因果学习

2022年5月7日 9:01

[IBM Causal Inference 360 Toolkit](https://ci360.mybluemix.net/).

Released in 2019, the toolkit is the first of its kind to offer a comprehensive suite of methods, all under one unified API, that aids data scientists to apply and understand causal inference in their models

Causal models要干的事就是，如何在没有A/B测试的情况下理解因果关系

Weight models operate by using the covariates to balance between the treatment groups, such that the weighted distribution of covariates between groups is more similar. Once each individual has its own weight, we can also weight the outcomes in each group to obtain unbiased potential outcomes. Then we can compare (say, subtract) the estimations of potential outcomes to obtain a causal treatment effect, in the sense that we canceled the effect of the covariates (now that they are similarly distributed across groups) and left only with the effect of the treatment. A special most-common type of weight models is **propensity models**, which estimate the weights by first estimating the probability of being treated given the covariates (also called a propensity score). They have vast theoretical literature backing them up. For our use, since they estimate probabilities, they can be evaluated more rigorously as we' ll mention below.
Outcome models, the second family of causal models, predicts the potential outcomes directly from the covariates and treatment assignment. This allows them to perform individual-level prediction of potential outcome, while weight models can only predict in the sample level (e.g., average within the treatment groups). However, the individual-level prediction is less reliable since there are more assumptions needed for it to hold.
Doubly robust models are an additional type of causal models that combine both weight and outcome models. There are multiple ways to do so, hence there are multiple models. The premise of doubly robust models is that their estimation is, well, more robust. Theoretically, it is sufficient that either one of the weight or outcome models be unbiased for the composite estimate to be unbiased.

来自 <https://ci360.mybluemix.net/resources#guidance>

什么是因果关系
- 哲学家约翰·洛克(John Locke)(Locke，1975)："会产生任何想法的事务，不论是简单或复杂，我们都称为因，而被生产出来的，就称为果。
- 遵循“连续性或相关性的规律”(Regularity of Succession or Correlation)

保罗·拉扎斯菲尔德(Paul F.Lazarsfeld)(Lazarsfeld，1959)进一步提出了一个被广泛接受的因果关系定义，他将因果关系的判定描述为遵循三个标准:(1)两个变量之间因果关系必须有时间顺序关系(Temporal Order)，这意味着在时间序列上，原因必须在结果之前，如果A是原因，B是结果，那么，A必须发生在B之前；(2)两个变量之间必须在经验上具有相关性；(3)更为重要的是，两个变量之间观察到的因果关系不能够被第三个变量解释，即两个变量之间不能够是伪关系(Spurious)。
- 通过“反事实框架”(Counter factual Framework)来探索因果关系

密尔(Mill，1973)指出，如果一个人吃一盘特别的食物，随后死亡了，这意味着，这个人不会死亡，如果他不吃这盘特别的食物。这说明密尔在比较在同一个人的两种潜在结果，一个是死亡(吃一盘食物)，一个是不死亡(不吃一盘食物)，这样一个人可以推断吃一盘食物导致了死亡。
潜在结果模型，也叫Rubin Causal Model：在因果推理的理论模型中，潜在结果模型(Potential Outcomes Model)是其中最重要的理论模型之一，其核心是比较同一个研究对象(Unit)在接受干预(Treatment)和不接受干预(Control)时结果差异，认为这一结果差异就是接受干预相对于不接受干预的效果。
https://zhuanlan.zhihu.com/p/484963104

杨洲推荐的经典
http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf
模式识别和机器学习

休谟问题（Humean problem）：关于因果与归纳的反思
休谟指出：我们无从得知因果之间的关系，只能得知某些事物总是会连结在一起。因果关系是信念的产物。
因果推理获得的知识构成了人类生活所依赖的大部分知识，按照休谟的观点，支撑我们现实的一切必然事物都将失去其必然性；太阳不一定会绕着地球转，被火烧着也不一定会痛，宇宙未必有其背后必然规律。
在他的怀疑主义、不可知论下的世界中，知识的底层保障是信念 。
不过对我们影响不大， 因为休谟还説过“习惯是生活的君王”， 他告诫他亲爱的读者：尽管这一切都没有必然性，依照日常习惯，你们最好不要

把手伸进火里——。

休谟指出，我们对于因果的概念只不过是我们期待一件事物伴随另一件事物而来的想法罢了。“我们无从得知因果之间的关系，只能得知某些事物总是会连结在一起，而这些事物在过去的经验里又是从不曾分开过的。我们并不能看透连结这些事物背后的理性为何，我们只能观察到这些事物的本身，并且发现这些事物总是透过一种经常的连结而被我们在想象中归类。
我们之所以相信因果关系并非因为因果关系是自然的本质，而是因为我们所养成的心理习惯和人性所造成的。”因果联系之间其实存在一个信心或盼望的跳跃，比如我们认为明天必定会来到是基于经验中对昨天今天与明天的“知识”，不自觉的把明天作为今天的结果，这是盼望和信心，不是知识

休谟的看待世界的态度是静态的、割裂的。比如谈论因果关系，他先假定出两个毫不相关的事物A和B，这两者直接并没有关系，但由于人类长期观察它们之间的伴随，于是给它们添加上一种关系。这种关系是人为的，不具有必然性的
康德一开始就否认事物之间独立存在的。康德讲：对象被感官以现象的样子呈现给我们，是因为我们具有主观感性条件——空间与时间。空间、时间并不是外在的经验性概念，而是先天存在于我们认识结构中的表象。所以当我们认识现象时，是因为空间和时间要求事物以此样子呈现出来。而休谟所说A和B也不是完全独立的，它们必然的伴随因果关系存在。因为康德认为：事物和关系必然同时的出现。

Causality (also referred to as **causation**, or **cause and effect**) is influence by which one event, process, state, or object (*a cause*) contributes to the production of another event, process, state, or object (an *effect*) where the cause is partly responsible for the effect, and the effect is partly dependent on the cause. In general, a process has many causes,[1] which are also said to be *causal factors* for it, and all lie in its past. An effect can in turn be a cause of, or causal factor for, many other effects, which all lie in its future. Some writers have held that causality is metaphysically prior to notions of time and space.[2][3][4]

来自 <https://en.wikipedia.org/wiki/Causality>

辨证唯物主义因果律：即任意宇宙状态都是其之前宇宙状态积累的结果，任意运动状态均是其前运动状态积累的结果。即什么样的因，对应什么样的果，其具有最为广泛的普遍性。

支持更高水平智能的六大知识维度

来自 <https://mp.weixin.qq.com/s?__biz=MzA5ODEzMjIyMA==&mid=2247686184&idx=2&sn=cef7dae79174d1866171f3b920c0ebe1&chksm=909a11bba7ed98adc0ed600fddd3fc6ea9e0de8f2f64f3294e68e5fbb847a57aaaf0de3625e6&mpshare=1&scene=1&srcid=0524GKt9pZOSpOKRJGJnBldm&sharer_sharetime=1653384052612&sharer_shareid=836710033d48c8adf14fa9377f6b569e&exportkey=AUX963WzMjXUAp2j8oa8KNc%3D&acctmode=0&pass_ticket=N%2FTjwyRQSugko09jothVqvvciyzCh9c8B%2FQVb3dekFJ5oMLQi0ab8fuzd6UY6BHf&wx_header=0#rd>

知识维度中有两个维度反映了对世界的看法，一个是**描述性维度**，描述性维度对世界上存在的事物进行了概念性的抽象
另一个是**现实世界及其现象的动态模型**。
此外，**故事**提升了人类在共同信仰和神话基础上的理解和交流复杂故事的能力。
语境和来源归因以及价值和优先级是元知识维度，这些维度带来了基于条件的有效性和知识的不断叠加。
概念参考是结构基础，跨维度、模态和参考而存在。
这六个知识维度结合在一起，可以让人工智能不仅仅停留在事件相关性上，而是获得更深入的理解，因为这六个知识维度的潜在概念是持续的，可以解释和预测过去和未来的事件，甚至允许计划和干预，并考虑反事实的现实——因此文中使用了“深度知识（deep knowledge）”一词。

Towards causal representation learning

经典的因果学习
- 变量和结构已知，研究集中在如何在没有A/B测试的情况下理解因果关系
- 例如 IBM Causal Inference 360 Toolkit

因果表示学习
- 从观察数据中学习潜在的因果关系

共因原理：没有因果关系就没有相关
- 一事件要么是另一事件的原因，要么是它的一个结果，要么两事件共享同一原因

独立因果机制

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathbf{PA}_i).$$ 简化 $$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_{i+1}, \ldots, X_n),$$

https://zhuanlan.zhihu.com/p/355009051
https://zhuanlan.zhihu.com/p/355180872
【综述长文】因果关系是什么？结构因果模型入门 - 知乎 (zhihu.com)
《因果学习周刊》第2期：因果表征学习 - 知乎 (zhihu.com)