The Seven Tools of Causal Inference

Centro de Investigación en Computación, IPN.

Marco Antonio Cardoso Moreno

El aprendizaje de máquina (/machine learning/) ha tenido un auge importante en los últimos años, que trajo consigo una alta expectativa sobre este tipo de sistemas y cuándo estos podran presentar un nivel de inteligencia parecido al de los seres humanos. Sin embargo, estos sistemas se enfrentan a ciertas problemáticas que les impide lograr este propósito: la ausencia de robustez o adaptabilidad, es decir que estos sistemas no tienen una capacidad de adaptarse a nuevas circunstancias para las cuales no hayan sido entrenadas para lidiar; y la poca explicabilidad que estos modelos ofrecen para poder interpretar y entender de dónde surgen sus resultados, por lo que la mayoría de dichos modelos siguen considerándose como *cajas negras*.

Además de los problemas ya planteados, Judea Pearl considera que un entendimiento de *causa-efecto* es un aspecto para fundamental si se pretende lograr una emulación de la inteligencia humana. A partir de este punto es que Pearl propone una jerarquía de tres niveles de inferencia causal, con la intención de formalizar esta parte del pensamiento humano. Los tres niveles de la inferencia causal de Pearl son los siguientes:

1. Asociación: intervienen las relaciones puramente estadísticas con los datos al desnudo.
   Preguntas características de este nivel son: ¿qué es?, ¿cómo cambiaría mi creencia en ver
   X?. En este nivel se trabaja con los datos al desnudo.
2. Intervención: este nivel implica un proceso más elaborado que el simplemente *ver* los
   datos, sino que implica cambiar lo que vemos o , en otra palabras, en este nivel se requiere
   de la participación del sujeto. Preguntas características de este nivel son:
   ¿qué pasa si?, ¿qué pasa si yo hago *X*?
3. Contrafactuales: es un modo de razonamiento que se remonta a la filosofía de David Hume
   y John Stuart Mill, y al que se le ha dado una semántica amigable con la

computadora en las
últimas décadas. Una pregunta típica en la categoría de contrafactual es:
"¿Y si hubiera actuado de manera diferente?" por lo tanto, requiere un
razonamiento
retrospectivo. Preguntas características de este nivel son: ¿por qué?, ¿fue *X* lo que
causó
Y?, ¿qué hubiera pasado si hubiera actuado diferente?

Pearl, además, dice que las preguntas del nivel de asociación pueden ser contestadas únicamente si se tiene información de los niveles superiores. Esta jerarquía de tres niveles,
así como sus restricciones, explica por qué los sistemas de aprendizaje de máquina basados
en asociaciones no son capaces de razonar sobre acciones, experimentos y explicaciones
causales.

Pearl propone siete herramientas para la inferencia causal, mediante las cuales se puede
lograr un razonamiento autónomo.

1. Codificar asunciones causales: Transparencia y comprobabilidad.
2. *Do-calculus* y el control de las confusiones.
3. Algoritmización de los contrafactuales.
4. Análisis de mediación y evaluación de efectos directos e indirectos.
5. Adaptabilidad, validez externa y sesgo de selección de la muestra.
6. Recuperación de datos faltantes.
7. Descubrimiento causal.

Un punto de énfasis es que los investigadores deben entender que los modelos utilizados
para llevar a cabo este tipo de tareas son conceptuales y, por lo tanto, no requieren compromiso con una forma particular de las distribuciones involucradas. Además, hay una
relación directa entre la validez de las inferencias realizadas y la veracidad de la estructura
asumida. Si la verdadera estructura difiere de la asumida, y los datos encajan igualmente
bien, pueden resultar errores sustanciales que a veces se pueden evaluar a través de un

análisis de
sensibilidad.

El razonamiento causal es uno de los componentes indispensables del pensamiento humano, que debe ser formalizado y algoritmizado para, así, lograr alcanzar la meta de producir inteligencia artificial al nivel de la inteligencia humana. Sin embargo, la manera de poder dar consecución a este objetivo es, únicamente, mediante una total comprensión del pensamiento humano, de tal modo que a partir de este conocimiento se pueda generar una teoría de la inteligencia humana que, a su vez, pueda ser algoritmizada.