

# 因果发现与归纳问题

## Causal Discovery and The Problem of Induction

杨仁杰 /YANG Renjie

(首都师范大学政法学院, 北京, 100048)  
(College of Political Science and Law, Capital Normal University, Beijing, 100048)

**摘要:**归纳问题是科学哲学与认识论中的核心问题之一。它的目的是寻找从已知预测未知、从有限经验推广到一般规律的归纳推断的合理性的基础。自从休谟提出归纳怀疑论的现代表述以来, 尽管人们尝试了许多方式试图对其进行回应, 但它们都没有得到普遍接受。本文论证的观点是, 在机器学习领域所通行的验证因果发现算法的推理模式提供了一种回应归纳怀疑论的特殊方法。这种方法不但具有与传统的科学哲学确证理论不同的特征, 并且还能够揭示出归纳推断问题的更为精细的结构。

**关键词:** 因果发现 归纳问题 机器学习 逻辑可靠性理论

**Abstract:** The problem of induction is a central issue in both philosophy of science and epistemology. It is about the validation of inductive inference from observed data to unobserved conclusion. Since Hume raised the problem in its modern form, there were many attempts to respond but none is widely accepted. This paper focuses on the problem of the machine learning discovery of causal relations among a finite set of variables from purely observed data. It is argued that the way machine learning researchers validate automatic causal discovery algorithms provides a novel approach toward the philosophical problem of inductive skepticism. This new approach shows that there are additional fine-grained structures of inductive inferences that is more appropriate to characterize scientific inferences than the traditional understanding of induction.

**Key Words:** Causal discovery; The problem of induction; Machine learning; Logical reliability theory

中图分类号: N0 文献标识码: A DOI: 10.15994/j.1000-0763.2020.08.003

### 一、因果关系与归纳问题

关于因果关系的判断在科学研究中非常普遍, 它们在科学论文和研究报告中随处可见, 如潜意识提示可以改变人类的决策行为和偏好, 童年时期的逆境如贫困、灾难经历等可能会导致记忆与学习能力的降低等。<sup>[1]</sup>而另一方面, 哲学家们仍然没有找到一个公认的关于因果性的形而上学理论。<sup>[2]</sup>我们是否能够在尚未完全确定因果性的定义的情况下做出关于因果关系的判断? 如何在没

有完整的形而上学基础的前提下论证某个具体的因果判断是否合理? 休谟的归纳怀疑论或许是对所有建立因果判断的努力的最致命的攻击。它所针对的不是某个具体的自然世界中的因果关系, 而是人类是否能够进行合理的因果判断的可能性。信念与知识之间的主要区别在于, 知识是建立在充分的理由和充足的证据基础上的, 而信念则不必如此。对于一个因果判断, 例如一个运动的小球碰撞一个静止的小球导致后者开始运动, 如果我们考察它的证据基础会发现, 能够直接观察到的关联与原因和结果之间的必然联系相距甚远,

基金项目: 首都师范大学新兴交叉学科建设项目“人工智能哲学”(项目编号: 01955071)。

收稿日期: 2019年11月10日

作者简介: 杨仁杰(1985-)男, 黑龙江哈尔滨人, 首都师范大学政法学院讲师, 研究方向为计算社会科学。Email: 6687@cnu.edu.cn

因此仅仅依据这些观察现象不足以建立任何因果关系。休谟认为所有从观察经验到因果关系的推断 (inference) 都没有合理的基础。他的论证可以分为两个部分。首先, 因果推断不是演绎推断, 因为在下一次的观察中原因并不导致结果出现, 这种情况在逻辑上是完全可能的。因此不能用演绎的方法来论证因果推断的合理性。这意味着因果推断是一种归纳推断。第二, 我们如何判断哪些作为归纳推断的因果推断是合理的, 哪些不是? 似乎能够想到的所有论证某个因果推断是合理的理由都要用到归纳推断本身。例如我们可能会说, 从观察到的反复关联可以推论出因果关系, 因为在过去这种推断方式都奏效了, 所以在未来也应该会奏效。然而这种关于过去与未来的一致性原则似乎同样需要归纳式的辩护: 一致性原则在过去一直是成立的, 因此在未来也应该会成立。而目的恰恰就是为归纳推断找到充分的理由。因此这一思路会导致循环论证。<sup>[3]</sup>

休谟的归纳怀疑论针对的不仅是因果推断, 而是所有那些结果包含的内容比前提更多的扩展型推断 (ampliative inference)。广义的归纳推断包括所有扩展性推断, 它比通常称作狭义归纳推断的枚举型归纳 (enumerative induction) 范围更广。在本文中所提到的归纳推断, 除特殊说明外, 均采取这种广义的定义。如果归纳怀疑论是对的, 那么归纳推断没有合理性基础, 因为归纳推断要求我们从观察到的现象推论出关于未观察现象的结论, 但不存在使得这种推断成立的非归纳式的理由。其严重后果在于, 我们无法建立任何关于因果关系的科学论断的理性基础。这是因为科学不仅仅是对于已经观察到的现象的总结。除了需要预测未观察到的现象之外, 科学工作还必须建立关于不可直接观察的对象的理论或规律, 例如引力场、亚原子粒子、抗体或经济稳定性等。<sup>[4]</sup> 这意味着科学推断在本质上是扩展的, 从而也是归纳的, 因此会受到归纳怀疑论的威胁。

在科学哲学领域, 对休谟的归纳怀疑论的回应有很多种, 例如比较流行的贝叶斯确证理论以及演绎主义的回应等。本文讨论一种与这些传统科学哲学理论不同的另外一种回应方法。在机器学习领域, 因果发现问题探讨的是如何根据有限个变量的观察数据推断出这些变量之间的因果关联。近年来, 机器学习研究者们提出了很多能

够解决某些特殊的因果发现问题的算法。他们对这些算法是否正确的验证方式提供了一种回应归纳怀疑论的特殊方法。本文论证这种方法不但具有与传统的科学哲学确证理论不同的特征, 并且还能够揭示出归纳推断的更为精细的结构。论文第二节首先对因果发现问题的机器学习理论和算法的基本思路做一简要介绍, 然后在第三节分析和总结机器学习领域验证因果推断算法的推理 (reasoning) 模式, 指出它所蕴含的与以往不同的另外一种关于归纳推断的理论, 以及这种理论如何回应归纳怀疑论。论文的第四节将上述理论与其它相关理论进行比较并指出其优势所在: 能够揭示出归纳推断问题的更为精细的结构。结论部分阐述未来的研究方向和建议。

## 二、机器学习与因果发现

在机器学习领域, 统计推断与因果推断的一个共同主题是寻找事件或属性之间的关系。然而与统计相关关系不同, 因果关系的建立需要关于对系统进行干预之后的信息。例如, 生物学家试图了解睾酮与个体的暴力行为之间的关系, 统计数据显示所有不同的人类文化中男性总是更加暴力的一方, 并且睾酮水平比较高的个体更倾向于做出暴力行为。然而这些统计数据并不足以建立睾酮与暴力之间的因果关系。我们还需要对生物被试进行“去除”和“替换”的实验信息: 如果在对男性阉割之后暴力行为减少, 则意味着暴力与男性生殖系统相关; 再为阉割后的男性重新摄入替换的睾酮, 如果男性暴力行为增加, 那么就证实了睾酮与暴力之间的因果关系。<sup>[1]</sup> 这个例子说明了科学家通常所理解的因果关系的含义。统计预测考虑的是在观察到变量X的情况下对变量Y的预测, 而因果预测则是关于改变X的值之后在Y上会有怎样的后果。

因果关系的建立与统计关系的建立相比难度大大增加, 原因在于因果关系考察的实际上是反事实判断。例如在医学中需要验证氟西汀 (Fluoxetine) 是否能够治疗抑郁症。<sup>[5]</sup> 为简化讨论, 我们用 $X=1$ 表示服用氟西汀,  $X=0$ 表示不服用;  $Y=1$ 表示有抑郁症状,  $Y=0$ 表示没有抑郁症状。统计推断问题关心的是当我们观察到X取某个值时Y的取值分布, 因此只需要考虑X和Y的概率分布就足够了。然

而因果推断问题则需要考虑另外两个量:将X的值设置为1时Y的分布,以及将X的值设置为0时Y的分布。将这两个随机变量分别记作Y1和Y0,它们被称作潜在结果(potential outcome),因为通常在实验中我们每次只能选择一种干预,因此只能观察到其中一个潜在结果;另一个则观察不到,因为关于它的判断是反事实条件判断。而因果推断恰恰就是要估计这两个潜在结果,因为它们之间的差是对X与Y之间的因果关系强度的一个度量。

尽管因果推断比统计推断困难,有两种理想化的方法可以估计因果关系的强度。澄清这两种方法有助于我们对因果关系的理解。仍然考虑氟西汀与抑郁症的例子。假设我们需要通过做实验来估计氟西汀对于某个被试s的抑郁症状是否有疗效,也即估计在s身上所体现的X与Y之间的因果效应强度(causal effect)。即使我们无法在同一时刻对s的X值做出两种不同的干预,我们可以找一个与s完全一样的另一个被试s',在其身上做另一种与s不同的干预,用s'的观察结果来估计s的对应的潜在结果,从而能够比较s的两个潜在结果Y1和Y0之间的差别。如何保证两个不同的被试s和s'在所有相关的方面是完全一样的?有两种理想化的方法。第一种是进行随机试验,将一个被试群体完全随机地分配在两个组,这样在平均的效果上两个被试组没有明显差别。当无法进行实验而只能采用观察数据时,可以通过控制X和Y的共同原因的方式来分组观察,这些共同原因通常被称为干扰因子(confounder)。例如假设越是富裕的人越倾向于使用氟西汀,并且他们有更好的社会支持来帮助治疗抑郁症,那么即使氟西汀对抑郁症没有疗效,我们仍然可以观察到X和Y是正相关的。在这里,一个人的富裕程度就是干扰因子,对干扰因子的控制意味着将富裕和贫穷的人均匀分布在两个不同的被试组里,这样两边就不会有平均意义上的差别了。可以想象,如果我们在只采用观察数据时能够控制所有的干扰因子,那么仍然可以估计和比较两个被试组的潜在结果从而估计因果效应强度。当然这是非常理想化的条件,在实际科学研究中几乎不可能达成。以上就是对科学家们通常所理解的因果性概念的澄清。

在上述因果推断问题里,我们已经知道了X和Y发生的时间顺序,因此目的就只在于估计X对Y的因果效应。因果发现问题是指在不假设任

何时间顺序的情况下,如何寻找一组变量V之间的因果结构。与休谟的归纳问题直接相关的一个因果发现的特例是,如何只通过对V中的变量进行观察来估计V中各变量之间的因果关系。根据休谟的归纳怀疑论,我们无法从V的观察数据中直接读取因果关系,因此机器学习研究者所采用的策略是建立“桥接原则”(bridge principle),<sup>[2]</sup>将V的因果结构和V的可观察性质连接起来。

因果马尔可夫假设和因果忠实性假设是在因果发现领域常用的两个桥接原则。这两个原则阐述了概率的条件独立关系与因果图模型中的某种特定的分隔关系是如何相互对应的。<sup>[6]</sup>为了直观展示这两个原则如何帮助解决因果发现的问题,仍然考虑氟西汀和抑郁症的例子,并在其中加入第三个变量Z,它表示对于患者富裕程度的度量,如社会经济地位。如果我们只考虑抑郁症的初期情况,那么抑郁症和氟西汀还来不及对患者的富裕程度产生影响,那么就可以将Z作为X和Y的可能共同原因,而不考虑X和Y对Z的影响。在这种情况下,X,Y,Z之间所有可能的因果结构如图一所示。因果马尔可夫假设告诉我们,如果X和Y之间既没有直接的因果关系,也没有任何共同的原因,那么X和Y是概率独立的。现在假设我们观察到X和Y是概率独立的,那么与这一观察结果相一致的因果结构包括图一中(1)、(2)、(3)、(8)、(12),因为在其它图中所示的因果关系将会使得X和Y之间出现概率相关。(8)和(12)可能导致X和Y独立的原因是,即使在其中表示的所有因果关系的强度都大于0时,仍然存在某些特殊情况,例如在图一的(8)中Z对X的直接影响和Z通过Y对X的间接影响刚好相反从而相互抵消,从而使得X和Y之间保持概率独立。因果忠实性假设的作用是排除所有类似的因果效应相互抵消的情况,因此只有(1)、(2)、(3)是可能的。然而在这三种情况中X和Y都没有直接的因果关联,由此可以判断X不是Y的原因,Y也不是X的原因。由此我们就从V的观察数据当中找到V中变量X,Y之间的因果关联,其中因果马尔可夫条件和因果忠实性条件在观察结果和因果结构之间起到了桥接的作用。

基于上述思路,机器学习研究者们建立了一系列能够实现因果发现的算法。理论研究发现,如果V的因果结构是非循环(acyclic)的,即没有



任何变量可以通过其它变量间接成为自身的原因,并且是因果充分的 (causal sufficient), 即任意两个V中的变量的共同原因都在V中, 那么对于任意两个不同的因果结构, 只要它们对V的同一个联合分布满足因果马尔可夫假设和因果忠实性假设, 它们就有同样的不考虑方向的因果关联和同样的“V结构”, 其中V结构是指两个相互没有因果关联的变量共同导致另一个变量的局部因果结构。<sup>[7]</sup> 这一结论在V的所有可能的因果结构组成的集合中定义了一个等价关系, 每个等价类里的因果结构都拥有同样的不考虑方向的因果关联和同样的V结构。这种等价类被称为马尔可夫等价类。例如在图一(5)的例子中, (5)和(9)构成了一个等价类, (6)单独构成一个等价类。这意味着如果我们运用算法找到真实因果结构的马尔可夫等价类, 那么我们就可以将V的真实因果结构的可能范围大大缩小。一些机器学习算法如PC算法可以通过V的联合分布中的条件独立关系找到包含V的真实因果结构的马尔可夫等价类, 这类算法在某些特定的参数假设下, 如V中的因果关联是线性高斯时, 可以找到条件独立关系所蕴含的所有因果信息。<sup>[8]</sup>

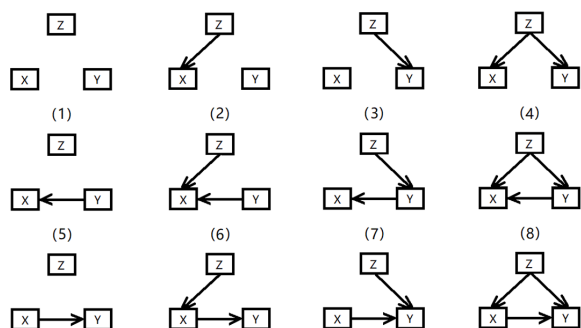


图1 三个变量的因果结构

在图一中我们看到, 如果真实的因果结构是(5), 那么通过PC算法我们只能找到(5)和(9)所构成的等价类, 而无法判断哪个是真实的因果结构。如果我们改变线性高斯的参数假设, 存在一些其它的因果发现算法可以帮助我们识别出这些更精细的因果结构。例如, 日本研究者清水 (Shimizu) 证明了当两个变量之间的因果结构是线性非高斯的, 可以根据其中一个变量与另一个变量的误差项之间的独立关系来识别出这两个变量之间的因果关联的方向, 并且这种情况不需要假定因果忠实性。机器学习研究者们基于这一思想开发了LinGAM算法及一系列推广。<sup>[9]</sup> 类似的结果还可以推广到非线性并且误差项为可加模型的

因果结构。除因果忠实性假定之外, 其它假定也可以被弱化或取消。例如, CCD算法取消了非循环的假定,<sup>[10]</sup> FCI算法取消了因果充分的假定等。<sup>[6]</sup> 在这些较弱的假定之下, 能够通过V的联合分布得到的关于V的因果信息将会减少。上述因果发现算法的目的是尽可能多地找到这些能够抽取出来的因果信息。可以证明在适当的条件下这些算法找到的信息都是正确的因果信息。

因果发现的机器学习算法并不仅仅具有理论价值。它们的最终目标是建构能够成功应用于科学研究的准确有效的计算方法。一些因果发现算法已经被用于流行病学、心理学、教育学、生物学等学科的研究和论文中。<sup>[6]</sup> LinGAM算法被用于fMRI脑成像数据的功能结构分析; PC算法广泛用在基因数据分析和生物科学研究中; 基于不同于条件独立关系的代数关系建立的因果因子分析被用来建立心理测量模型。<sup>[11]~[13]</sup> 因果发现的机器学习算法正在成为科学方法论中的一个组成部分。

### 三、归纳推断的计算理论

因果发现的机器学习算法要解决的问题本质上是归纳推断问题。按照休谟的观点, 对因果发现问题的任何回答都没有理性基础, 我们用来辩护因果判断的根据无非是心理习惯或习俗。但在上一节中所介绍的因果发现算法并非如此。对于人工智能的一个常见的误解是, 人工智能算法只是人类的某些心理过程在机器上的实现, 所以可以将机器学习的算法看成是人类在学习时的心理过程的某种复制。这种想法是错误的。大多数机器学习算法如果想要成功发表, 都需要在数学上严格证明它们能够给出正确的结果。而人类心理在进行判断和决策时常常发生误判。<sup>[14]</sup> 同心理过程相比, 人工智能算法有更高的规范性。对于上一节所提到的所有因果发现算法, 机器学习研究者都在演绎推断的严格意义上证明了它们在正确的输入下能够输出正确的结果。这些证明的最终实现需要三个主要的组成部分: 因果性的概念, 因果发现算法的正确性所需要假设的条件, 以及在何种意义上这些算法是正确的。下面详细分析这三个方面如何与归纳怀疑论相关。

在第二节里我们看到, 科学家们所理解的因果性的概念要求我们估计同一个个体的不同的潜

在结果之间的差别。然而通常的实验设定只能观察到对象个体的一个潜在结果,因此其它的潜在结果就需要通过其它的个体进行估计。这是一个从已观察量到未观察量的归纳推断问题。此外,无论是实验研究还是观察研究,研究对象只是一个有限样本。将有限样本的规律推广到其它样本同样也是从已知到未知的归纳推断。这些归纳推断问题都受到归纳怀疑论的威胁,但科学并不能因此停止发展。如果科学家们想开发一种有效的药品,他们不可能对每个人类个体进行检验,而只能运用之前积累起来的关于人类生物学和医学的知识来对某些样本设计实验。这些先在知识类似休谟的一致性原则,只不过更加详细和明确。如果不假定任何先在知识,就无法学到任何新知识。

在证明因果发现算法的正确性时所做的桥接假设和模型假设也是一致性原则的组成部分。它假定观察到的对象和未观察到的对象都满足例如因果马尔科夫假设,因果充分条件等。怀疑论者会询问科学家们是如何知道这些假定是正确的。实际上在很多情况下,科学家们不但不假定它们是正确的,而且会明确表明它们是错误的,<sup>[15]</sup>但即使在错误的假定下用机器学习算法来分析数据仍然是有帮助的。例如如果在运用PC算法来寻找因果结构时发现,某两个变量之间存在两个不同方向的因果关联,那么这一结果可能暗示在这两个变量之间有隐藏的共同原因,因此需要用不假设因果充分的算法来分析。<sup>[6]</sup>这也是机器学习研究者不断努力弱化假设并寻找有效算法的原因。设计并验证这些算法的正确性并不是为了彻底解决归纳怀疑论,而是为科学研究中的归纳推断提供越来越丰富的有效的数据分析工具。这些工具可以帮助科学家们找到更多的自然界中可能存在的因果关系。

一个常见的误解是,归纳怀疑论只关涉那些与一致性原则相关的桥接假设和模型假设,于是只把注意力集中在如何论证如因果马尔科夫假设、因果忠实性假设是否合理。实际情况远非如此。即使是在这些假设都成立的前提下,在归纳推断中仍然存在更为丰富的与归纳怀疑论相关的精细结构。这些假设本身是不足以证明因果发现算法的正确性的,还需要一种对于算法如何趋近正确结果的收敛标准。以PC算法为例,在包括因果马尔科夫假设等一系列假设都成立的条件下,如果

输入所有正确的关于V的条件独立关系,可以证明PC算法能够输出包含V的真实因果结构的马尔科夫等价类。这里受到归纳怀疑论影响的是,联合分布中的条件独立关系并不是可以直接观察的对象,而必须通过统计假设检验的方式进行估计。由于假设检验只能观察有限个样本,从有限样本的性质到总体性质的推断是一个典型的归纳推断。条件独立假设检验通常在大样本极限下收敛到正确结果,因此PC算法也只能在大样本极限下收敛到正确的因果信息,而不能保证在有限步骤内得到正确的结果。这意味着,证明PC算法的正确性不仅需要桥接假设和模型假设,还需要一种适当的收敛标准。如果换一种更强的收敛标准,例如要求算法在某个有限的样本量N之内就达到正确结果,那么就无法证明PC算法是正确的。

总结以上的介绍和分析,可以看到机器学习研究者对因果发现算法是否正确的验证过程遵循如下推理模式:从一些背景知识出发,假定某些桥接原则和模型假设,设立一个如何趋近正确结果的收敛模式,根据要解决的因果发现问题设计适当的算法,最后判断算法是否能够按照设定的收敛模式收敛到正确的结果并给出数学证明。不仅因果机器学习遵循这种模式,统计机器学习也是如此。这是由于机器学习领域的核心对象是算法,因此对于归纳推断的计算方法的要求与对一般算法的要求是一致的。我们可以在可计算性的数学理论中找到同样的理论结构。这种基于算法的归纳推断理论在历史上由哲学家赖辛巴哈(H. Reichenbach)、普特南(H. Putnam)、凯利(K. Kelly),计算机科学家奥舍森(D. Osherson)<sup>[16]-[19]</sup>等人建立并发展起来,它常常被称作逻辑可靠性理论(logical reliability theory),也称作形式学习理论(formal learning theory)。这一归纳推断理论隐含在大量科学研究背后。每次我们在经验研究的报告中看到t统计量、p值、或者卡方检验等统计结果时,其内在的理论基础都是逻辑可靠性理论。正如葛莱默(C. Glymour)<sup>[20]</sup>所评价,机器学习实质上是另外一种做科学哲学研究的方式。因此应该得到科学哲学家们更多的重视。

#### 四、逻辑可靠与确证

自从休谟提出归纳怀疑论的现代表述以来,



人们尝试了许多方式试图对其进行回应。这一节里我们将其中有代表性的理论与上文所总结的逻辑可靠性理论做一初步比较。

传统的科学哲学中比较常见的对归纳怀疑论的回应方式与演绎推断有关。演绎推断可以保证如果前提为真那么结论必然为真，尽管归纳推断不具有这种必然性，但可以将归纳推断看作是演绎推断的一种弱化形式：前提为真不能必然保证结论为真，但可以为结论为真提供某种支持。这种支持比必然的逻辑蕴含关系更弱，但没有弱到可以忽略不计，通常可以用某种概率测度来刻画。这种证据支持的观点是很多科学哲学确证理论的核心思想。<sup>[21]</sup> 贝叶斯确证理论是其中较为典型的一种。比较常见的贝叶斯确证理论采用概率的主观解释，将其看成是对某个信念的相信程度的度量。主观贝叶斯确证理论的主要思路是，科学家们为某个具体的科学理论赋予一定的先验相信度，用先验概率来表示，满足概率公理；然后科学家们不断收集数据作为证据，运用贝叶斯定理计算以新证据为条件的后验概率，并将这些后验概率解释为科学家在考虑新证据之后对原有科学理论应该具有的新的相信度。证据对于理论的支持或确证程度可以通过后验概率与先验概率之间的差别来度量。贝叶斯主义对归纳怀疑论的回应来自于对支持或确证概念的精确刻画，这种刻画不但可以解释科学史中很多具体的科学推断背后可能的逻辑，并且可以通过避免必输赌局（sure-loss bet）的决策论结果来进行辩护。<sup>[21]</sup>

关于贝叶斯确证理论有很多相关的评价和讨论，这里只考虑它与逻辑可靠性理论之间的根本区别。这里需要澄清一个常见的误解。贝叶斯确证理论和贝叶斯统计学是两个完全不同的理论，不能混淆在一起，尽管它们的名字都包含贝叶斯，并且都采取对概率的主观解释。贝叶斯确证理论的目的是阐释证据对于理论的确证关系的精确含义，贝叶斯统计学是研究在采用概率的主观解释后如何对随机现象进行统计推断的数学理论。贝叶斯统计学完全可以用来实现不同于贝叶斯确证的其它确证模式，例如假设-演绎确证。<sup>[22]</sup> 尽管逻辑可靠性理论并不假定概率的主观解释，但仍然可以用贝叶斯统计学来设计实现因果发现的算法，其基本思路与上文对逻辑可靠性理论的刻画相一致。因此贝叶斯确证理论与逻辑可靠性理论

的区别并不在于是否运用贝叶斯定理或是否采用某种特定的概率解释。两者之间的本质区别在于，贝叶斯确证理论只考虑了一种归纳推断的具体方法，即由贝叶斯定理实现的从先验概率到后验概率的信念更新。而与之相对地，逻辑可靠性理论考虑的是在不同的背景知识和不同的收敛准则之下，关于某个具体的归纳推断问题的所有有效的算法，其中有效的含义是在合适的输入下能够给出正确的输出结果。基于贝叶斯定理的信念度更新只是所有这些无限多种可能的归纳推断计算方法中的其中一种，它在某些情况下有效、某些情况下无效。<sup>[21]</sup> 逻辑可靠性理论提供了一个刻画各种具体的归纳推断方法的理论框架，这是贝叶斯确证理论无法做到的。

例如，统计学家们熟悉的一个常识是概率相关并不意味着因果相关。逻辑可靠性理论可以对统计推断与因果推断的本质区别做出精确的数学刻画。在统计学中，可以设计算法对统计模型参数做出估计，并且计算出估计量的置信区间，比如用最小二乘法估计线性高斯回归模型的回归系数。这意味着在统计模型假设之下，可以算出保证估计值与真实值之间小于某个特定距离的概率为95%所需要的样本数量。这对与实验设计非常有帮助。不仅如此，还可以比较不同统计模型下，随着样本量增加，同一个估计算法的估计值收敛到真实值的速率的大小。比如非参数模型的所有估计方法收敛到真实值的速率都比对应的参数模型要慢，因为非参数模型的模型假设比较少，因此能够从数据中提炼出的信息也比较少。而因果发现问题与所有统计推断都不相同，因为对纯因果结构的任何估计算法都没有置信区间，这也意味着不存在因果发现算法收敛到真实因果关系的收敛速率。<sup>[23]</sup> 这些通过逻辑可靠性理论所揭示的归纳推断的精细结构是贝叶斯确证理论无法刻画的。

## 五、结 语

本文从机器学习领域所通行的验证因果发现算法的推理模式总结出了一种回应归纳怀疑论的特殊方法。这种方法不但具有与传统的科学哲学确证理论不同的特征，并且还能够揭示出归纳推断问题的更为精细的结构。机器学习领域的研究

需要不断弱化已有的桥接原则和模型假设,并设计更有效的因果发现算法。哲学领域的研究则需要进一步澄清机器学习算法背后的逻辑可靠性理论与其它确证理论之间的精确关系,以及逻辑可靠性理论如何回应科学哲学中的经典论题如相对主义、新归纳之谜以及亚决定性问题。最后,本文指出在运用机器学习方法回应归纳问题时仍然需要继续深入研究的三个课题:

(1) 隐结构因果发现。在社会与行为科学中,科学家们经常用隐变量模型来刻画具体的社会现象。这些隐变量常常是不可直接观察的,而需要借助量表对它们进行度量。当我们掌握关于研究现象的大量观察数据时,是否能够设计有效的机器学习方法,不但能够从数据中发现背后的隐变量,并且可以发现这些隐变量之间的因果关系?如何运用逻辑可靠性理论为这一特殊的归纳问题提供理论基础?

(2) 有限样本归纳问题。机器学习算法通常只有在大样本极限下趋近于正确结果,如何根据有限样本进行理论模型选择是一个尚待解决的问题。对于统计机器学习方法,科学家们常常根据收敛速率来选择算法,但是因果结构模型的发现算法不存在收敛速率。如何为基于有限样本的因果模型选择提供有效方法和理论依据?

(3) 科学发现的哲学。与科学理论的发现相比,传统的科学哲学研究更加注重科学理论的确证问题。因果发现的机器学习方法为科学发现的哲学研究提供了新的思路。如何建立一个既能刻画确证问题也能刻画发现问题的一般科学哲学理论?

本文认为,在机器学习与归纳问题的交叉领域,上述三个课题是比较重要并且有很好发展前景的研究方向。通过本文的分析与论证,我们揭示出了另外一种做哲学的方法,即通过当代的计算方法来回应古老的哲学问题。哲学的任务不仅仅是提出问题,更为重要的是对这些问题的解决。通过人工智能与机器学习的科学研究发展出来的计算方法为重要的哲学问题的解决提供了新的思路与研究方向。

#### [参考文献]

- [1] Sapolsky, R. *Behave: The Biology of Humans at Our Best and Worst*[M]. New York: Penguin Press, 2017, 194-197.
- [2] Eberhardt, F. 'Introduction to the Epistemology of Causation'[J]. *Philosophy Compass*, 2009, 4(6): 913-925.
- [3] Hume, D. *An Enquiry Concerning Human Understanding*[M]. Oxford/New York: Oxford University Press, 1999, 17.
- [4] Klee, R. *Introduction to the Philosophy of Science: Cutting Nature at Its Seams*[M]. Oxford: Oxford University Press, 1996, 4-5.
- [5] Altamura, A. C., Moro, A. R. 'Clinical Pharmacokinetics of Fluoxetine'[J]. *Clinical Pharmacokinetics*, 1994, 26(3): 201-214.
- [6] Spirtes, P., Glymour, C., Scheines, R. *Causation, Prediction and Search*[M]. Cambridge: MIT Press, 2000, 29.
- [7] Pearl, J., Verma, T. 'Equivalence and Synthesis of Causal Models'[A], Bonissone, P., Henrion, M., Kanal, L., Lemmer, J. (Eds) *Proceedings of Sixth Conference on Uncertainty in Artificial Intelligence*[C], Berlin: Springer, 1991, 220-227.
- [8] Geiger, D., Pearl, J. 'On the Logic of Causal Models'[A], Kanal, L., Lemmer, J., Levitt, T., Schachter, R. (Eds) *Proceedings of Forth Conference on Uncertainty in Artificial Intelligence*[C], Berlin: Springer, 1988, 3-14.
- [9] Shimizu, S., Hoyer, P., Hyvärinen, A., Kerminen, A. 'A Linear Non-Gaussian Acyclic Model for Causal Discovery'[J]. *Journal of Machine Learning Research*, 2006, 7: 2003-2030.
- [10] Richardson, T., Spirtes, P. 'Automated Discovery of Linear Feedback Models'[A], Cooper, G., Glymour, C. (Eds) *Computation, Causation, and Discovery*[C], Cambridge: MIT Press, 1996, 253-303.
- [11] Ramsey, J., Hanson, S. H., Glymour, C. 'Six Problems for Causal Inference from fMRI'[J]. *NeuroImage*, 2010, 49(2): 1545-1558.
- [12] Glymour, C., Zhang, K., Spirtes, P. 'Review of Causal Discovery Methods Based on Graphical Models'[J]. *Frontiers in Genetics*. 2019, 10: 524.
- [13] Yang, R., Spirtes, P., Scheines, R., Reise, S. P., Mansoff, M. 'Finding Pure Submodels for Improved Differentiation of Bifactor and Second-Order Models'[J]. *Structural Equation Modeling: A Multidisciplinary Journal*, 2017, 24: 402-413.
- [14] Kahneman, D. *Thinking, Fast and Slow*[M]. New York: Farrar, Straus and Giroux, 2011.
- [15] Smaldino, P. 'Models Are Stupid, and We Need More of Them'[A], Vallacher, R., Read, S., Andrzej, N. (Eds) *Computational Social Psychology*[C], Routledge:

- London, 2017, 311–331.
- [16] Reichenbach, H. *The Theory of Probability*[M]. Oxford: Oxford University Press, 1949.
- [17] Putnam, H. 'Degree of Confirmation and Inductive Logic'[A], Schilpp, A. (Ed) *The Philosophy of Rudolph Carnap*[C], Lasalle, Ill.: Open Court, 1963, 761–783.
- [18] Kelly, K. *The Logic of Reliable Inquiry*[M]. Oxford: Oxford University Press, 1996.
- [19] Osherson, D., Michael, S., Scott, W. *Systems That Learn*[M]. Cambridge: MIT Press, 1986.
- [20] Glymour, C. 'Android Epistemology: Computation, Artificial intelligence, and the Philosophy of Science'[A], Salmon, J., Earman, J., Glymour, C., Lennox, J., Machamer, P., McGuire, J. E., Norton, G. J., Salmon, C. W., Schaffner, F. K. (Eds) *Introduction to the Philosophy of Science*[C], Prentice–Hall, Inc, 1992, 364–403.
- [21] Kelly, K., Glymour, C. 'Why Probability Does Not Capture the Logic of Scientific Justification'[A], Hitchcock, C. (Ed) *Contemporary Debates in Philosophy of Science*[C], Blackwell Publishing Ltd, 2004, 94–114.
- [22] Gelman, A., Shalizi, R. C. 'Philosophy and the Practice of Bayesian Statistics'[J]. *British Journal of Mathematical and Statistical Psychology*, 2013, 66: 8–38.
- [23] Robins, M. J., Scheines, R., Spirtes, P., Wasserman, L. 'Uniform Consistency of Causal Inference'[J]. *Biometrika*, 2003, 90(3): 491–515.
- [责任编辑 王巍 谭笑]