJE | BEBÉ | NIÑO | APRENDIZAJE AUTOMÁTICO | RED NEURFNEUAL |