工程科学学报, 第 44 卷, 第 7 期: 1231−1243, 2022 年 7 月
Chinese Journal of Engineering, Vol. 44, No. 7: 1231−1243, July 2022
https://doi.org/10.13374/j.issn2095-9389.2021.07.04.002; http://cje.ustb.edu.cn

因果推断三种分析框架及其应用综述

马忠贵，徐晓晗✉，刘雪儿

北京科技大学计算机与通信工程学院, 北京 100083
✉通信作者，E-mail: g_runeko@163.com

摘　要　介绍因果推断所涉及的基本概念及其三种分析框架: 反事实框架、潜在结果模型和结构因果模型. 首先, 从反事实框架介绍因果效应的发端; 然后, 从基于反事实的两个因果推断分析框架: 潜在结果模型和结构因果模型, 来分别阐述两个分析框架所涉及的关键理论和应用方法. 其中, 潜在结果模型使用数学和可计算的语言对因果理论进行阐述, 是一种将假设、命题和结论清晰化表达的计算模型, 其在原因和结果变量已知的前提下定量分析原因变量对结果变量的因果效应, 并对缺失的潜在结果进行补齐, 使观察性研究的效果接近试验性研究. 结构因果模型则是一种基于图论的因果推断方法, 它将事件分为观察、干预和反事实三个层级, 并通过 do 运算将干预和反事实层级的因果关系都降维成可以通过统计学手段解决的问题. 最后, 探讨了现今多领域内因果推断的应用场景, 并总结了三种分析框架的异同点.

关键词　因果效应; 因果推断; 反事实; 潜在结果模型; 结构因果模型
分类号　TG142.71

Three analytical frameworks of causal inference and their applications

MA Zhong-gui, *XU Xiao-han*✉, *LIU Xue-er*

School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
✉ Corresponding author, E-mail: g_runeko@163.com

ABSTRACT　Causality is a generic relationship between an effect and a cause that produces it. The causal relationship among things has been a research hotspot; however, the complexity of causality is sometimes far beyond our imagination. Although some causality problems seem easy to analyze, finding an exact answer may not be easy. Nevertheless, through the continuous innovation and development of empirical research methods in recent decades, we have had several clear analytical frameworks and effective methods on how to define and estimate causality. Exploring the causal effects among things is a promising research topic in many fields, such as statistics, computer science, and econometrics. With Joshua D. Angrist and Guido W. Imbens winning the Nobel Prize in economics for their methodological contributions to the analysis of causality in 2021, causal inference is expected to thrive in these fields. This paper briefly introduces the basic concepts involved in causal inference and its three analytical frameworks, namely, counterfactual framework (CF), potential outcome framework (POF), and structural causal model (SCM). Firstly, we introduce the origin of causal effects according to CF. Secondly, based on the counterfactual theory, two analysis frameworks are considered (POF and SCM), and we introduce the associated key theories and methods. The SCM explains the causal theory through mathematics and computable language, and it is a calculation model that clearly expresses hypotheses, propositions, and conclusions. It quantitatively analyzes the pair of cause variables under the premise that the cause and effect variables are known. The POF makes up for the missing potential results, such that the effect of the observational research is close to experimental research. The SCM is a causal inference method based on graph theory. It divides events into three levels: observation, intervention, and counterfactual. Through the “do” operation, the causal relationship at the intervention and counterfactual levels could be reduced to low-dimensional problems, which can be solved *via* statistical methods.

收稿日期: 2021−07−04
基金项目: 中央高校基本科研业务费专项资金资助项目（FRF-DF-20-12, FRF-GF-18-017B）

Finally, the current application scenarios of causal inference in many fields are discussed in this paper, and the three analysis frameworks are compared.

KEY WORDS causal effect; causal inference; counterfactual; potential outcome model; structural causality model

为什么需要研究因果关系？因为有三件事需要在厘清原因的情况下才能更好地做到，那就是：解释、预测和干预. 合理的解释可以为探索世界提供支撑，准确的预测可以可靠地描述事件结果. 有时我们可能需要用一些理由去解释事件发生的原因，不仅想知道为什么发生，更希望可以利用其中某些信息来促进或者避免某些结果的产生，也就是对原本的事件施加干预（可以是一项行动、措施或政策）去得到特定的结果.

无论是哲学、自然科学还是社会科学领域，研究因果关系一直是人类持之以恒探索的终极目标. 从 Sabine 和 Russell[1] 对亚里士多德“四因说”中因果概念的论述：事物的出现所必需的条件都被称为原因；到文献 [2] 中所研究的关于确定现象因果联系的“穆勒五法”，再到文献 [3] 中论述了 Hume 提出的从“是”能否推出“应该”，也即“事实”命题能否推导出“价值”命题. 哲学领域对因果关系相关概念做出了透彻论述，因果关系的哲学思想发展史纵贯两千余年，因果推断的统计方法至今依然在社会学、计量经济学、流行病学等诸多科学领域发挥余韵，并展现出了巨大的潜力. 培根曾提出“真正的知识是根据因果关系得到的知识”，如何找到一种科学普适的方法探寻事物间的因果关系，随着人类认知的发展不断精深，依旧是一项不小的挑战.

目前因果推断的方法主要可以分为基于实证的方法和基于数据观察的方法. 其中，实证方法是进行因果关系推断的黄金标准，其干预决策是随机的，公认最有效的是随机对照试验（Random controlled trials, RCT），也称为 A/B 测试（A/B Test）. 随机对照试验将参与者随机分配到对照组或实验组，并将在试验中对照组和实验组唯一的预期差异视为实验的结果. 然而，随机对照试验虽然是分析因果关系的绝佳环境，但其受到伦理限制、个体不依从等因素影响，往往具有不可操作性，其试验范围也不可能遍及所有真实场景，因此在很大程度上限制了因果推断的应用. 基于数据观察的观察性研究（Observational study）也是一种常用方法. 研究人员在没有任何干扰的情况下观察受试者并得出数据，从观测数据中得到他们的行动及其结果，但不能得到他们采取特定行动的动机. 基于数据观察方法的核心问题就是如何基于已有的观测数据得到反事实结果，这是颇具挑战性的一项工作，主要原因如下：(1) 根据获取到的观测数据只能得到事实结果，无从得知其反事实结果；(2) 在无干预的观测环境中，试验往往不是随机分配的，这可能会因观察对象群体分布不同有较大偏差. 为了合理地从数据中推断因果关系，研究者们构建出了基于潜在结果模型和结构因果模型两种主流方法的分析框架，我们将针对不同的分析框架分别进行介绍.

本文先介绍因果关系和因果推断过程中所涉及到的基本概念，以辅助对后文因果推断的理解；然后我们将分别阐述现阶段因果推断中三种主流分析框架：反事实框架、潜在结果模型和结构因果模型；再介绍因果推断在各个学科领域的研究应用；最后总结三种分析框架的主要特点，并提出因果推断未来的应用前景及其巨大潜力.

1 因果推断的相关概念

1.1 相关不是因果

事物间的因果关系常常是我们经常要面对和分析的问题，研究因果的意义在于：在许多领域，我们需要理解数据并据此做出进一步的行动和决策. 比如对于我们而言，常常会想要知道“学历越高就会找到越好的工作吗？”政府可能想知道“增设离婚冷静期会对离婚率有影响吗？”医生可能想知道“某种药剂的使用会增加患者康复的几率吗？”这些问题的核心就是因果效应，即：X的变化会对Y造成影响吗？如果会，Y受影响的程度要如何度量？在研究因果关系的基础上，还要进一步挖掘因果关系产生的原因及其造成的影响.

从时间序列的角度，经济学家 Granger[4] 给出了因果关系的文字描述：如果利用X可以更好地预测Y，那么就可以说X是Y的原因. 经过后来的研究不难发现，这段描述中存在着一些谬误；严格来说，这句话描述的是相关关系，而非因果关系. 那么相关关系和因果关系究竟有何不同？其实，相关表示一种一般关系，即：当两个变量同时呈现出增加或减少的趋势时，它们就是相关的；而因果关

系中原因会导致结果，结果部分取决于原因．两个变量之间即便没有相关关系也可能具有因果关系，反之有相关关系也可能没有因果关系．例如，一项研究表明，通常吃早饭的人比不吃早饭的人体重轻，因此得出结论：不吃早饭有利于减肥．但事实上，不吃早饭和体重轻之间可能只是相关，而并非因果关系．从事实的角度出发为这个现象寻求一种解释，可能只是因为每天吃早饭的人习惯于保持一种健康的生活方式，定期运动、睡眠规律、饮食健康，最终才拥有了更加理想的体重．在这类情况中，拥有更健康的生活方式是吃早饭和轻体重的共同原因，因此也可以将其视为吃早饭和轻体重之间因果关系的混淆因素．

因此，相关不是因果．虽然相关关系在统计学中取得了一系列成果，但因果关系可以拓展传统统计学解决新问题所需的必要内容，并可延伸到其他学科，因此更具研究价值．

1.2 因果推断的基本概念

在因果推断中有一些基本概念，某种程度上在几类分析框架中是通用的，也是理解因果推断的基础，以下将分别进行介绍．

(1) 同一个研究对象 (Unit)：即在施加干预以研究因果关系时选定的研究对象，可以是一个物理对象也可以是一个对象的集合；在潜在结果模型中，不同时间点下的研究样本是不同的研究对象．(2) 干预 (Treatment)：干预指对一个样本采取的行为，用$W \in \{0,1,2,...,N_w\}$表示干预，目前大多数因果推断采用二元干预，即采用了干预 ($W=1$) 的样本划为干预组；未进行干预 ($W=0$) 划为对照组．(3) 潜在结果 (Potential outcome) 和事实结果 (Observed outcome)：在现实世界中，对于每一个研究对象，其在每一种干预下都存在一个可能的结果，即潜在结果；而在真实观测数据中出现的结果称为事实结果．(4) 效应 (Effect)：效应即评判干预与否所导致结果差别的指标，通过对各个研究对象干预与否的潜在结果的比较得出．(5) 分配机制 (Assignment mechanism)：哪些结果可以被观察到主要取决于干预的分配机制，即哪些研究对象对应采取了哪些干预．对于一个二值干预，$X=1$代表干预组，$X=0$代表对照组，在接受干预分配$X=x$后结果变量表示为Y_x，表示接受相应干预后的潜在结果．

1.3 因果推断的分析框架

近几十年来，因果推断一直是许多领域的关键性研究课题，在各个领域都涌现出了令人瞩目的研究成果．2008 年，诺贝尔经济学奖获得者 Heckman[5] 提出了政策评价中出现的三个基于因果推断且极具挑战性的难题：

(1)评价历史上出现的干预对结果的影响；

(2)预测在一个环境中曾经经历过的干预在其他环境中的影响；

(3)预测历史上从没有经历过的干预在各种环境中的影响．

通常认为，哲学和统计学得益于定量化数据记录对各个学科的普适性以及统计学以数据为分析对象的特点，先后提出了三种分析框架，即反事实框架、潜在结果模型和结构因果模型．反事实框架介绍因果效应的起源，潜在结果模型和结构因果模型是在反事实理论的基础上进一步发展，并成为了发现因果关系和评价因果效应时理论最成熟、应用最广泛的两种因果推断分析框架．随着实证研究方法的不断创新发展，对于如何界定因果关系以及推断事物间的因果关系已经有了比较成熟的理论，以下将详细阐述．

2 反事实框架

在因果关系的研究中，对于因果关系的界定几个世纪以来哲学家们都没有给出一个明确的定义，这主要是因为因果关系中原因和结果的定义在某种程度上都是以彼此为阐述条件，即需要结果来定义原因，也需要原因来判定结果，使得二者的关系纷繁复杂、扑朔迷离．

在很长一段时间内，哲学中关于因果推理的主要范式是遵循“连续性或相关性的规律”，将因果推断看成是一个挖掘事物规律的过程．无论是 Cook 等[6] 提出的判定因果关系的三项原则，还是 Lazarsfeld[7] 提出的因果判定方法，都强调因果关系中“规律性”的影响；随后，越来越多的学者认识到通过连续性或相关性的规律并不一定能得出真正的因果关系，使得哲学中开始出现通过反事实框架（Counterfactual framework）来探究因果关系的方法．Hume[8] 于 18 世纪最早提出基于反事实框架讨论因果关系并给出了反事实的文字化阐述，Lewis[9] 在 Hume 的研究基础上给出了反事实框架的符号化表达，结合可能世界语义学和反事实来刻画因果依赖性，并形成了“界定可比较相似性→用相似性来说明反事实→用反事实来定义反事实依赖性→用反事实依赖性来阐述因果依赖性→用因果依赖性来解释因果性”的逻辑链条．Lewis 提出的因果依赖命题是对 Hume 因果关系的正式概括——“若事件 A 没有发生，则事件 B 也不

会发生”, 一旦这一反事实命题成立, 则可得出“若事件 A 发生, 则事件 B 发生”的命题自然成立. 至此, Lewis 完成了从因果依赖性向因果性的跨越, 他指出: “如果 A 和 B 是两个现实事件且满足若 A 不发生则 B 不发生, 则可以确定事件 A 是事件 B 的原因”, 这一结论给出了因果关系中对于原因和结果比较明晰的界定方法, 为因果关系的理论思考提供了一种明确的道路.

3 潜在结果模型

在因果推断的理论体系中, 潜在结果模型(Potential outcomes framework)是其中最重要的理论模型之一. 潜在结果模型由哈佛大学知名统计学者 Rubin 提出, 因此该模型又称 Rubin 因果模型. 潜在结果模型的核心是对同一个研究对象, 比较其接受干预和不接受干预的效应. 对于接受干预的研究对象而言, 其不接受干预是一种“反事实”状态, 而对于不接受干预的研究对象而言, 其接受干预就是一种“反事实”状态. 对于“反事实”框架的概念, Rubin 却并不认同, 他认为对于一个研究对象, 其结果出现与否主要取决于分配机制(Assignment mechanism), 事实上我们只能看到一种结果, 但并不意味着另一种结果不存在, 这并不是一个非黑即白的概念, 因此用潜在结果去描述事件是一种更加恰当的方式.

3.1 基本概念

潜在结果模型跳脱出因果推理的正统思想, 转而着重哲学中反事实框架的影响, 通过借鉴统计学中随机对照试验和潜在结果的概念, 构建了因果推断的新分析框架. 潜在结果模型的核心假设是“没有假设就没有因果”, 以下将分别介绍潜在结果模型中涉及到的一些必要概念, 以更好地理解这种分析框架.

潜在结果框架分析中, 通常说因果将干预和研究对象联系在一起, 干预就是原因, 干预所导致的结果就是效应. Imbens 和 Rubin 在文献 [10] 中提出潜在结果的含义, 即: 给定一个研究对象和一系列干预, 将每一对“干预–结果”界定为一个潜在结果. 在潜在结果模型中, 文献 [11] 将因果效应定义为同一个研究对象潜在结果之差, 令 Y 表示研究对象 i 接受干预或对照的结果, 则因果效应(Individual causal effect, ICE)可定义为 $\mathrm{ICE}(i) = Y_1(i) - Y_0(i)$. 尽管 Rubin 等学者清楚地定义了一个研究对象的因果效应, 但是通常对于同一个研究对象, 不可能既接受干预又接受对照, 自然也不能够同时观测到两种结果. 统计学中往往关注总体的统计特征, 利用潜在结果还可以得到总体的平均因果效应(Average causal effect, ACE). 假设所有研究对象都接受干预 $X = 1$ 的平均结果为 $E(Y_1)$, 所有研究对象都接受对照 $X = 0$ 的平均结果为 $E(Y_0)$, 则可通过 $\mathrm{ACE}(i) = E(ICE) = E(Y_1 - Y_0) = E(Y_1) - (Y_0)$ 来表示平均因果效应的两个平均结果之差.

统计学家 Fisher 给出了识别平均因果效应的方法—随机对照试验, 即将干预随机分配给研究对象 i. 例如, 通过抛硬币决定研究对象 i 是否接受干预, 这个随机决定的过程与潜在结果无关, 可以保证潜在结果 (Y_1, Y_0) 与干预分配机制完全独立, 进而得到 $E(Y_x) = E(Y|X = x)$, 平均因果效应可以通过在干预组与对照组中观测到的结果变量期望之差 $\mathrm{ACE} = E(Y|X = 1) - E(Y|X = 0)$ 得到, 计算式中不再含有潜在结果变量 Y_1 和 Y_0, 可以通过传统的统计方法分别估计 $E(Y) = E(Y|X = 1)$ 和 $E(Y) = E(Y|X = 0)$ 来推断平均因果效应. 在实际研究中, 随机对照试验往往需要比较理想的试验条件. 一旦随机对照试验条件不成立, 这种通过计算干预组与对照组之间差值来得到平均因果关系的方法就面临着内生选择性偏差[12]、差别化干预效应偏差等问题的困扰. 文献 [13] 定义了观察性研究的概念, 其不再满足随机分配的条件, 在此种情形下如果忽略协变量的作用, 仅通过随机对照试验方法估计因果效应就会产生偏差, 乃至造成统计学悖论.

3.2 关键假设

在哲学和统计学领域, 许多优秀的学者为潜在结果模型的研究发展做出了卓越贡献, Hume[14] 和 Mill[15] 作为哲学家的典型代表最早从反事实框架的视角讨论因果关系; Fisher[16] 及 Neyman 等[17] 则各自从统计学家的立场出发, 分别提出从潜在结果和随机的视角来讨论因果关系. Fisher 提出了“随机对照试验”的概念, 而 Neyman 提出了“潜在结果”并将其应用于随机对照试验, Rubin 在文献 [18] 中进一步结合了“潜在结果”和“随机对照试验”这两个概念, 系统性地提出了潜在结果模型的理论假设、核心内容和推理方法. Neyman 利用数学语言描述了潜在结果框架下的因果效应, Rubin 将这一数学定义推广到观察性研究中, 潜在结果模型作为一种因果推断的重要分析框架, 其本身需要建立在一些基本假设和前提之上, 如果现实情况不能满足其基本假设, 潜在结果的结论就不成立, 本节我们将重点讨论潜在结果模型的三个基本假设.

3.2.1 研究对象干预值稳定性假设

研究对象干预值稳定性假设（Stable unit treatment value assumption, SUTVA）是一个先验假设，从广义上看，研究对象干预值稳定性假设强调两个要点：（1）研究对象之间不存在相互影响，互相独立；（2）每种干预的单一性，即不同层次的干预在 SUTVA 下不能归因为同一种干预[19].

Rubin[20] 在 1986 年指出无论分配干预 X 到研究对象 i 的分配机制是什么，也无所谓其他研究对象接受干预与否，研究对象 i 受到干预 X 影响而得到的结果总是不变的. 这一假设中既明确了其他研究对象的决策对参考研究对象接受干预与否没有影响，也包含了分配机制对潜在结果没有影响的假设.

STUVA 假设意味着每个研究对象所做出的决策并不受其他研究对象影响，然而在现实中这往往是不现实的，因为其是针对潜在结果模型的不完全假设，完全忽略了研究对象之间相互影响的间接效应. 针对 STUVA 的局限性，Sinclair 等[21] 提出了部分干扰假设（Partial interference assumption），即对干预分配机制进行两阶段随机对照设计，第一阶段给研究对象随机分层到不同干预分配策略，第二阶段将同层级内的研究对象进行随机干预或对照. 实验表明，当群体在空间、事件或社会性上充分分离的情况下，部分干扰假设可能是合理的.

3.2.2 可忽略性假设

观察性研究仅对观测数据进行观察，以推断变量间的因果效应，但这种方法不能由研究者决定是否针对某些研究对象采取干预或对照操作，因此观察性研究不再满足随机对照试验的条件.

在可忽略性假设（Ignorability of treatment assignment mechanism）中阐述了干预分配机制的可忽略性；即，在给定协变量 V 的条件下，干预 X 的分配独立于潜在结果 (Y_1, Y_0) 独立于 $X|V$. 可忽略性假设表明了两点：（1）如果给定两个研究对象共同的协变量 V，不管他们的分配机制如何，二者的潜在结果应该是相同的；（2）如果给定两个研究对象共同的协变量 V，无论他们潜在结果的值是什么，它们的干预分配机制应该是相同的. 因此，可忽略性假设又被称为无混杂假设，它意味着在具有相同协变量的研究对象之间，无论进行干预还是进行对照都不会影响潜在结果，也就是分配机制不会因为研究对象接受干预或者不接受干预的结果产生任何影响.

3.2.3 正值假设

如果在给定某些协变量的情况下，其干预分配机制是固定不变的，那么在这些依据协变量所做的分层中，至少有一种潜在结果是无法观察到的，在这种情况下，是无法通过潜在结果之间的差别来推断因果关系的. 例如，假设有两种药物 A 和 B，想要评估药物 A 对孕妇的影响，孕中期的孕妇总是被分配服用药物 A，那么在孕中期孕妇这个分组中研究药物 B 对其影响是没有意义的. 正值假设可用数学符号表达为：$0 < p(X = 1|V) < 1$，表明在基于协变量的分层中，每个研究对象接受干预或对照的概率都是正值. 正值假设表明了分配机制的可变性，这对估计干预效果有着不可忽视的重要意义.

在文献 [10] 中，可忽略性假设和正值假设一起被称为强可忽略性分配假设（Strongly ignorable treatment assignment）. $E(Y_1 - Y_0) = E[E(Y_1 - Y_0|V)] = E[E(Y|X = 1, V) - E(Y|X = 0, V)]$ 即可表示总体可识别平均因果效应，由可忽略性假设和正值假设条件相合得到. 可忽略性假设表明了随机对照试验和观察性研究的区别，当协变量的分布在干预组和对照组不一致时，就会产生一些偏差，我们将在下一部分进行讨论.

3.3 潜在结果的因果推理方法

观察性研究中如果忽略了协变量的作用，仅使用随机对照试验进行因果关系推断就会产生偏差，我们就将这种影响因果关系估计的变量称为混杂因素. 文献 [22] 基于相关关系的度量定义混杂因素为：如果两个变量之间相关关系的度量受到第三个变量的影响，那么称第三个变量为混杂因素. 文献 [23] 从潜在结果出发定义了混杂因素：$p(Y_1|X = 1) = p(Y_1|X = 0)$ 且 $p(Y_0|X = 1) = p(Y_0|X = 0)$，即如果干预总体的潜在结果 Y_0 和 Y_1 的分布与对照总体的潜在结果的分布相同，那么可以说在干预组和对照组之间无因混在因素而产生的混杂偏差.

本节将在 3.2 节所介绍的潜在结果模型的三个基本假设基础上，对潜在结果模型中现有的因果推理方法进行延展，根据其控制混杂因素的方式，我们重点介绍其中三种：匹配法、逆概率加权和分层方法.

3.3.1 匹配法

在可忽略性假设中，协变量在干预组和对照组之间的差别往往对因果推断的准确性起着举足轻重的作用，为了消除协变量分布在两组研究对象之间的差异，匹配法（Matching methods）常用于观察性研究中，其目的就是对每一个研究对象匹配一个具有相同或相近协变量取值的研究对象集

合, 使得通过匹配得到的数据在干预组和对照组之间有着相同的协变量分布, 然后再根据匹配数据进行因果推断. Cochran 和 Rubin[24] 在 1973 年提出用一个或多个协变量构建协变量集合, 但在一些因果关系中, 所涉及到的协变量较多, 难以判定根据哪些协变量构造的集合做匹配才能得到最准确的结果. 因此, 文献 [25] 中提出了倾向值匹配(Propensity score matching), 即将协变量的倾向值作为参考值来构造匹配集合, 并逐渐成为了观察性研究中最常用的匹配方法; 这种方法形式化定义倾向值为给定协变量时干预的条件概率, 并提出如果在给定协变量的情况下可忽略性假设成立, 那么在给定倾向值时可忽略性假设也成立, 因此可以用倾向值替代协变量在因果推断中进行分层或匹配, 从而避免了从众多协变量中遴选最适组合的困难. 文献 [26] 中根据给定研究对象集合, 提出根据研究对象匹配数构造匹配集合的方法, 并根据通用匹配估计平均因果效应. 如果在实际运算中无从得知真实的倾向值, 那么可以首先根据观察数据进行预先估计, 然后再用估计所得的倾向值做匹配, 常用的基于倾向值匹配的方法包括回归[27] 和决策树[28] 等.

3.3.2 逆概率加权

除匹配法之外, 逆概率加权估计方法[29] 也是一种基于倾向值的方法. 由于混杂因素的存在, 干预组和对照组的协变量分布不同, 会导致内生选择性偏差, 逆概率加权就是消除内生选择性偏差最有效的方法之一. 通过给观测数据中的每个研究对象分配适当的权重, 可以创建一个干预组与对照组具有相似分布的伪总体. 样本重加权方法涉及到的关键概念—均衡得分(Balancing score), 经均衡得分处理后的协变量与干预分配机制相独立, 倾向值就是均衡得分中的一种.

逆概率加权(Inverse propensity weighting, IPW)将权重 w 分配给每个研究对象: $w = X/\pi(x) + (1-X)/(1-\pi(x))$, 其中, X表示分配机制, $\pi(x)$表示倾向值. 文献 [30] 在重加权后, 可以将平均因果效应的逆概率加权估计为: $\mathrm{ATE}_{ipw} = \sum_{i=1}^{n} X_i Y_i/\pi(V_i) - \sum_{i=1}^{n}(1-X_i)Y_i/[1-\pi(V_i)]$. 文献 [31] 表明, 经不同规模的研究表明倾向值机制足以消除因协变量而产生的偏差; 然而在实际估计中, 逆概率加权的正确性高度依赖倾向值的正确性, 一旦倾向值计算出现偏差就会严重影响逆概率加权的准确性.

3.3.3 分层方法

分层方法(Stratification)又被称为子分类, 是调整混杂因素的代表性方法之一. 这种方法通过将整个总体划分为同质的子分层(Block)来调整因干预组和对照组之间差别造成的偏差. 理想状况下每个子层中干预组和对照组在协变量前提下的特定观测值是相似的, 因此, 同个子层中的研究对象可以看作遵循随机对照试验的分配机制, 也可以依据随机对照试验中的计算方法计算各子层内的平均因果效应. 与直接计算干预组和对照组结果差值的估计方法相比, 分层方法显著降低了平均因果估计的偏差, 但如何确定分层方式又是另一个研究要点.

等概率(Equal-frequency)方法[25] 通过倾向值对总体进行分层, 使得协变量在每一个子层中具有相等的倾向值, 总体的平均因果效应则通过每个子层中平均因果效应的加权平均进行估计. 然而这种方法会在某些权重过高或过低的子层中导致较大的方差, 针对这类问题, 文献 [32] 提出了一种对倾向值分层得到的子层进行逆概率加权的估计方法, 降低了等概率方法中出现的高方差问题.

4 结构因果模型

除潜在结果模型外, 因果推断中使用最多的一类模型就是结构因果模型(Structure causal model, SCM), Pearl[33] 阐述了这两类模型的等价性. 相比之下, 潜在结果模型更加精确, 而结构因果模型更加直观. 结构因果模型可以描述多个变量之间的因果关系, Pearl 基于贝叶斯网络提出了外部干预的概念, 并基于外部干预对因果关系形成了一种形式化表达方法, 开创了从数据中发掘因果关系和数据产生机制的方法.

4.1 图形化的因果关系

图论作为一种用途广泛的数学语言, 可以直观地描述事物之间相互影响的关系, 也能够经过简单运算解决因果问题. 在数学中, 有向图[34] 中节点X和Y之间的路径是指从X开始到Y结束的一系列由边首尾相接的节点, 路径上的第一个节点称为该路径上所有节点的祖先节点, 其他节点俱为祖先节点的后代节点[35]. 如果两个节点之间的路径能够沿着箭头方向追踪, 那么这条路径就称为有向路径. 当图中存在一个节点存在回到自身的有向路径时, 这个图称为有环图, 不存在环的有向图就是有向无环图(Directed acyclic graph, DAGs)[33].

从因果推断经典问题辛普森悖论[36] 可知, 某些决策无法仅从数据本身获得有效信息, 而是要细究数据背后的原因. 为了能够严格地处理这些

因果关系问题, Pearl 寻找出了一种能够借助图论这种数学工具形式化表述数据背后因果假设的方法, 即结构因果模型(Structural causal model, SCM). 其可用于描述现实世界关联特征及其相互作用, 具体而言, 结构因果模型描述了如何为感兴趣的变量赋值.

从形式上看, 结构因果模型由一组函数: $f=\{f_x: W_x \to X \mid X \in V\}$, 和两个变量集$U$和$V$构成, 其中$U$中的变量称为外生变量, V中的变量称为内生变量, 模型中的每一个内生变量都至少是一个外生变量的后代; 外生变量在图中表现为一个根节点, 它没有祖先节点, 特别不能是内生变量的后代. 如果知道每个外生变量的值, 那么利用函数f_x就可以完全确定每个内生变量的值.

我们主要讨论基于有向无环图的结构因果模型[37], 因此可以给出因果关系的图形化定义: 在图模型中, 如果变量X是变量Y的子节点, 那么Y是X的直接原因; 如果变量X是变量Y的后代节点, 那么Y是X的潜在原因(非传递性的特殊情况此处不予讨论).

在结构因果模型的理论体系中, 因果关系的推断依托于有向无环图的三种基本路径结构, 即链状结构、叉状结构和对撞结构, 三种结构具有不同的信息流转方式, 所有因果图都可以拆解为这三类结构的组合, 因此路径结构在结构因果模型的学习中占据着举足轻重的地位. 链状结构(Chain)就像一条链子一样, 可以表示为: $X \to Y \to Z$, 表示信息仅可单向流通; 叉状结构(Forks)$X \leftarrow Y \to Z$表示信息可以从中间分发到两端; 对撞结构(Collider)$X \to Y \leftarrow Z$表示中间同时接收两端节点的信息. 在结构因果模型中不管多么复杂的结构都可以拆分为以上三种结构的组合, 三种结构也可能导致不同的偏差. 链状结构会导致过度控制偏差, 叉状结构会导致混淆偏差, 对撞结构会导致内生选择性偏差; 在复杂因果模型的拆解分析中需要考虑全部因果路径, 才能推断出准确的因果关系.

4.2 因果关系的三个层级

因果关系已经成为了现代社会最重要的认知工具之一, 基于人类认知的发展, Pearl[33] 将因果关系划分为三个层级: 第一个层级是关联(Association), 它涉及到由数据定义的统计相关性, 观察性研究也正是依托于这一层级而实现的统计学方法; 第二个层级是干预(Intervention), 它不仅表明了通过观测数据能直观地看到规律, 更想知道如果对观察对象做出干预行为会导致什么结果; 第三个层级是反事实(Counterfactual), 这一层级是对过去所发生的行为的溯因和思考, 比如"如果某个研究对象没有采取 A 操作, 而是采取 B 操作, 与现在得到的结果会有何不同?"人类要想改变世界和创构世界, 就要迈上因果之梯的更高层级.

4.2.1 关联

图模型不仅能提供对因果关系的直观表述, 还能有效表达联合分布. 利用结构因果模型可以有效表示 n 个变量的联合分布: 通过确定表述变量之间关系的 n 个函数, 以及误差项的概率分析, 就可以得到联合分布概率以及各个边缘概率之间的关联. 对于任何有向无环图, 模型中变量的联合分布可以通过计算条件概率分布 P(子节点|父节点) 的乘积得出, 它不再需要创建一个概率表, 在海量模型数据运算中节约了大量的时间; 因此, 图的深层意义就是将一个高维分布估计问题降维成一个低维分布估计问题.

但是当我们无法得知变量之间具体的函数关系, 或无法获取误差项的分布时, 就可以应用下述 d-分离法则解决以上问题. 实际应用中基于图模型的因果关系表达往往更加复杂, 变量之间可能有多条路径连接, 且每个路径耦合多个基础路径结构; 因此, 应用贝叶斯网络中的 d-分离方法对一个复杂的因果图模型做解构. d-分离的判断方法是一对节点之间是否存在一条连通路径, 如果存在即为 d-连通; 如果不存在则为 d-分离, 也就是这两个变量相互独立. d-分离主要分为两类: ①不以特定节点为条件, 只有对撞节点可以阻断一条路径; ②以特定节点为条件, 包含链状和叉状结构的中间节点以及对撞结构除对撞节点外的其他节点. 使用 d-分离工具研究复杂图模型的优势在于 d-分离是非参数的, 仅需要依托图模型进行运算, 而不需要参考变量之间的函数关系, 并且 d-分离仅能实施局部性的检验模型, 而非进行全局性的检验, 这使得它可以识别假设模型中有缺陷的特定区域并对其进行修复完善, 从而得到一个全新的模型.

4.2.2 干预

关联是以观测数据为研究对象进行统计分析, 其并未改变数据的分布, 而干预就是要改变现有的数据分布. 在理想情况下, 随机对照试验输出变量的变化必然是由于输入变量而引起的. 在随机对照试验不可行的情况下, 研究者们采用观察性研究方法, 但这种方法很难将因果关系从相关关系中识别出来, 因此, 在实际应用中研究通过

干预方法预测干预措施的效果显得尤为重要. 对此, Pearl[38] 提出用 do 运算表达干预. 比如$P(Y=y|X=x)$表示在$X=x$的条件下$Y=y$的概率分布, 它反映了在X取值都是x的个体上Y的总体分布. 而$P(Y=y|\mathrm{do}(X=x))$表示通过干预使$X=x$时$Y=y$的概率分布, 它反映了将群体中所有个体的X取值都固定为x时$Y=y$的总体分布.

因果推断中的识别性和传统统计中的识别性定义是一致的. 统计学中, 如果两个不同的模型参数对应不同的观测数据的分布, 那么我们称模型的参数可以识别. 这里, 如果因果效应可以用观测数据的分布唯一地表示, 那么我们称因果效应是可以识别的, 因果关系可以通过 do 表达式和图模型从相关关系中识别出来. 在明确因果关系可识别后, 就需要通过施加相应的干预来研究其因果效应, 结构因果模型中对研究对象施加干预主要依托以下三种手段.

(1)校正公式: 在因果估计中, 通过计算平均因果效应以估计接受干预$\mathrm{do}(X=1)$和$\mathrm{do}(X=0)$之间的差别, 但是在没有因果关联的情况下, 无从直接通过观测数据本身估计因果效应. 因此需要通过借助图模型, 以对图进行处理的方式模拟干预, 即对哪个变量进行干预就把指向它的箭头去掉. 在校正公式中使用 P_{m} 表示修改后的图模型的概率, 因此也被称为操纵概率.

假设X是Y的原因, 且Z是X,Y的共因, 即因果图中存在$X\to Y$和$X\leftarrow Z\to Y$两条路径. 如果想要通过干预知道$P(Y=y|\mathrm{do}(X=x))$的值, 也即是$P_m(Y=y|X=x)$, 则将Z指向X的箭头移除. 在经过干预后, Z的边缘分布是不变的, 即$P_m(Z=z)=P(Z=z)$, 因为移除指向X的箭头是不影响Z的概率分布的; 同时, 经过干预后, 条件概率$P(Y=y|Z=z,X=x)$也是不变的, 因为不管X是自发变化还是被干预变化, Y对X和Z的响应函数不变. 因模型修改后X与Z是 d-分离的, 即$P_m(Z=z|X=x)=P_m(Z=z)=P(Z=z)$.

综上可知: $P(Y=y|\mathrm{do}(X=x))=\sum_z P_m(Y=y|X=x,Z=z)\mathrm{P}_m(Z=z)$. 最后利用不变性关系, 得到一个将干预降维表示的概率公式$P(Y=y|\mathrm{do}(X=x))=\sum_z P(Y=y|X=x,Z=z)P(Z=z)$, 这就是校正公式. 校正公式对$Z$的每一个取值$z$计算了$X$和$Y$之间的关系, 然后对这些值求平均, 这个过程就被称为“对Z的校正”.

(2)后门准则: 根据校正公式的讨论可以得知, 为了确定一个变量对另一个变量的因果效应, 需要对该变量的父节点变量进行校正; 但实际上变量往往会受到一些不可观测的父节点影响, 因此需要找到一个替代变量集合来用于校正. 那么这个问题就引发了另一个深层次的问题: 在什么条件下, 因果图足以描述给定数据集的因果效应? 这就涉及到了计算因果效应的一个重要干预工具—后门准则.

后门准则[39] 的标准定义为: 给定一个有向无环图G及其中一对有序变量(X,Y), 如果变量集合Z中的节点都不是X的后代节点, 且以Z阻断了X,Y之间的所有含有指向X的路径(即后门路径), 那么Z满足X,Y之间的后门准则, X对Y的因果效应可以通过校正公式进行计算. 后门准则的变量集合Z要求: ①可以阻断任何含有指向X的后门路径, 避免使X,Y相关但不传递X产生的因果效应; ②不以X的后代节点为条件, 避免阻断X,Y之间的因果路径; ③不以对撞节点为条件, 避免在X,Y之间产生新路径.

(3)前门准则: 在尝试采取随机对照试验以外的方法估计因果效应时, 后门准则提供了一种简便方法识别需要校正的变量集合, 但当变量X,Y之间存在不可观测变量无法阻断从X到Y的后门路径时, 其因果效应依然是不可识别的[40]. 为了处理这种特殊情况, Pearl 提出了前门准则: 给定一个有向无环图G及其中一对有序变量(X,Y), 如果变量集合Z中的节点① 切断了所有X,Y之间的有向路径; ②X到Z之间没有后门路径; ③所有Z到Y的后门路径都被X阻断; 那么可以通过前门准则进行校正.

例如, 图模型G中存在两条因果路径: $U\to Y$和$U\to X\to Y\to Z$. 假设U是不可观测变量, 且没有变量能够阻断$X\leftarrow U\to Y$这条伪路径, $P(Y=y|\mathrm{do}(X=x))$只能通过两次后门准则来识别. 首先, X到Z之间没有后门路径, 因此其因果效应可识别$P(Z=z|\mathrm{do}(X=x))=P(Z=z|X=x)$; 其次, 从$Z$到$Y$的后门路径$Z\leftarrow X\leftarrow U\to Y$以$X$为条件阻断, 再将两部分因果效应连接起来获得的整体因果效应, $P(Y=y|\mathrm{do}(Z=z))=\sum_x P(Y=y|X=x,Z=z)P(X=x)$ 去除 do 运算得到前门公式, 再将两部分连接起来获得X,Y的整体因果效应: $P(Y=y|\mathrm{do}(X=x))=\sum_z P(Y=y|\mathrm{do}(Z=z))P(Z=z|\mathrm{do}(X=x))$, 用以上两个公式替换, 去除 do 运算得到: $P(Y=y|\mathrm{do}(Z=z))=\sum_x P(Y=y|X=x,Z=z)P(X=x)$, 即前门公式.

基于校正公式的前门准则和后门准则在消除 do 运算的情况下对因果图进行干预, 仅通过已知的观测数据和变量分布就可以对变量间的因果关

系进行推断. 从理论上讲, 这两种方法已经基本涵盖了因果推断中的大部分可能性, 但是实际上, 我们无法通过观测数据得到所有Z的可能取值, 更无法对Z的全部值计算概率再求和, 在干预方法导致算力受限时, 可以通过逆概率加权方法克服校正方法的实际困难.

4.2.3 反事实

基于结构因果模型的反事实推理核心在于: 虽然现实情况下$X=1$, 但假如$X=0$时, Y会发生怎样的变化? 从上一部分我们知道可以通过 do 运算对因果图进行干预从而估计因果关系, 但是在反事实中, do 运算太过笼统, 不能明确区分干预与反事实, 因此使用新的表达方式来标记事实结果与反事实结果.

(1)反事实的定义与形式化表达.

定义一个完全确定的模型M并用G表达M中的函数集, 已知外生变量集U和内生变量集V, 其中$U=u$代表对某个外生变量的赋值. 反事实语句"在$U=u$的情况下, 若假设X取值x, 则Y会取值y"记作$Y_x(u)=y$, 其中X,Y均为V中的任意两个变量. 假设一个结构因果模型M, 其中$X=aU$, $Y=bX+U$, 那么计算反事实$Y_x(u)$, 即在$U=u$情况下, 假设X取值x, Y的取值情况; 用$X=x$替换掉$X=aU$得到修改模型M_x, 再代入$U=u$得到$Y_x(u)=bx+U$. 将反事实概念推广到任何结构因果模型M, 反事实$Y_x(u)$可形式化定义为: $Y_x(u)=Y_{M_x}(u)$, 表示模型M中的反事实$Y_x(u)$定义为修改后的子模型M_x中Y的解. 反事实一般遵循如下一致性规则: 如果$X=x$, 则$Y_x=y$.

反事实因果估计一般分为确定性模型和非确定性模型, 针对这两种因果模型的计算都可以分三步走. 其中, 针对确定性模型: ①溯因(Abduction),用证据$E=e$确定外生变量U的值; ②行动(Action),修改模型M并用$X=x$替换掉原来模型中变量X的表达式; ③预测(Prediction),使用修正后的模型M_x和U计算反事实结果Y值. 从时间出发对上述计算步骤进行解释, 首先根据当前证据e解释过去U, 再通过最低限度的干预来符合假设的$X=x$, 最后根据对过去的认识和新增的条件来预测未来. 类似的, 针对非确定性的模型总结的计算步骤为: ①溯因,通过证据更新$P(U)$获得$P(U|E=e)$; ②行动,修改模型M并用$X=x$替换掉原来模型中变量X的表达式; ③使用修正后的模型M_x和$P(U|E=e)$计算反事实结果Y值.

使用 do 运算表达干预是一种很好的方法, 那么能否在反事实中也沿用 do 运算呢? 首先, 反事实表达与 do 运算之间存在着巨大差异: do 运算刻画了干预之下的总体行为, 而$Y_x(u)$描述了一个特定研究对象$U=u$在干预之下的行为, 反映了计算总体水平与个体水平之间的差别. do 运算不能刻画反事实问题, 即$E(\mathrm{do}(X=1),Z=1)\neq E(Y_{X=1}|Z=1)$. 因为前者将$Z=1$看作干预后的条件在$X=1$和$X=0$两个前提下所满足的不同研究对象的集合, 而后者是在当前世界中定义$Z=1$的单独研究对象集合, do 运算不能描述后者. 因此, do 运算可以描述干预后的世界, 而反事实既能描述单一世界又能刻画跨世界的事件概率.

(2)反事实的图形化表达.

反事实作为结构方程的衍生理论, 也可以在与模型相关的因果图中得到表达. 如果修改模型M得到子模型M_x, 那么M_x中的结果变量Y就是原模型中的反事实Y_x. 模型修改要求移除所有指向修改变量X的箭头, 也就是说, 只有在修改后的模型中, 与变量Y相关联的节点替换为与Y_x相关联才成立.

事实上, 如果变量集Z满足$X\rightarrow Y$的后门条件, 那么在给定Z的条件下, 对于变量X的所有取值x, 反事实结果Y_x都与X独立, 即$P(Y_x|X,Z)=P(Y_x|Z)$. 根据以上定理结合一致性规则便可得到, $P(Y_x=y)=\sum_z P(Y_x=y|X=x,Z=z)P(z)$, 这就是熟悉的校正公式.

4.3 结构因果研究方法

结构因果模型主要解决的问题是识别变量间的因果效应, 其与潜在结果所解决的问题具有一定的相似性, 但可以比潜在结果模型更加精准地判断混杂因素[41]. 比如, 当一个变量与干预对象和结果变量相关时, 利用潜在结果方法无法判断出其是否是混杂因素, 但文献 [40] 基于结构因果模型方法提出了区分混杂因素和其他变量的方法文献 [39] 提出了基于结构因果模型的前门准则识别方法. 相较于潜在结果模型, 结构因果模型方法需要知晓现有因果模型, 而潜在结果模型无需得知结构模型, 但是需要遵循三个基本假设[42].

SCM 是对变量间因果关系的定性分析, 将结构因果模型进行参数化, 利用结构方程模型(Structure equation model, SEM)定量描述. 在 SEM 中结果变量Y可以用结构方程表达为$Y=f(X,a)$, 其中X是研究对象集合, a是误差项且独立于X. 文献 [43] 提出了当误差项不满足高斯分布或结构方程为非线性时, 研究对象与误差项之间的

因果关系是可识别的; 文献 [44] 将潜在混杂因素纳入因果推断的考量范畴, 并使用独立成分分析技术(Independent component analysis, ICA)进行推理模型选择. 针对似然函数存在的马尔可夫等价类问题, 文献 [45] 将 SEM 引入似然函数计算框架实现了似然函数和 SEM 的结合算法; 文献 [46] 提出了一种采用分治策略的混合加误差模型与条件独立性检测的因果方向推断方法. 针对存在隐含变量的因果关系推断, 文献 [47] 利用探索性因子分析得到相对独立的各个隐含变量, 再使用路径分析算法估计变量之间的因果关系. 针对高斯稀疏图模型, 文献 [48] 基于原有图模型构造了可逆的马尔可夫链, 在高维稀疏图上进行随机抽样; 文献 [49] 提出了一种基于互信息的高维数据因果推断算法, 将高维网络结构学习问题分解为每一个节点的因果推断问题. 文献 [50] 中比较了 DAG、SEM、贝叶斯网络和 TAN 贝叶斯网络四种模型在数据因果推断中的原理和应用价值, 并为因果推断模型选型提供了有力的参考依据.

5 应用

研究因果关系、挖掘因果关系的科学方法对各个学科领域都具有一定的普适性, 因果推断在计量经济学、计算机科学等领域中都得到了颇为丰硕的研究成果.

早期计量经济学的主要目标是运用概率统计方法对经济变量之间的因果关系进行定量分析, 更加偏重总结、估计和假设检验, 并不十分关注预测. 机器学习则因其实操属性更加片中预测而非因果推断. 进入大数据时代后, 二者的联系开始加强, 传统计量方法在样本量少且维度低的数据中应用效果较好, 但无法很好地处理大规模和高维异构数据. 大数据时代的丰富数据为从概率论立场研究因果关系提供了新视角, 海量样本数据有助于从根本上克服由于抽样偏颇所引起的内生选择性偏差, 使得对因果关系的检验比有限样本的抽样数据更为文件可靠. 2015 年, Bareinboim 和 Pearl[51] 提到了因果推断和数据融合问题, 提出将机器学习融合到高维数据因果推断中的初步构想. 文献 [52] 从机器学习与传统因果推断计量方法相结合入手, 总结机器学习在策略评估和事后分析中的应用. 文献 [53] 利用机器学习算法循环将各协变量划分成多个子层, 计算每个子层内的平均因果效应后, 再加权得出总体平均因果效应, 利用随机森林估计倾向得分. 文献 [54] 则引入协变量平衡倾向得分, 通过模型的干预分配来优化协变量平衡, 显著改善了倾向得分匹配和加权方法的性能. 目前大多数研究都利用机器学习“数据驱动”的特点, 全面考虑各种模型以进行反事实推断, 如文献 [55] 中提出了一种双重选择方法, 分别筛选与干预分配和结果变量都相关的协变量, 并对两组协变量进行最小二乘回归, 相较于简单正则化回归, 该方法改善了平均因果效应的估计效果. 文献 [56] 则分别使用决策树、随机森林、K 近邻和神经网络来揭示数据模式的流程, 可以更好地阐明交互作用和非线性效果, 作为传统因果推断的补充. 文献 [57] 提出了一种双稳健的交叉拟合估计器以估计平均因果关系, 在大样本场景下具有更好的统计特性. 文献 [58] 提出了一种新的基于机器学习的无监督学习方法—基于视觉常识区域的卷积神经网络, 将学习目标由常规的似然性转换为基于因果干预, 仅使用特征连接来支持各种高级任务.

因果推断在计算机科学中也具有广泛应用, 可与计算机视觉、推荐系统等方向结合. 在计算机视觉中, 现有的视觉问答(Visual question answering, VQA)更多地倾向于依赖语言而未能充分了解视觉和语言的多模态知识. 文献 [59] 以 Pearl 所构建的概率图模型为基础, 综述了现今主流多模态统计学习方法, 利用大数据背景下多模态数据对同一对象的描述形式多源异构、内在语义一致的特点研究更为有效的跨模态匹配. 文献 [60] 根据因果关系提出基于反事实框架的 VQA 模型, 能够通过从总因果效应中减去直接语言效应来减少对答案的提问和降低语言偏差的提问. 文献 [61] 通过建立因果推断框架寻求输入样本造成的直接因果效应, 降低了深度学习图像分类中因长尾效应的存在而导致的结果偏差, 可以有效解除矛盾效应并提高识别准确率. 文献 [62] 在弱监督的语义分割中基于图像背景和阶级标签之间的因果关系, 提出上下文调整(CONTA)方法, 以消除图像级分类中的混淆偏差, 从而为后续分割模型提供更好的基础. 在零次学习(ZSL)和开放式识别(OSR)中, 常见的挑战是由已知类别上进行训练所的结果推广到未知类别, 但已知和未知的类别之间的识别率严重不平衡; 文献 [63] 为 ZSL 和 OSR 提出一个反事实框架, 有效改善了已知和未知的分类失衡问题, 模型在整体性能方面得到显著提高. 针对基于关注的视觉语言模型中的混淆效果难以消除的问题, 文献 [64] 基于前门准则提

出新的因果机制, 通过减轻混淆效果提高注意力机制的性能. 针对推荐系统存在的选择偏差, 文献 [65] 通过提出一种新的域自适应算法, 采用小型无偏的数据集来纠正选择偏差. 尽管推荐系统能够产生高质量的建议, 但由于使用黑盒预测模型, 通常不能提供直观的解释; 文献 [66] 在维持推荐模型的预测准确性的基础上, 将来自用户交互历史的因果规则作为黑盒顺序推荐机制的解释, 利用反事实方法提取推荐模型的个性化因果关系, 为黑盒推荐模型的行为提供个性化和更有效的解释.

6 结论

通过对反事实框架、潜在结果模型和结构因果模型的分别阐述, 不难看出这三类分析框架既有共同点, 也存在差异. 反事实框架更多的是一种文字描述, 所涉及到的符号表达比较有限, 更多依赖文字表达因果关系理论, 哲学中的反事实更多是对休谟思想的清晰化阐述; 但反事实框架基本仍停留在哲学思想层面, 即便部分学者使用反事实框架描述因果关系, 但其依然是根据潜在结果模型或结构因果模型来构造.

潜在结果模型强调对同一研究对象施加干预或不施加干预的效应进行比较从而得出该干预所产生的因果效应, 它的成立依赖于三个关键假设: 第一、研究对象干预值稳定性假设, 即研究对象之间相互独立彼此干预与否都不会影响其他研究对象; 第二、可忽略性假设, 即给定协变量的前提下, 干预分配机制不会对潜在结果产生影响; 第三、正值假设, 即在基于协变量的分层中, 研究对象接受干预或对照的概率都是正值. 基于以上假设, 潜在结果模型又有匹配、逆概率加权和分层等因果推断方法. 潜在结果模型将观察性研究和试验性研究统一在一个框架下进行结合, 在有限的条件下使观察性研究贴近试验性研究, 也是判断因果关系的重要标准, 是因果理论体系中的重要构成部分.

结构因果模型基于图论对多变量之间因果关系进行图形化表达, 其在观察、干预、反事实三个层级对因果关系进行分析研究, 它是一种可以描述数据产生机制和外部干预的形式化语言, 通过在贝叶斯网络上引入外部干预, 来定义外部干预的因果效应并描述多个变量之间的因果关系. 与潜在结果模型相比, 结构因果模型不仅能定量评价变量之间的因果效应, 还可以定性评价混杂变量, 便于从数据中挖掘隐含的因果关系, 在人工智能不确定性推理方面取得了突破性进展.

在大数据时代, 数据的收集和分析在越来越多的学科中都日益重要, 因果推断分析框架也在此情景中逐步健全并演化. 因果推断打破了对数据相关关系的盲目迷信, 强调在数据挖掘的基础上要建立因果模型以提升认知. 因果推断在计量经济学、社会学和计算机科学等诸多学科领域展现出了蓬勃的发展态势. 从理论研究角度来看, 用数据学习因果推断仍具有较大潜力, 如研究在因果推断中增加或放宽假设限制时如何进行建模交互, 并使不同因果推断模型之间存在形式联系; 将机器学习和因果推断进一步融合, 使用因果知识改进机器学习算法, 包括对黑盒深度学习算法进行因果解释, 及学习更具鲁棒性和公平性的因果感知模型. 从应用和评估角度来看, 需要在更多领域应用中明确对“干预”和“效果”的广义解释, 整合部分实证研究和观察性研究得到可解释结果; 利用多模态数据得到具备复用能力的可扩展模型, 并使用真实数据创建、明确评估指标和目的基准.

2021 年 10 月 11 日, Angrist 和 Imbens 因“对因果关系分析的方法学贡献”而获得诺贝尔经济学奖. Angrist 和 Imbens 在上世纪 90 年代中期就论证了自然实验方法论对因果推断的准确性. 传统实证方法强调方法的正确使用, 而不聚焦数据收集; Angrist 和 Imbens 则强调在被动收集数据之前是否可以利用自然实验得到对因果关系的直接判断. 实证研究和因果关系方法论之间密不可分, 实证研究是因果关系方法论的试炼场, 并大力推动着因果关系的发展, 诺贝尔经济学奖对因果关系方法论的关注也昭示着从挖掘数据到真正理解数据, 因果推断必然会在各个研究领域大放异彩.

参 考 文 献

[1] Sabine G H, Russell B. A history of western philosophy and its connection with political and social circumstances from the earliest times to the present day. *Am Hist Rev*, 1946, 51(3): 485

[2] Zhang D S, Sun G J. On the methodological characteristic and value of mill's five canons. *J Central China Norm Univ* (*Humanit Soc Sci*), 2001, 40(6): 19
(张大松, 孙国江. 论穆勒五法的方法论特征与价值. 华中师范大学学报(人文社会科学版), 2001, 40(6): 19)

[3] Ying Q, Xu D S. Hume and the birth of analytical action philosophy—from the point of “causal explanation”. *Tianjin Soc Sci*, 2021, 12(2): 57
(应奇, 徐东舜. 休谟与分析性行动哲学的诞生—从“因果解释”视角看. 天津社会科学, 2021, 12(2): 57)

[4] Granger C W J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 1969, 37(3): 424

[5] Heckman J J. Econometric causality. *Int Statistical Rev*, 2008, 76(1): 1

[6] Cook T D, Campbell D, Shadish W. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin, 2002

[7] Lazarsfeld P. Problems in methodology. *Sociology Today*: *Problems and Prospects*. New York: Basic Books, 1959

[8] Hume D. *An Enquiry Concerning Human Understanding, and Selections from A Treatise of Human Nature.* Chicago: Open Court Publishing, 1912

[9] Lewis D. Causation. *J Philos*, 1973, 70(17): 556

[10] Imbens G W, Rubin D B. *Causal Inference for Statistics, Social, and Biomedical Sciences.* New York: Cambridge University Press, 2015

[11] Rubin D B. Bayesian inference for causal effects: The role of randomization. *Ann Statist*, 1978, 6(1): 34

[12] Morgan S L, Winship C. *Counterfactuals and Causal Inference Methods and Principles for Social Research*. 2nd Ed. Cambridge: Cambridge University Press, 2014

[13] Cochran W G, Chambers S P. The planning of observational studies of human populations. *J Royal Stat Soc Ser A*, 1965, 128(2): 234

[14] Hume D, Beauchamp T L. *An Enquiry Concerning Human Understanding*. Oxford: Clarendon Press, 1999

[15] Mill J S. *A System of Logic*: *in Collected Works of John Stuart Mill*. Toronto: University of Toronto Press, 1973

[16] Fisher R A. Statistical methods and scientific inference. *J Institute Actuaries*, 1957, 83(1): 64

[17] Neyman J S, Dabrowska D M, Speed T P. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statist Sci*, 1990, 5(4): 465

[18] Rubin D B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*, 1974, 66(5): 688

[19] Miao W, Liu C C, Geng Z. Statistical approaches for causal inference. *Sci Sin* (*Math*), 2018, 48(12): 1753
(苗旺, 刘春辰, 耿直. 因果推断的统计方法. 中国科学:数学, 2018, 48(12): 1753)

[20] Rubin D B. Statistics and causal inference: Comment ifs have causal answers. *J Am Stat Assoc*, 1986, 81(396): 961

[21] Sinclair B, McConnell M, Green D P. Detecting spillover effects: Design and analysis of multilevel experiments. *Am J Political Sci*, 2012, 56(4): 1055

[22] Bickel P J, Hammel E A, O'Connell J W. Sex bias in graduate admissions: Data from Berkeley. *Science*, 1975, 187(4175): 398

[23] Greenland S, Pearl J, Robins J M. Confounding and collapsibility in causal inference. *Statist Sci*, 1999, 14(1): 29

[24] Cochran W G, Rubin D B. Controlling bias in observational studies: A review. *Sankhyā*:*The Indian Journal of Statistics*, *Series A*, 1973, 35(4): 417

[25] Rosenbaum P R, Rubin D B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, 70(1): 41

[26] Abadie A, Imbens G W. Matching on the estimated propensity score. *Econometrica*, 2016, 84(2): 781

[27] Bahadori M T, Chalupka K, Choi E, et al. *Causal regularization*[J/OL]. *ArXiv Preprint* (2017-2-23) [2021-6-11].https://arxiv.org/abs/1702.02604

[28] Lee B K, Lessler J, Stuart E A. Improving propensity score weighting using machine learning. *Stat Med*, 2010, 29(3): 337

[29] Rosenbaum P R. Model-based direct adjustment. *J Am Stat Assoc*, 1987, 82(398): 387

[30] Imbens G W. Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev Econ Stat*, 2004, 86(1): 4

[31] Robins J M, Rotnitzky A, Zhao L P. Estimation of regression coefficients when some regressors are not always observed. *J Am stat Assoc*, 1994, 89(427): 846

[32] Hullsiek K H, Louis T A. Propensity score modeling strategies for the causal analysis of observational data. *Biostatistics*, 2002, 3(2): 179

[33] Pearl J. *Causality*: *Models, Reasoning, and Inference,* 2nd Ed. New York: Cambridge University Press, 2009

[34] Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Francisco: Morgan Kaufmann, 1988

[35] Spiegelhalter D J, Dawid A P, Lauritzen S L, et al. Bayesian Analysis in Expert Systems. *Statistical Science*, 1993, 8(3): 219

[36] Blyth C R. On Simpson's paradox and the sure-thing principle. *J Am Stat Assoc*, 1972, 67(338): 364

[37] Lauritzen S L, Spiegelhalter D J. Local computations with probabilities on graphical structures and their application to expert systems. *J Royal Stat Soc*:*Ser B* (*Methodol*), 1988, 50(2): 157

[38] Pearl J. *Causality.* New York: Cambridge University Press, 2000

[39] Pearl J. Causal diagrams for empirical research. *Biometrika*, 1995, 82(4): 702

[40] Greenland S, Pearl J. Adjustments and their consequences—collapsibility analysis using graphical models. *Int Stat Rev*, 2011, 79(3): 401

[41] Greenland S, Pearl J, Robins J M. Causal diagrams for epidemiologic research. *Epidemiol* (*Camb Mass*), 1999, 10(1): 37

[42] Zhao X L. *Most Useful Science of Econometrics*. Beijing: Peking University Press, 2017
(赵西亮. 基本有用的计量经济学. 北京: 北京大学出版社, 2017)

[43] Shimizu S, Hoyer P O, Hyvarinen A, et al. A linear non-gaussian acyclic model for causal discovery. *J Mach Learn Res*, 2006, 7(4): 2003

[44] Hoyer P O, Shimizu S, Kerminen A J, et al. Estimation of causal effects using linear non-Gaussian causal models with hidden variables. *Int J Approx Reason*, 2008, 49(2): 362

[45] Cai R C, Qiao J, Zhang Z J, et al. Self: Structural equational

embedded likelihood framework for causality discovery // *The Thirty-Second AAAI Conference on Artificial Intelligence(AAAI-18)*. New Orleans, 2018: 1787

[46] Mai G Z, Peng S G, Hong Y H, et al. Causation inference based on combining additive noise model and conditional independence. *Appl Res Comput*, 2019, 36(6): 1688
(麦桂珍, 彭世国, 洪英汉, 等. 混合加噪声模型与条件独立性检测的因果方向推断算法. 计算机应用研究, 2019, 36(6): 1688)

[47] Fei N N, Yang Y L. Estimating linear causality in the presence of latent variables. *Clust Comput*, 2017, 20(2): 1025

[48] He Y, Jia J, Yu B. Reversible MCMC on Markov equivalence classes of sparse directed ayclic graphs. *Ann Statist*, 2013, 41(4): 1742

[49] Zhang H, Hao Z F, Cai R C, et al. High dimensional causality discovering based on mutual information. *Appl Res Comput*, 2015, 32(2): 382
(张浩, 郝志峰, 蔡瑞初, 等. 基于互信息的适用于高维数据的因果推断算法. 计算机应用研究, 2015, 32(2): 382)

[50] Qin Q L, Li Q, Yan X X, et al. A comparative study on the four causal diagram models for causal inference in observation study. *Chin J Heal Stat*, 2020, 37(4): 496
(覃青连, 李峤, 颜星星, 等. 四种因果图模型在观察性研究因果推断中的比较研究. 中国卫生统计, 2020, 37(4): 496)

[51] Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *PNAS*, 2016, 113(27): 7345

[52] Abadie A, Cattaneo M D. Econometric methods for program evaluation. *Annu Rev Econ*, 2018, 10(1): 465

[53] Wyss R, Ellis A R, Brookhart M A, et al. The role of prediction modeling in propensity score estimation: An evaluation of logistic regression, bCART, and the covariate-balancing propensity score. *Am J Epidemiol*, 2014, 180(6): 645

[54] Imai K, Ratkovic M. Covariate balancing propensity score. *J R Stat Soc B*, 2014, 76(1): 243

[55] Bloniarz A, Liu H Z, Zhang C H, et al. Lasso adjustments of treatment effect in randomized experiments. *PNAS*, 2016, 113(27): 7383

[56] Choudhury P, Allen R, Endres M. Developing theory using machine learning methods. *SSRN Journal*, 2018: 3251077

[57] Zivich P N, Breskin A. Machine learning for causal inference: on the use of cross-fit estimators. *Epidemiology*, 2021, 32(3): 393

[58] Wang T, Huang J Q, Zhang H W, et al. Visual commonsense representation learning *via* causal inference // 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*). *Seattle*, 2020: 1547

[59] Chen P, Li Q, Zhang D Z, et al. A survey of multimodal machine learning. *Chin J Eng*, 2020, 42(5): 557
(陈鹏, 李擎, 张德政, 等. 多模态学习方法综述. 工程科学学报, 2020, 42(5): 557)

[60] Niu Y, Tang K H, Zhang H W, et al. Counterfactual VQA: A cause-effect look at language bias // 2021 *IEEE/CVF Computer Vision and Pattern Recognition*. Virtual, 2021: 12695

[61] Tang K H, Huang J Q, Zhang H W. Long-tailed classification by keeping the good and removing the bad momentum causal effect // *Neural Information Processing Systems*. Vancouver, 2020: 12991

[62] Zhang D, Zhang H W, Tang J H, et al. Causal intervention for weakly-supervised semantic segmentation // *Neural Information Processing Systems*. Vancouver, 2020: 12547

[63] Yue Z Q, Wang T, Zhang H W, et al. Counterfactual zero-shot and open-set visual recognition [J/OL]. *arXiv preprint* (2021-3-1) [2021-6-11].https://arxiv.org/abs/2103.00887v1

[64] Yang X, Zhang H W, Qi G J, et al. Causal attention for vision-language tasks // *Computer Vision and Pattern Recognition*. Virtual, 2021: 9842

[65] Bonner S, Vasile F. Causal embeddings for recommendation// *Proceedings of the* 12*th ACM Conference on Recommender Systems*. Vancouver, 2018: 104

[66] Xu S Y, Li Y Q, Liu S C, et al. Learning post-hoc causal explanations for recommendation [J/OL]. *arXiv preprint* (2021-2-23) [2021-6-11].https://arxiv.org/abs/2006.16977