Causality (因果关系)

提出原因：

当前深度学习领域有以下两个问题：

① 模型解释性差。不论是(i) attention mechanism，(ii)用线性浅层模型来模拟深层模型，(iii) CNN的可视化 等均不能从本质上提高模型可解释性

② 深度学习永远在做"拟合"，只能解决"关联"问题，无法解决因果问题。

⇒ 若AI想从浅层AI跨进深层AI，就必须能理解因果关系

Judea Pearl 提出的认知世界模型：

Supervised Learning 1 ⇒
Unsupervised Learning 1.5 ⇒
Reinforcement Learning 2 ⇒

Level (Symbol)	Typical Activity	Typical Questions	Examples	
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?	关联问题
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?	干预问题
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection 想象与回顾	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?	反事实推理问题

Fig. 1. The Causal Hierarchy. Questions at level i can only be answered if information from level i or higher is available.

研究因果关系将对AI领域产生八大重要改变：

第一点是机器做的假设以人类容易理解的方式（因果图）呈现出来，从而让模型更加透明，也让测试模型的推论的后人能够更精准的去检验模型的鲁棒性。
第二点是通过因果推断，去除混杂因素的影响。有了因果推断，就不必人，来根据常识去掉那些可能影响相关性的混杂因素，从而在更复杂的坏境下，做到端对端的学习。
第三点是算法化的回答反事实的问题。如果机器能够做这样的思考，那AI思考的模块化程度就会进一步提高，需要的训练数据也会减少，对于跨领域的迁移学习也会有所助力。
第四个助力是区分直接和间接的诱因。区分了直接的诱因与通过第三方作用间接的影响，就能够判定数据中那些异常点处在间接影响的链条上，受到未知因素的影响，属于噪音，而对于处于直接因果链条上的，则异常不应该被视作是噪音，而是可以证伪模型的“黑天鹅”。
第五个助力是模型具有跨领域的适用性，能够通过其他领域来验证该模型的鲁棒性。如果智能体是通过因果推理，来决定下一回合的policy，那这个思考过程就更像人类做决定时的所思所想，由此类比推出，智能体也会具有更好的domain adaptation。
第六个助力避免sampling bias。人类在对机器建模时，会展现出认知偏见。如果机器具有了公理化的因果推理，那通过反事实的问题，就可以指出人类可能受到了采样偏见的影响。这指出了人机协作的新的可能性
第七个助力是通过因果模型，来判定数据集中是否存在数据缺失的问题。
第八个助力是去发现因果关系。现实中存在着诸多类似孟德尔随机的自然形成的与随机双盲实验等价的场景，通过让模型具有因果推断能力，就能够发现未知的因果关系。