

La caja negra

La aplicación de la Inteligencia Artificial va a ser una realidad en los próximos años, aunque el problema estribará en el conocimiento de la toma de decisiones. La inteligencia artificial se basa en el concepto de redes neuronales, pero para la mayoría de las personas trabajará como una caja negra, donde son conocidos los inputs de entrada, se conocerán los outputs de salida, pero se desconoce el proceso por el que se ha producido una salida concreta. Sabemos que en la mayoría de los casos la respuesta es correcta, pero el conocimiento que se obtiene no queda en nuestras manos, sino que está codificado dentro de la propia red.

La idea básica de una red neuronal consiste en simular el comportamiento de un cerebro biológico mediante pequeñas unidades computacionales, las 'neuronas', dispuestas en capas conectadas mediante enlaces digitales, las 'sinapsis'. La primera capa de neuronas captura datos externos (y distribuye dicha representación de los datos a la siguiente capa a través de los lazos sinápticos, que se refuerzan o debilitan siguiendo un proceso definido por reglas matemáticas. Las representaciones se van transformando según las neuronas propagan la información hacia adelante y, eventualmente, la capa superior nos proporciona una respuesta. Este proceso, repetido miles de veces mediante un mecanismo de retroalimentación, permite a la red neuronal aprender las diferencias entre uno y otro para clasificar con mayor exactitud.

El problema es la existencia de falsos positivos y de falsos negativos, que desconocemos como se han generado, y que hay que repasar muchas veces para intentar corregir los problemas.

Para ilustrarlo intentaremos definir unos cuantos ejemplos, en algunos casos justificado por el sesgo de datos, y en otros por el comportamiento extraño del algoritmo.

Un ingeniero programador de sistemas de conducción autónoma gobernado por Inteligencia Artificial, estuvo probando el algoritmo en el laboratorio usando simulaciones. Cuando estuvo seguro se pasó a la fase del mundo real, pasando de caminos apenas transitados a carreteras públicas con mayor velocidad conforme aumentaba la robustez del sistema. El sistema había conseguido una conducción suave, segura, impecable, casi mejor que la de algunos humanos. El ingeniero confiado por la ausencia de incidentes decide transitar por nuevos paisajes, y cuando se acercó a un puente en la carretera, a la vera de un acantilado, nota un ligero desvío a la izquierda. El pensó “nada grave, algo que, seguro que el sistema corregirá pronto”, pero ese movimiento se convirtió en un giro brusco del volante, del cual pudo reaccionar a tiempo para frenar el coche en seco antes de despeñarse.

Al volver al laboratorio, el ingeniero, cuyo nombre es Dean Pomerleau, investigó las posibles causas del accidente y llegó a una conclusión sorprendente: al parecer la red neuronal que gobierna las actividades del sistema había aprendido a utilizar la hierba de los márgenes del camino como una guía para orientarse, y se había quedado completamente confundida cuando esta desapareció en la vereda del puente.

Este caso, se pudo determinar el mal funcionamiento debido al conocimiento del Sr. Dean, pero no será el caso de la mayoría de los usuarios.

Otro caso del uso erróneo del algoritmo fue el caso de Brisha Borden.

En 2014, en una población de Florida, USA, Brisha Borden llegaba tarde a recoger a su hermanastra del colegio cuando vio una bicicleta de un niño de 6 años, sin candado, y un scooter. Borden y una amiga cogieron la bicicleta y el scooter y se los llevaron. Pronto se dieron cuenta de que eran demasiado grandes para esos transportes y se dispusieron a dejarlos, pero la madre del niño las había visto llevárselos, además, un vecino había llamado a la policía. Las dos chicas fueron arrestadas y acusadas de robo y hurto menor por los dos artículos que estaban valorados en un total de 80 dólares.

Por otro lado, el verano anterior, Vernon Prater, de 41 años, fue detenido por robar herramientas por valor de 86,35 dólares en una tienda cercana. Prater era un criminal reincidente. Ya había sido condenado por robo e intento de robo a mano armada, por lo que había cumplido cinco años de prisión. Además, tenía otros antecedentes por delitos menores cometidos cuando era menor de edad.

Sin embargo, algo extraño cuando un algoritmo de Inteligencia artificial fue el encargado de predecir la probabilidad de que cada uno cometiera un delito en el futuro. Borden, por ser de raza negra, fue calificada como de alto riesgo y Prater, por ser de raza blanca, fue calificado como bajo riesgo

Dos años más tarde, una vez estudiado el caso, se dieron cuenta de que el algoritmo se había equivocado, Borden no había vuelto a ser acusada de un nuevo crimen. En cambio, Prater estaba cumpliendo una condena de ocho años de prisión por irrumpir en un almacén y robar miles de dólares en electrónica. El algoritmo se revisó.

En noviembre del 2019, la empresa Apple fue señalada por discriminación sexista tras descubrirse que su tarjeta bancaria Apple Card estaba dando ventajas en el crédito a los hombres frente a sus esposas, estando ambos en las mismas condiciones financieras. Apple respondió a las críticas argumentando que su compañía no tiene, ni ha tenido en ningún momento de su historia, comportamientos sexistas y todo lo achacó al “inexplicable algoritmo”

Este nivel de dependencia, en muchos casos, nos llevará a la paradoja que a alguien le negarán un préstamo hipotecario y cuando pregunte por qué, sólo sabrán decirle por el sistema lo ha determinado sin que la persona conozca la realidad de la toma de decisiones. La persona que utilice el sistema sabrá que en la mayoría de los casos previos la elección ha sido la correcta pero no conocerá la razón de la negación.

- En el caso médico, es bueno o malo el concepto de caja negra.
- Debería haber una supervisión periódica de los resultados y la corrección de los algoritmos.
- Cuando se puede considerar que un sistema está lo suficientemente validado para el entorno de decisión médica.
- Debería haber máxima transparencia en las empresas para definir la red neuronal utilizada
- ¿Debería regularse el uso de la Inteligencia Artificial?