

The Seven Tools of Causal Inference

Centro de Investigación en Computación, IPN

Laboratorio de Ciencias Cognitivas Computacionales

Miguel Angel Soto Hernandez

Gracias al éxito de la Inteligencia Artificial (IA), se han desarrollado una gran cantidad de aplicaciones, lo que conlleva a la creación de expectativas más grandes por parte de los usuarios. Sin embargo, la IA se encuentra con muchos problemas al tratar de llenar estas expectativas. Uno de los principales problemas es la adaptabilidad o robustez, ya que los investigadores señalan que estos sistemas de IA son incapaces de reconocer o reaccionar a nuevas circunstancias para las que no han sido entrenadas con anterioridad. Otro gran problema es la explicabilidad, ya que la mayoría no puede explicar el porqué de las decisiones que toma o las predicciones que hace. El último problema se refiere a la falta de comprensión de lo que la IA está realizando.

El artículo argumenta que el razonamiento causal es un componente indispensable del pensamiento humano que debe formalizarse y algoritmizarse para lograr la inteligencia artificial a nivel humano. Para lograr esto se explican los impedimentos para llegar a lograr esto en forma de jerarquía de tres niveles, y demuestran que la inferencia al nivel 2 y al nivel 3 requiere un modelo causal del entorno de uno. Estos tres niveles son los siguientes:

1. **Asociación:** en esta parte se invocan las relaciones puramente estadísticas con los datos al desnudo. En este nivel nos podemos hacer preguntas como: ¿Qué es? ¿Cómo cambiaría mi creencia en ver X?
2. **Intervención:** aquí no es simple y sencillamente ver los datos y ya, sino que implica no cambiar lo que vemos. Las preguntas típicas de este nivel son: ¿Qué pasa si? ¿Qué pasa si yo hago X?
3. **Contrafactuales:** es un modo de razonamiento que se remonta a los filósofos David Hume y John Stuart Mill y al que se le ha dado una semántica amigable con la computadora en las últimas dos décadas. Una pregunta típica en la categoría de contrafactual es: "¿Y si hubiera actuado de manera diferente?" por lo tanto, requiere

un razonamiento retrospectivo. Las preguntas que nos haremos en este nivel serían:
¿Por qué? ¿Fue X lo que causó Y? ¿Qué hubiera pasado si hubiera actuado diferente?

Esta jerarquía de tres niveles, y las restricciones formales que conlleva, explican por qué los sistemas de aprendizaje automático, basados solo en asociaciones, no pueden razonar sobre acciones, experimentos y explicaciones causales. Aunado a esto, se describen siete tareas cognitivas que requieren el uso de herramientas de estos dos niveles de inferencia y se demuestra de igual manera como es que estos se pueden utilizar en el marco de "Modelos Causales Estructurales" (SCM). Las tareas que se describen son:

1. Codificación de supuestos causales: Transparencia y comprobabilidad.
2. Do-cálculo y control de la confusión.
3. La algoritmotización de los contrafactuales.
4. Análisis de mediación y evaluación de efectos directos e indirectos.
5. Adaptabilidad, validez externa y sesgo de selección de muestras.
6. Recuperarse de los datos faltantes.
7. Descubrimiento causal.

Finalmente se recalca que es importante que los investigadores tomen nota de que los modelos utilizados para llevar a cabo estas tareas son estructurales (o conceptuales) y no requieren compromiso con una forma particular de las distribuciones involucradas. Por otro lado, la validez de todas las inferencias depende críticamente de 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. Así como también es importante que se tenga en cuenta que las limitaciones teóricas del aprendizaje automático sin modelos no se aplican a las tareas de predicción, diagnóstico y reconocimiento, donde las intervenciones y los contrafactuales asumen un papel secundario.