

# The Limitations of Opaque Learning Machines

Centro de Investigación en Computación, IPN.

Marco Antonio Cardoso Moreno

Judea Pearl es profesor de ciencias de la computación y director del Laboratorio de sistemas cognitivos en la universidad UCLA. En este artículo comienza la discusión mencionando que él trabajó inicialmente con el enfoque de la cibernética, del cual destaca que, a pesar de no contar con todo el poder computacional de las máquinas de Turing, era un paradigma "transparente"; esta transparencia es un aspecto de gran importancia en los sistemas computacionales, y que con las nuevas tendencias en inteligencia artificial, como Deep Learning, se está perdiendo dada la naturaleza de estos algoritmos, cuyos procesos consisten básicamente en la aproximación de funciones.

Una consecuencia de esta falta de transparencia es que el usuario final, generalmente, no tiene mucha idea de por qué una red neuronal profunda entrega sus resultados; aspecto que no suele importar mucho cuando la red parece tener un buen desempeño, pero en el caso contrario, aquel de un desempeño pobre, no se cuenta con herramientas para entender el por qué del mal desempeño, ni de como plantear una posible solución. De aquí surge un debate entre dos posturas: si la transparencia es necesaria o si, por el otro lado, la opacidad que viene con ciertos sistemas no debe preocuparnos. Pearl prefiere la transparencia, y decide buscar obtenerla mediante el análisis causal. Con base en esto, ha descubierto que existen ciertas barreras en los paradigmas actuales de Machine Learning que de no ser superadas, no nos permitirán lograr el desarrollo de inteligencia a un nivel como el humano.

En concreto, una de las características que definen a la inteligencia humana es la capacidad de memoria, y de almacenar representaciones de nuestro entorno. Esta capacidad nos permite hacernos (y responder) preguntas del tipo "qué pasa si...?". En el caso de lo que hoy se conoce comercialmente como inteligencia artificial, esto está muy lejos de ser cierto para los modelos computacionales que, como ya mencioné, funcionan principalmente como aproximadores de funciones.

Finalmente plantea que la inferencia causal ha tenido logros en la algoritmización tanto del nivel jerarquico de intervención como del contrafactual.

Por mi parte, la frase con la que cierra el artículo (y que a continuación cito), resume perfectamente hacia donde estamos yendo con los paradigmas actuales de machine learning; considero que es necesario un replanteamiento sobre la dirección en la que queremos avanzar, y si ésta coincide con aquella en la que nos movemos actualmente.

"Opaque learning systems may get us to Babylon, but not to Athens."