



Una dicotomía engañosa y una paradoja ética





RICARDO BAEZA-YATES

Profesor de Investigación del Instituto de Inteligencia Artificial Experiencial de Northeastern University, además de Profesor Titular a tiempo parcial en los Departamentos de Tecnologías de la Información y de las Comunicaciones de la Universitat Pompeu Fabra en Barcelona y Ciencias de la Computación de la Universidad de Chile, donde además es Investigador Senior del Instituto Milenio Fundamentos de los Datos. Entre 2006 y 2016, fue vicepresidente de investigación de Yahoo! Labs, primero desde Barcelona y luego en Sunnyvale, California. Es ACM e IEEE Fellow. En Twitter lo encuentras como @PolarBearby.

Hace poco más de un año, el 20/02/2020 —¿fecha aleatoria?—, Geoff Hinton, uno de los padres del aprendizaje profundo, tuiteó lo siguiente [1]:

Suponga que tiene cáncer y tiene que elegir entre un cirujano de IA de caja negra que no puede explicar cómo funciona, pero tiene una tasa de éxito del 90% y un cirujano humano con una tasa del 80%. ¿Quiere que el cirujano de IA sea ilegal?

Esta provocativa (doble) pregunta incluye una dicotomía engañosa, ya que una persona racional no decidiría basándose sólo en un promedio que ni siquiera sabe cómo se calculó. Engañosa porque hay una tercera posibilidad que es mucho mejor: *quiero un cirujano humano con apoyo de IA*. Pero el verdadero dilema

está en la legalidad del cirujano de IA, lo cual es confuso, porque si fuera ilegal, la elección sería retórica.

Los comentarios en este tuit son en su mayoría en contra del cirujano de IA por muchas razones, la mayoría de las cuales se incluyen a continuación. Para empezar, supongamos que el robot médico no es ilegal y sólo tenemos estas dos opciones. ¿Cuál es el mejor? Dando el beneficio de la duda al robot médico, la respuesta depende de la *complejidad del cáncer y la normalidad del paciente*, es decir qué tan diferente es usted de la población usada en los datos de entrenamiento. Entonces, si conocemos estos datos y usted es un caso estándar, puede elegir con seguridad al cirujano de IA. En todos los demás casos, es mejor seguir con un médico humano que pueda generalizar y lidiar con problemas inesperados en función de su experiencia. El diagrama de la Figura 1 muestra estas dos dimensiones.

Supongamos que: (1) el robot médico usó *buenos datos de entrenamiento* y aprobó todas las regulaciones legales (por ejemplo, regulación para dispositivos médicos); (2) conocemos la *distribución del error con respecto a la complejidad del caso* (aunque la mayoría de los sistemas de IA ni siquiera reportan el error promedio); y (3) conocemos los *sesgos y debilidades del sistema* con respecto a la diversidad de pacientes (por ejemplo, enfermedades actuales, peso, presión arterial, etc.). Sí, estoy *suponiendo muchas cosas*.

Como casi no existen estudios de distribución del error de predicción en función de la complejidad de la instancia del problema, supondremos que el error relativo es proporcional a la complejidad del caso dividida por el tamaño de los datos de entrenamiento usados para ese nivel de complejidad (ver Figura 2). Esto supone que los datos de entrenamiento son homogéneos, lo que difícilmente es cierto en la práctica, ya que normalmente hay menos datos para

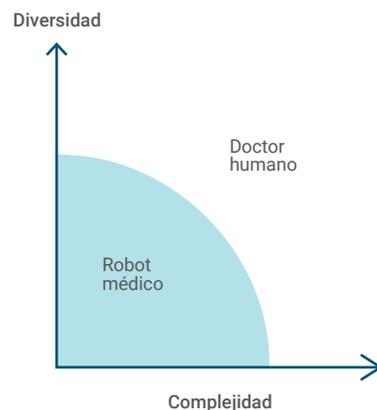


Figura 1. Preferencia de un doctor humano o un robot médico, según la complejidad del cáncer (eje x) y la normalidad del paciente (eje y) a operar.

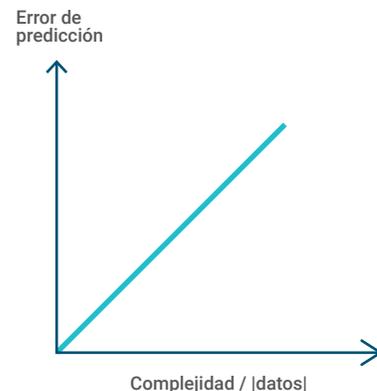


Figura 2. Error en la predicción del éxito de un robot médico, en función de la complejidad de la cirugía y el tamaño de los datos de entrenamiento.

instancias más complejas. Observe que estoy usando el error de predicción en un sentido amplio, pero en nuestro caso particular sería el error en la predicción de tener una cirugía exitosa o, en otras palabras, salvar al paciente.

Para el médico humano consideremos que (1) tiene mucha experiencia, lo que le permite transferir sus conocimientos a eventos inesperados en casos complicados,

Para tener mejores robots médicos para todos, debemos correr más riesgos con algunos pacientes, lo que tiene varias facetas éticas [...].

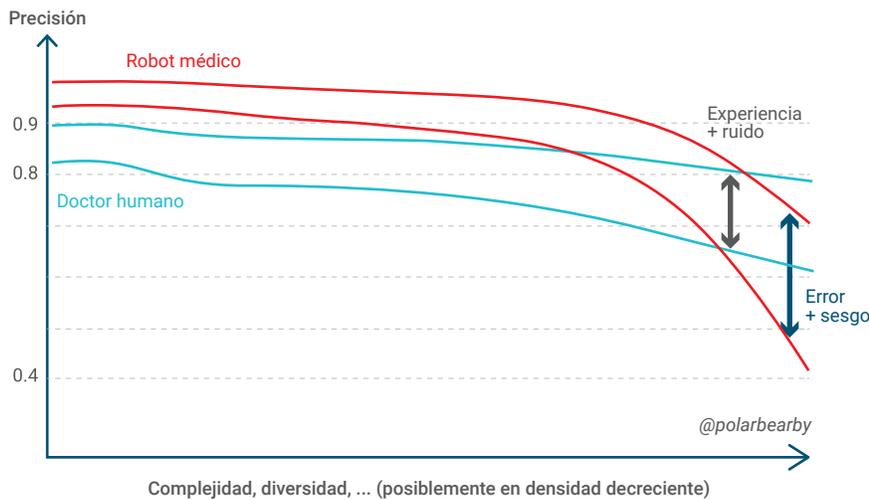


Figura 3. Precisión de éxito de una cirugía (eje y) en función de los diversos factores involucrados (eje x) para ambos doctores.

así como evitar sesgos de diversidad de la anatomía humana; pero (2) es humano y, por lo tanto, su desempeño varía según su estado de ánimo, las horas de sueño y otros factores. Estos son ejemplos del ruido que existe en las decisiones humanas y que tiene un costo tan o más alto que el de los sesgos [2, 3].

Dado lo anterior, la precisión de tener una cirugía exitosa en función de estos factores para ambos médicos viene dada por el gráfico de la Figura 3. Por lo tanto, si sabe dónde usted se encuentra en este gráfico, según la complejidad de su cáncer y lo normal/diferente que usted sea, puede tomar la decisión correcta utilizando el primer diagrama. Es decir, si no coincide con el “caso estándar”, mejor elegir un médico humano. Si se encuentra en el lado derecho del gráfico para las dos dimensiones utilizadas en el ejemplo, su riesgo es mayor y es por eso que se obtiene un cuarto de círculo.

Sin embargo, en la práctica no tenemos tanta información. Primero, deberá hablar con el médico humano para saber cuán complejo es su caso, ya que el cirujano de IA no puede explicar nada. Hinton diría que el médico inventará una historia, no una explicación, como en su entrevista en *Wired* [4] que fue refutada por ocho expertos en *Forbes* [5]. Incluso si creemos esto, las personas necesitan historias para sentirse seguras y tomar decisiones importantes. La vida, lamentablemente, incluye también las percepciones además de la realidad, basta con ver lo que está sucediendo con el novel coronavirus.

Ahora, supongamos que el médico humano le dice que su caso es bastante estándar (de lo contrario, ya sabe que probablemente esté más seguro con el médico humano). Ahora necesita conocer las características de los datos de entrenamiento, pero los trabajos que describen la investigación que conduje-

ron al cirujano de IA no son tan detallados y cuando le pregunta al fabricante, le dice que los datos son privados (según las regulaciones legales que el fabricante de cirujanos de IA debe cumplir, no porque no quiera que nadie lo audite, particularmente con respecto al sesgo de diversidad).

En este punto, su única opción es hacer una suposición informada basada en cuán “distinto” es usted. Si es bastante distinto, consulte con el médico humano. Si usted es un caso típico (por ejemplo, un hombre blanco en los Estados Unidos), puede asumir que los datos de entrenamiento representan muchas de esas personas y puede elegir con seguridad al robot médico.

Sin embargo, según la ley de Murphy, usted está en el medio de la región derecha del gráfico, donde la línea roja inferior y la amarilla superior ya se cruzaron. Mala suerte, los trabajos de investigación tampoco tienen la distribución de errores, por lo que realmente no conoce las líneas rojas. Pero, tiene suerte con las líneas amarillas, ya que encuentra un foro con muchas opiniones sobre el médico humano, pudiendo estimar su experiencia y ruido. No ocurre lo mismo con el cirujano de IA porque es bastante nuevo.

Ahora se encuentra en un dilema, ya que sabe que el cirujano de IA tiene una ventaja del 10%. Pero debido al sesgo de diversidad desconocido, la incertidumbre de un error de IA negativo y la fe en un ruido positivo, el médico humano podría ser igualmente bueno. *Piensa esto tres veces y elige al médico humano.* ¿Por qué? Sus instintos se lo dicen. Pero también sabe que el médico puede explicar más tarde lo que sucedió y en el caso de un problema, su familia podría estar psicológicamente mejor, ya que culpar a una máquina no es fácil. Más aún si el robot no puede explicar nada. Además, no ha tenido tiempo de leer todas las cláusulas minúsculas de su seguro médico y tampoco quiere arriesgarse a dejar a su familia con una enorme deuda de salud.



Si los robots médicos no son ilegales, al menos necesitan una regulación estricta con respecto a los datos de entrenamiento, pruebas estándar contra resultados sesgados y algún nivel de explicación.



¿Podemos responder ahora si el cirujano de IA debería ser ilegal? Probablemente debería ser legal pero no es una pregunta sencilla. Una ley de este tipo implica temas éticos, empatía y otros rasgos humanos. Por supuesto, si los robots médicos no son ilegales, al menos necesitan *una regulación estricta con respecto a los datos de entrenamiento, pruebas estándar contra resultados sesgados y algún nivel de explicación*, incluso si necesitan inventar historias. También deben advertirnos cuándo no utilizarlos, ya que tomar la decisión correcta, como hemos visto, no es trivial. Esto es hoy estándar en los medicamentos (por ejemplo, mujeres embarazadas, personas con pre-

sión arterial alta, alergias, etc.). Si ese es su caso, ni siquiera tendrá que elegir, el médico humano le dirá directamente que es un caso de riesgo para el robot médico. Sí, quiero este tipo de explicaciones, ¡y mejor si vienen de un doctor empático! (muy probablemente una mujer, un sesgo positivo).

Pareciera estar todo claro, sin embargo, hay una *paradoja ética* escondida en nuestra discusión anterior. Para mejorar el cirujano basado en IA, ellos necesitan aprender y para eso necesitamos tener grandes maestros humanos que puedan generar datos de entrenamiento asombrosos. Para ello, necesitan practicar en los casos más

complejos, por lo que ésta es una razón social para preferir cirujanos humanos en los casos más arriesgados. Pero para llegar a este punto, los médicos humanos deben adquirir experiencia en casos estándares, lo que implica que también deben realizar cirugías cuando el cirujano de IA sería una mejor opción. Ésta es la paradoja, para tener mejores robots médicos para todos, debemos correr más riesgos con algunos pacientes, lo que tiene varias facetas éticas que como objetivo final tienen el bienestar común de todos. Lo más justo sería hacerlo al azar, pero no es tan sencillo en un mundo capitalista lleno de sesgos. Éste ya es un tema para filósofos y economistas. ■

REFERENCIAS

- [1] Geoff Hinton, <https://twitter.com/geoffreyhinton/status/1230592238490615816>, 2/2020.
- [2] Daniel Kahneman, Andrew M. Rosenfield, Linnea Gandhi, and Tom Blaser Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making, *Harvard Business Review*, <https://hbr.org/2016/10/noise>, 10/2016.
- [3] Daniel Kahneman, Olivier Sibony, Cass R. Sunstein. *Noise: A Flaw in Human Judgment*. Little, Brown Spark, 5/2021.
- [4] Tom Simonite: Google's AI Guru Wants Computers to Think More Like Brains, *Wired*, <https://www.wired.com/story/googles-ai-guru-computers-think-more-like-brains/>, 12/2018.
- [5] Hesse Jones: Geoff Hinton Dismissed the Need for Explainable AI: 8 Experts Explain Why He's Wrong, *Forbes*, <https://www.forbes.com/sites/cognitiveworld/2018/12/20/geoff-hinton-dismissed-the-need-for-explainable-ai-8-experts-explain-why-hes-wrong/>, 12/2018.