

doi:10.13582/j.cnki.1672-7835.2021.02.006

因果推理研究的认知进路： 基于人工智能视角

黄闪闪^{1,2}

(1.天津理工大学 马克思主义学院,天津 300384;2.南京大学 博士后流动站,江苏 南京 210023)

摘要: 因果关系是人工智能实现从感知到认知需突破的关键问题,因果推理对于构建可解释的机器学习模型至关重要。用概率解释因果关系是一种自然主义观点。因果贝叶斯网络的运用表明,发现因果关系的人工智能归纳法面临因果马尔可夫条件的客观性解释难题。为保证归纳的因果模型符合我们理解的因果关系,因果认知理论体现了一种认知转向的非实在论立场。因果推理的认知进路昭示着概率理论从单一向多样化发展,主观与客观从对立到融合的可能性。

关键词: 因果关系的概率解释;因果马尔可夫条件;认知进路

中图分类号: B814, N031 **文献标志码:** A **文章编号:** 1672-7835(2021)02-0044-07

因果性是哲学中的传统问题,但在人工智能领域并未引起足够重视,还是一个较新的领域。人工智能的核心基础是机器学习,其本质是一个通过假设导出可泛化结果的归纳学习过程。早期的人工智能是基于符号逻辑的演绎推理,1980年以后的人工智能算法则是基于概率(贝叶斯网络)的归纳推理。因果推理^①和人工智能之间的交集体现在:机器学习问题实质上是因果性研究中的发现问题,基于假设模型来桥接因果关系和数据;因果思维解决了迁移学习等问题,让人工智能能够回答为什么的问题,是实现感知智能到认知智能的重点工作之一。

为了让人工智能更值得被信任,因果推理是充要条件。因果推理结合了演绎逻辑和归纳逻辑的推理模式,其研究倾向于计算机编程、人工智能模型、概率等领域。随着对大数据、云计算以及基于语境的机器学习研究的进一步深入,收敛交叉映射算法(convergent cross mapping)和压缩感知(compressed sensing)等新方法的涌现,因果性理论及应用空间不断扩大,其研究内容也不断深入,

带动了整个科学认知的发展和进步。在人工智能领域,因果模型提供了数据分布特性的简洁描述,已经证明因果知识可以促进各种机器学习任务,包括半监督学习和转移学习。因果贝叶斯网络提供数据中重要关系结构的图形化、易于解释的表示形式。在现代归纳逻辑的视域中,以因果关系和概率为中心,讨论因果推理的本体论解释和认识论解释,探索因果推理研究的认知进路。

基于逻辑视角,我们在此讨论因果关系的概率框架和人工智能中的因果模型。目前贝叶斯网络的推理算法仍然是研究的热点。按照复杂程度,因果模型主要包括因果图、因果贝叶斯网络(CBN)和因果结构模型(SCM)。以因果贝叶斯网络为代表的人工智能(AI)归纳法面临因果马尔可夫条件的因果和概率解释难题。现代归纳主义认为,基于实验和观察数据的算法进行假设,适当的数据会产生正确的因果关系。通过归纳程序找到某个“最佳”的因果模型,其概率独立性由因果马尔可夫条件蕴含,并与从数据推断而来的独立性是一致的。但是,因果马尔可夫条件在许多

收稿日期:2020-09-20

基金项目:教育部人文社科基金青年项目(18YJC72040001);中国博士后科学基金面上项目(2019M651798)

作者简介:黄闪闪(1986—),女,湖北黄石人,博士、副教授,南京大学博士后流动站研究人员,主要从事归纳逻辑与科学方法论研究。

①这里理解的因果推理基于条件性因果关系,主要探讨这组关系成立的相应条件。

归纳法中不成立。例如,约束归纳(constraint-based induction),贝叶斯归纳(Bayesian induction),信息理论化归纳(information-theoretic induction),以及谢弗(Shafer)的因果猜想等。

接着我们从哲学上介绍因果关系的几种本体论解释和认知理论。传统哲学方法从本体论出发,试图把因果关系还原成其他概念,根据非因果性的概念来解释因果关系,坚持一种实在论立场。因果关系的机械解释把因果关系还原成物理过程;概率解释把因果关系还原成物理的概率关系;反事实解释把因果关系还原成反事实法则;主体解释把因果关系还原成主体获取目标的能力。基于认识论看待因果性问题,则强调证据种类和方法用于推理因果关系。威廉姆森(Williamson)的客观贝叶斯主义,提出介于心理解释和物理解释之间的因果关系。毕比(Helen Beebe)的休谟式投射主义(Humean projectivism),强调推理对因果“类律”关系的重要性。瑞斯(Julian Reiss)的推理主义(inferentialism),指出科学共同体在局部或特定语境中做出因果断言。斯庞(Wolfgang Spohn)的排序函数理论本质上是一种认知概率,它表明因果关系是认知主体所相信的东西,与推理相关,这是认识论的核心。

因果推理的认知进阶不尝试将因果关系还原成主体的经验,因果关系不是作为本体论的一个元素,而是作为描述世界的一个范畴,以此做出成功的推理和解释。认知理论支持一种反实在论的因果关系,表明因果关系来自科学实践,基于观察、干预等行为做出科学推理。癌症流行病中结合频率主义和客观贝叶斯主义解释的因果模型,以及因果推理在P-log程序等人工智能领域中的运用,昭示着因果推理的认知发展趋势可能有:一是概率理论种类从单一化向多样化发展,二是主观与客观从对立到融合。

一 因果关系的概念框架:脉络与评价

鉴于统计工具在科学中的重要性,用概率解释因果关系是一个科学因果关系的自然起点。概率因果关系的代表人物主要有赖欣巴哈(Hans Reichenbach),古德(I. J. Good),萨普斯(Patrick Suppes)和埃尔斯(Ellery Eells)。

逻辑经验主义者赖欣巴哈受康德先验论和爱因斯坦时空相对论的影响,1951年在《科学哲学

的兴起》中提出因果关系概念的重要观点。赖欣巴哈认为,概率命题都是关于经验世界的综合命题,不能证明为真。他把先验概率原则添加到康德的先验因果关系原则中。一个关于概率和充分因果原则的命题是一条关于经验知识的必要先验原则。赖欣巴哈把他的分析扩展到物理测量中的误差概率。因为所有的物理测度都可能会有误差,只有误差出现在概率分布上时,自然法则知识才是可能的。这是一个不能被经验证明的综合命题。所以经验命题要求把个体事件的先验概率原则和普遍规律结合起来。

古德在20世纪60年代早期用物理概率来定量解释概率因果关系,并提出一个不确定性的因果关系概念。因为原因不总是产生结果,同样的原因在某些情况下可能产生不同的结果,一个结果在不同情况下可以通过不同原因产生。古德想通过概率论的形式化获得这个概念,使它能定量获得。古德对因果关系进行了分类,并为其提供不同的概率测度。一种是谈论一般术语,即一个事件引起另一个事件的倾向。例如抽烟增加总体中罹患肺癌的机遇。另一种是谈论具体因果,即一个事件实际引起另一个事件的程度。比如,张三抽烟使得他更容易罹患肺癌。

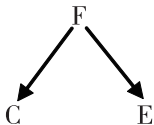
萨普斯在1970年出版的《概率因果论》中,解释了因果关系的概念,即原因影响结果的概率。根据他的观点,如果像Y这样的事实跟随像X这样的事实出现很高的概率,并且如果没有第三个事实对X和Y之间的概率关系负责,那么X称为Y的原因。同时,他定义了表面正因和表面负因,既涵盖原因使得结果更有可能出现的直观情形,又包括原因使结果不大可能的负面原因。所谓表面正因,就是给定原因的结果概率大于这个结果本身的概率。而表面负因则是给定原因的结果概率小于这个结果本身的概率。在科学实践中,我们很可能更关注负因。比如在医学中,我们更专注降低疗效概率的原因。这样萨普斯提供了一个更广泛的概念,不仅仅局限于概率上升,同时包括整个概率变化,它可能是正向或负向的。

萨普斯还用幕隔(screen off)的概念来处理原因的表面化问题。幕隔最早由赖欣巴哈提出,他分析了时间箭头,并试图通过物理现象的因果关系来确定时间的方向。萨普斯对事件之间的幕隔进行了定义。事件C是事件E的表面原因,那么有个更早的事件F幕隔了C和E。具体而言,F

早于 C, C 早于 E; C 和 F 同时发生的概率大于 0; 给定 F 发生的 C 的概率大于 0, 虽然 C 和 E 是相关的, 但是这种相关性会消失, 如果我们引入因素 F。采用图形将上述关系形式化:

$$F \longrightarrow C \longrightarrow E$$

还有一种幕隔的典型情况, F 是 C 和 E 的共因, 即 F 是 C 和 E 的共因。



确定正确的因果结构正是物理学、社会科学、医学等因果建模者不断努力解决的问题, 其中寻找幕隔是非常重要的。例如手指发黄与肺癌有关, 所以手指发黄是肺癌的表面正因。但是医学知识告诉我们肺癌的真正原因一定是其他原因, 例如抽香烟。抽香烟完全解释了文献中手指发黄和肺癌之间的相互关系。我们说抽香烟幕隔手指发黄和肺癌。

埃尔斯对应古德的因果关系分类, 深入讨论类型因果关系和标志因果关系 (token causation) 的概念。前一个概念关于总体水平, 后一个概念关乎个体水平。埃尔斯用形式化方法定义类型-水平因果关系的概率提高:

$$P(E | C \wedge K_i) > P(E | \neg C \wedge K_i)$$

这里 C 和 E 表示原因和结果, K_i 表示背景语境。例如抽烟导致肺癌, 如果它提高了肺癌的概率, 鉴于所有相关背景、可能的原因、混杂因素等。它明显关于一个总体或类型。这是一个相当强的要求, 因为我们很少能够控制假定原因周围的整个环境。但是这个模型包含了因果关系的边界和可能条件^①。在此基础上, 可以对标志概率提高进行推理, 通过确定标志事件概率可能变化的关键点, 并绘制一个概率轨迹。

因果关系的概率框架主要有两个局限。一是概率方法还不足以解释复杂的因果关系, 而这些复杂的因果关系并不能仅仅用事件之间的概率关系来形式化。二是时间是预先设定的, 而不是正确分析的。这种预设的问题在于, 我们很多时候正好不知道正确的时间顺序。但是, 概率推理对科学中的因果推理非常重要, 它把因果关系从确定性中解放出来, 并将其与基于概率和统计的实

际科学实践联系起来。正因和负因, 幕隔和因果关系水平的概念^②, 有利于科学中的方法论实践, 以便能够有效地利用因果关系进行预测和控制。

二 贝叶斯网络中的因果马尔可夫条件: 解释恰当性问题

现代归纳主义认为, 基于实验和观察数据的算法进行假设, 适当的数据会产生正确的因果关系。珀尔 (Pearl) 20 世纪 80 年代引进贝叶斯网络作为表征工具, 使人们能够将概率信念组合成一个一致的整体。虽然其中一些信念可以从贝叶斯网络中直接读出, 但是许多都是由这种表征暗含的, 并且需要明确的计算工作。计算和明确这种信念就是贝叶斯网络中的推理问题。目前贝叶斯网络的推理算法仍然是研究的热点。按照复杂程度, 因果模型主要包括因果图、因果贝叶斯网络 (CBN) 和因果结构模型 (SCM)。

(一) 因果模型

最简单的因果模式是因果图, 它用点和箭头组成的图表表示几个变量之间的因果关系, 即用一个潜在的有向无环图 (DAG) 来解释因果关系。因果图只显示了定性因果关系。围绕因果图可以构建复杂一些的贝叶斯网络, 用图形表征感兴趣的变量之间的因果关系。建构一个 V 上的贝叶斯网络, 一是通过包含一个 A 到 B 的箭头建构一个有向无环图 C, 当且仅当 A 是 B 的直接原因; 二是依据确定 S 中的对应概率表。贝叶斯网络 (C, S) 中的图形给出一个因果解释, 所以这个网络称为因果贝叶斯网络 (CBN), 简称为因果网。符号 C 表示存在因果解释。因果结构模型 (SCM) 是对贝叶斯网络的一种泛化。它结合了结构方程 (SEM)、虚拟事实模型 (RCM)、概率图模型 (主要是贝叶斯网络), 并将其应用于因果分析。各类常用因果模型, 都可以看作 SCM 的子类。这些方程表达其中每个变量同它的直接原因变量之间的关系, 写作 $A_i = f_i(Par_i, E_i)$, 其中 f_i 是一个函数, E_i 是一个误差变量, 每个误差变量都在概率上独立。

(二) 贝叶斯归纳

导出因果关系的贝叶斯方法由 Cooper, Heck-

^①Gillies, D. and Sudbury, A. "Should causal models always be Markovian? The case of multi-causal forks in medicine". *European Journal for Philosophy of Science*, 2013, 3(3): 275-308.

^②Illari, P. and Russo, F. *Causality: Philosophy Theory meets Scientific Practice*. Oxford: Oxford University Press, 2014, p.81.

erman, Herskovitz 和 Meek 提出。其基本思想是,归纳最大化后验概率的因果图 C:

$$p(C|D) = \frac{p(C)p(D|C)}{\sum_c p(C')p(D|C')}$$

这里的 D 是已观察过去案例数据的数据库。现在

$$p(D|C) = \int p(D|C, S_c)p(S_c) dS_c$$

具有贝叶斯网络中伴随 C 的概率规范 S_c 上的积分。(这个方法要求 p 不仅在 V 中的变量上定义,而且还在因果图、概率规范和数据库上定义。)假定因果马尔可夫条件, C 和 S_c 来自一个可以计算 $p(D|C, S_c)$ 的贝叶斯网络。为进一步计算提供帮助,假定概率说明符本身是概率上独立的,它们的先验分布采取狄利克雷分布的形式。

(三) 因果马尔可夫条件的解释问题

围绕因果图构建贝叶斯网络的策略,最早由珀尔 20 世纪 80 年代提出,已经广泛应用在概率专家系统的构造中。珀尔的 CBN 概念在三个方面都是开创性的。一是,他提出一个框架,这里的不确定性因果关系是随机变量之间的初始关系;二是,他表明相关和经典独立性关系如何以自然方式从这些因果关系中产生;三是,他表明这种产生如何忠于我们关于因果关系和纯粹相关之间差异性的直觉。

这种构建贝叶斯网络的方式,要求因果关系的先验知识,并取决于三个关于因果关系本性的关键假设。一是,假定直接因果关系的概念是变量之间的关系;二是,假定 V 上的因果图将是无环的;三是,假定 (C, S) 是个贝叶斯网络,换言之,假定因果图将满足在研的概率函数的马尔可夫条件。即每个变量是在概率上独立于它的无效性,以它的直接原因为条件^①。

发现因果关系的 AI 归纳法面临因果马尔可夫条件的难题。除了 CBN,因果网包含作为一个内在假设的因果马尔可夫条件。而在结构方程模型中,因果马尔可夫条件是每个变量表示为其直接原因和误差变量的函数的后承,鉴于进一步的假设,即所有误差变量都是概率独立的。但是因果马尔可夫条件在常见的因果关系和概率解释中无效。如果这个条件失败了,归纳法将只会假定错误的因果关系。具体而言,其真假可能取决于

人们怎么解释概率和因果关系。

因果马尔可夫条件不能解释为纯粹客观的。为了从统计关系中收集因果关系,发现因果关系的 AI 归纳法通常假定因果马尔可夫条件持有物理因果关系和物理概率。换言之,因果马尔可夫条件被解释为物理因果关系和物理概率之间的关系。在这种客观解释中,这个条件与共同原因原则密切相关,它诉诸共同原因原则背后的直觉。共同原因原则断定任何概率依赖必须有一个因果说明。如果两个变量是概率依赖的,那么一个变量引起另一个变量,或者它们是幕隔摆脱依赖的共同原因的结果。萨普斯论证,因果关系导出概率依赖性,并且幕隔用于区分共同结果变量和因果变量。然而,在变量是概率依赖的任何情况中,如果这种依赖性并不是由变量之间的因果关系来解释的,都是一个反例。概率依赖实际上不仅是通过因果关系产生的,而且是偶然的,或者因为变量是通过意义、逻辑关系、数学关系联系起来的,由于它们是由非因果的物理定律联系起来的,或者因为它们受到局部定律或边界条件的约束。因此,为这个条件寻求恰当的因果解释,才能保证归纳的因果模型符合我们理解的因果关系。

三 因果性研究的认知进路:解释和推理

因果关系的概率定义通常是形式的。为了更好地运用概率的形式概念,我们必须知道应该如何解释概率。传统哲学方法从本体论出发,试图把因果关系还原成其他概念,根据另一个非因果性的概念来解释因果关系,坚持一种实在论立场。纯粹的物理或客观解释不能保证因果马尔可夫条件的有效性。从认识论出发看待因果关系,不在于试图解释因果概念,而是基于科学实践对世界的推断和解释,体现一种认知转向的非实在论立场。

(一) 本体论的因果解释

许多概率解释把因果关系还原成物理的概率关系。除此之外,机械论把因果关系还原成物理过程。反事实解释把因果关系还原成反事实法则。主体解释把因果关系还原成主体通过影响他们原来来获取目标的能力。

因果关系的机械解释旨在理解连接原因和结

^①这里无效性和一个变量的直接原因由因果图 C 决定。

果的物理过程,把因果陈述解释为这类过程的某种陈述。例如道尔(Phi Dowe)认为一个因果过程是一个从头到尾占有一个守恒量的过程。这种方法的主要局限在于它的适用范围非常狭窄,因为大多数因果断言显然和物理学中的守恒量无关。

反事实解释把因果关系还原成虚拟条件句或反事实条件句: E 因果地依赖 C, 当且仅当(1)假如 C 发生, 那么 E 将会发生(或者它发生的机遇将会显著提高), 以及(2)假如 C 不发生, 那么 E 将不会发生(或者它发生的机遇将会显著降低)。那么因果关系被看作因果依赖的传递闭包: C 引起 E, 如果 E 因果地依赖 C, 或者如果 E 因果地依赖某个 D, 并且 C 引起 D。反事实条件句被赋予了可能世界的语义。这种解释的困难在于: 一是刘易斯(David Lewis)对可能世界的依赖; 二是相似性的测度问题。如果遵循刘易斯通过相似性来测度亲密性的建议, 是否存在一种客观方法来确定哪些世界最接近我们自己的世界, 这是值得怀疑的, 在这些方面的选择和权衡是一个主观问题。

主体解释根据主体通过控制原因以实现目标的能力来分析因果关系, 即 C 引起 E, 当且仅当, 引起 C 是主体引起 E 的有效方式。这里引起 C 的策略被认为是有效的, 如果一个理性决策理论将其规定为引起 E 的方式。这种解释的主要支持者有普赖斯(Huw Price)和孟席斯(Peter Menzies), 但是他们在这些概率的解释上不一致, 孟席斯坚称它们是机遇, 而普赖斯似乎有个贝叶斯概念^①。所以我们不清楚他们把因果关系看成一个物理概念, 还是一个心理概念。困扰主体方法的主要问题从概率和反事实方法所面临的问题中继承而来。一是, 主体方法假定主体概率的因果依赖不总是有效的; 二是, 主体解释诉诸虚拟条件句, 反事实解释的效用的顾虑同样适用于主体方法。

(二) 认识论的因果解释

基于认识论看待因果性问题, 强调证据种类和方法用于推理因果关系。因果关系不是作为本体论的一个元素, 而是作为描述世界的一个范畴, 以此做出成功的推理和解释。这有点像某些主体理论的变体, 但这里的因果关系关乎更广泛的推

断和推理, 而不是仅仅关乎我们的行动和高议。

威廉姆森(Williamson)认为因果关系介于完全主观心理解释和物理解释之间, 所以提出了介于主观主义和物理概率之间的客观贝叶斯主义解释^②。概率是一个主体的合理信念度, 即精神实体; 信念度被固定当作主体的背景知识函数, 即客观的。威廉姆森认为, 主要有心理与物理、主观与客观两种概率解释之分。心理概率解释是指概率是心理的, 即认识论的或者个人主义的, 此时概率被解释为主体心理状态的特征; 否则就是物理的。所谓主观概率解释, 就是指概率是主观的或者主体相关的, 如果具有相同背景的两个主体在某个概率值上意见分歧, 但二者都不是错的; 否则就是客观的。为了保证因果马尔可夫条件有效, 一是, 因果关系是心理而非物理的; 二是, 因果关系是客观而非主观的。不同贝叶斯主义的区别主要体现在逻辑哲学, 具体集中在概率解释上。逻辑主义把概率解释为命题之间的客观关系。主观主义把概率解释成遵守连贯性约束的信念度。威廉姆森的客观主义则将概率解释为心理实体之间的客观关系。需指出, 威廉姆森的客观贝叶斯主义根源于拉姆齐(Ramsey)和菲耐蒂(De Finetti)的主观主义^③, 有些学者将其看成一种主观解释。但是威廉姆森强调主观贝叶斯主义中信念度的“连贯性”对概率公理的遵守, 对理性来说是充分必要的; 而客观贝叶斯主义则认为只是必要的, 还需要应用最大熵原则。客观主义摆脱了杰恩斯(Edwin Jaynes)所谓的心灵投射谬误(the mind projection fallacy)^④。威廉姆森写了很多关于因果信念的文章, 并指出很多因果理论对因果关系的认知方面关注不足。从某种意义上说, 客观贝叶斯主义解释是一种认知概率解释。

毕比(Helen Beebe)的休谟式投射主义(Humean projectivism)是对休谟的现代解释, 沿着所谓的“投射主义”路线。根据投射主义解释, 休谟坚持我们的因果思维和谈话是我们推理习惯的一种表达。通过观察 a, 我们推断 b 将出现, 我们把推理“投射”到这个世界上^⑤。简而言之, 我们看

^①Menzies, P. and Price, H. "Causation as a secondary quality", *British Journal for the Philosophy of Science*, 1993(44): 197-203.

^②Williamson, Jon. *Bayesian Nets and Causality*. New York: Oxford University Press, 2005, p.65.

^③Darrell P. R. "On the Proximity of the Logical and 'Objective Bayesian' Interpretations of Probability". *Erkenn*, 2008(69): 335-349.

^④Jaynes, E.T. *Probability theory: the logic of science*. Cambridge: Cambridge University Press, 2003, p.22.

^⑤Beebe, H. "Hume on causation; The projectivist interpretation". In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited* (ed. H. Price and R.Corry), Oxford: Clarendon Press, 2007, p.225.

到一个原因 C,我们推断结果 E 将很快出现,并做出 C 是 E 的一个原因的推理。我们把这个推理“投射”到世界上,认为 C 和 E 本身是因果联系的。这种现代新休谟主义要求原因和结果之间的“类律”关系,即便我们知道这种关系意味着什么,也很难保证它在特定情况下是正确的,从而解决认识论问题。毕比的观点之所以是认知的,在于她强调推理对我们因果归因的重要性。她认为我们通过投射“必须”(must)做出更成功的推论,比那些只会预测的主体更成功,因为追踪这种必需的感觉要容易很多。“能够思考或说出黑球必须移动,与能够思考或说出黑球将移动之间有重要的区别吗?我认为是的。因为,说出或思考黑球必须移动时,我们设想自己有很好的理由认为黑球将会移动。”^①毕比的投射的一个重要部分是解释休谟,她表明把“必须”的感觉投射到世界,基于我们观察的规律性。

瑞斯(Julian Reiss)的推理主义(inferentialism)反对因果多元论,他从一个普遍理论中接受了推理主义的概念,这个理论认为语句意义由其推理联系构成。瑞斯没有定义一个普遍的推理主义,只是指出推理和因果关系有关。“我对因果关系的推理主义的理解不是作为一个成熟的因果关系理论,而是作为一个理论框架,在这个框架中,人们可以用哲学上富有成果的方式来处理与因果关系有关的各种问题。”^②对于瑞斯来说,相关推理没有外在的总体标准。一个语句与一个因果断言是推理联系的,当且仅当提出断言的科学共同体有权从因果断言推断它,或者从它推断因果断言(连同这个共同体持有的其他语句)。需要对这个描述说明三点:有权做出推断的科学共同体的本质;有权的本质;以及做出的推理的本质。“有权”这个标准一定是局部和有语境的,不仅仅因为它们刻画推理的指定科学领域所特有的,还因为它们对调查目的是特定的^③。根据这种观点,不同共同体之间的分歧可能无法解决,由语境

驱动。“因果关系根本上是主观的,因为证明‘原因’或同源词在断言的出现,是断言以某种方式与其他类型的断言在推理上相联系,而这种推理联系在人类的推理实践中起着作用。这就是因果关系部分客观的原因,这些推理实践不可能是任意的。(原文如此)而是由客观事实构成的。”^④

斯庞(Wolfgang Spohn)的排序函数理论^⑤是根本上与因果关系相关的一种认知概率,排序函数的思想非常类似于概率提高,使因果关系有些与认知有关的。排序函数提供一个动态信念理论。排序函数行为在某种方式上类似于概率测度,所以斯庞认为,我们也能把排序函数看作分析确定性因果关系的一种尝试,就像概率理论分析概率因果关系一样。具体而言,C 是 E 的一个原因,当且仅当 C 和 E 同时出现,C 先于 E,鉴于获得环境,C 提高了 E 的形而上学或认识论状态。斯庞要求原因总是在时间上先于结果。“形而上学或认识论状态”故意在因果关系的某些客观特征与某些认识特征之间含糊其词,例如客观概率和主观概率、斯庞的排序函数信念度之间。斯庞认为我们应该尝试主观转向,但是这个理论的研究是通过排序函数完成,而不是对概率的特殊解释。在主体因果关系的约束上,他通过对命题的排序函数中附加信念来做到这点,即假定具有某些客观的真值条件,这基于函数对象的实现。斯庞认识主体因果关系的主观相对论在某种意义上是一种休谟式的举动,因为因果必然性的本性在世界上是找不到的,而在于主体如何将感觉数据的概念联系起来。因此,因果关系是认知主体所相信的东西,与推理相关,这是认识论的核心。

四 因果概率解释的发展:多样性和融合可能性

因果推理的认知进路不打算将因果关系还原成主体的经验,它是一种反实在论的因果关系。因果性研究涉及一个更基本的哲学问题,即因果

^①Beebe, H. "Hume on causation: The projectivist interpretation". In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited* (ed. H. Price and R. Corry), Oxford: Clarendon Press, 2007, pp.245-246.

^②Reiss, J. "Causation in the sciences: an inferentialist account". *Studies in History and Philosophy of Biological and Biomedical Sciences*, 2012a(43):770.

^③Reiss, J. "Causation in the sciences: an inferentialist account". *Studies in History and Philosophy of Biological and Biomedical Sciences*, 2012a(43):771-772.

^④Reiss, J. "Causation in the sciences: an inferentialist account". *Studies in History and Philosophy of Biological and Biomedical Sciences*, 2012a(43):775.

^⑤Spohn, W. *The Laws of Belief. Ranking Theory and its Philosophy Applications*. Oxford: Oxford University Press, 2012, p.95.

性是一种客观实在还是一种认知关系^①。科学共同体通过观察、干预等行动,基于因果信念和证据做出科学推理。瑞斯和威廉姆森都明确允许因果信念的建立是动态的,而不是静态的,随着时间的推移,随着额外证据的不断收集而逐渐修正。因果关系作为一个合理连贯的概念是对人类心智的一种建构。威廉姆森在早期研究中明确使用了“心智依赖”的表达。对于推理主义来说,推理基础和目标共同作用来决定因果关系的意义,在因果关系的认知解释下,推理基础和目标的作用是严格分离的。正如我们将看到的,推理目标用于提供一个独立的规范性成功标准,在这个标准中,因果断言建立在推理基础上^②。

因果关系的认知进路与科学实践相关,昭示着概率理论种类从单一化向多样化发展。例如,在关于癌症的因果模型中解释概率。威廉姆森等人认为,相互竞争的解释是没有对错的,而是更适合或更不适合特定语境,以及我们对它们的要求。

癌症流行病学中因果模型的概率应该给予一个频率主义附带客观贝叶斯解释^③。他将频率主义和贝叶斯主义结合起来。这种选择的主要原因在于,需要处理两种不同类型的概率推理,群体层面和个体层面。

因果推理在人工智能等认知科学领域的应用,表明主观与客观从对立到融合的可能性。例如P-log中的因果推理^④。P-log程序的目标是创造一种知识表征(KR)语言,允许包含逻辑和概率的常识性知识的自然和容变(elaboration tolerance)表征。P-log的逻辑框架是回答集逻辑语言(ASP),ASP源于声明式语言,标准逻辑语言的语法和语义,析取数据库,以及非单调逻辑。根据容变的立场,为了系统合并新知识,P-log概率基础的自然选择是因果贝叶斯网络,其概率基础的自然选择是主观主义解释。一个P-log程序的ASP部分可以用于描述可能信念,而概率部分允许知识工程师量化这些信念的程度。

The Cognitive Approach to Causal Reasoning: From the Perspective of Artificial Intelligence

HUANG Shan-shan^{1,2}

(1. School of Marxism Studies, Tianjin University of Technology, Tianjin300384, China;

2. Postdoctoral Research Station, Nanjing University, Nanjing 210023, China)

Abstract: Causality is a key breakthrough in the realization of artificial intelligence from perception to cognition, and causal reasoning is of great importance for constructing interpretable machine learning models. Probabilistic causation is widely accepted as the most natural explanation of causal relations. But its application in causal Bayesian networks yields challenges of the Markov hypothesis, which often produces causal relations incoherent to our practical and metaphysical knowledge. To ensure that inductive causal models conform to our understanding of causality, the proposed cognitive approach to causality that is particularly beneficial by adopting a non-realist stance, we have liberated ourselves from the need of ontological justification. The idea is that by replacing probabilistic causal explanation with cognitive one, we would unify the multitude of causal explanations, and eventually unify subjectivity and objectivity in philosophy.

Key words: probabilistic causal explanation; causal Markov condition; cognitive approach

(责任校对 朱正余)

①顿新国:《因果理论的概率论进路及其问题》,《哲学研究》2012年7期。

②Williamson, Jon. "How can causal explanations explain?". *Erkenntnis*, 2013, 78(s2): 257-275.

③Williamson, Jon and Russo, F. "Interpreting probability in causal models for cancer." In *Causality and Probability in the Sciences* (Texts in Philosophy) (v5). (ed. Russo, F. and Williamson, Jon), London: College Publications, London, 2007, p.218.

④Goldszmidt, M. "Causal and Probabilistic Reasoning in P-Log". In *Heuristics, Probability and Causality: A Tribute to Judea Pearl* (ed. Dechter, R. and Geffner, H. et al). College Publications, 2010, p.337.