发现因果的算法：被忽视的计算社会学工具箱

乔天宇　赵一璋

摘　要　社会学者长期以来聚焦于因果识别的方法及其应用，却对因果发现的相关进展缺乏关注。在回顾因果推断两种理论传统的基础上，从算法原理、应用路径和方法关联三个方面，系统介绍因果发现的经典算法，探讨它们在研究实践中的应用方式，以及与其他计算社会学方法结合的可能性。面对大语言模型的崛起，因果发现算法的应用将赋予计算社会学以新的学科能力，助力数据驱动的知识生产，为计算社会学更深入理解和解释复杂社会现象、探索因果关系提供丰富的算法工具箱，也对推动数智时代人文社会科学知识生产范式变革具有深远意义。

关键词　因果推断　因果发现　算法工具　数据驱动　计算社会学

作者乔天宇，北京大学社会学系助理教授（北京 100871）；赵一璋，清华大学社会科学学院副教授（北京 100084）。

中图分类号 C91　　**文献标识码** A　　**文章编号** 0439-8041(2025)06-0107-16

2022 年底，OpenAI 公司发布 ChatGPT，大语言模型由此进入人们的视野。它如同一股强劲的浪潮，不仅在日常工作和生活中不断渗透，还将对科学研究活动产生深远影响，甚至可能引发科学研究范式的根本性变革。可以预见，大语言模型将在社会科学研究中扮演愈发重要的角色。[①] 然而，在社会学者畅想大语言模型即将改变社会科学研究未来的同时，他们对计算社会学的兴趣和关注似乎正在降低，这可能要归因于计算社会学过去聚焦于大数据的传统。人们惊叹于大语言模型的表现，但也意识到对数据进行处理与挖掘的能力或将被大语言模型替代，由此，难免要反思在大语言模型崛起的时代，计算社会学价值何在的问题。

计算社会学自兴起以来，数据挖掘是其中重要研究路径之一。它主要致力于应用机器学习算法，从高维度、大规模数据中提取信息与知识，发现社会模式与社会规律。[②] 这种基于大数据、从数据出发、由数据驱动的研究路径给社会学研究带来了新希望。不同于以统计推断（即得到参数的无偏估计并进行统计检验）作为主要目标的传统统计分析技术，机器学习的主要目标在于追求更准确的预测，并寻找泛化能力更强（即有较强样本外预测能力）的模型。这能避免传统统计显著性检验的局限性以及人们长期诟病的 p 值

① Bail, C. A., "Can Generative AI improve social science?" *Proceedings of the National Academy of Sciences*, 121(21), 2024, e2314021121.

② 邱泽奇：《数字社会与计算社会学的演进》，《江苏社会科学》2022 年第 1 期；Molina, M., and F. Garip, "Machine Learning for Sociology," *Annual Review of Sociology*, 45, 2019, pp. 27-45.

操纵问题。[①] 另外，很多机器学习算法（如决策树、随机森林、神经网络）善于挖掘非线性关系，也能弥补常用的回归类工具囿于分析线性关系的局限。国内有学者认为，机器学习在社会学领域的应用将开启一种社会预测的新范式，让社会学重拾对可预测性的关注。[②]

然而，可解释性差是应用机器学习算法时面临的主要挑战。模型预测能力的提升往往会以牺牲可解释性为代价。应用随机森林、梯度提升等集成学习算法及神经网络模型通常能得到更好的预测效果，但这些算法透明性差，人们难以理解它们是如何做出预测决策的。由于这些模型本身不具备可解释性，它们也常被称为“黑盒模型”。追求解释和理解的社会科学研究者在使用这些模型研究社会现象时，常会面临预测准确性与可解释性之间的张力。于是，如何在此类研究中整合解释与预测，更好地发挥机器学习的优势，引发了不少的讨论与关注。[③] 近年来，计算机科学家在可解释人工智能研究领域取得了不小的进展，以夏普利加性解释（SHapley Additive exPlanations，SHAP）为代表的模型后解释方法（Post-hoc Interpretation Method）能在一定程度上解决“黑盒”机器学习算法的局限性，帮助人们理解算法的预测决策。[④] 将这些进展应用到社会学研究中能够帮助研究者以数据驱动的方式确定合适的函数形式[⑤]，进而更好地挖掘数据中蕴含的知识，辅助理论知识的生产[⑥]。然而尽管如此，应用机器学习算法得出的关联性结论归根到底还只是相关关系，即便能够实现准确预测并同时具备可解释性，也并不能据此推断因果。可预测不构成因果，这是以机器学习算法为主导的计算社会学难以避免的最大局限所在。

探求因果关系是包括社会科学在内的一切科学追求的目标，社会学也不例外。社会学自19世纪末诞生之日起，就对因果解释有着孜孜不倦的追求，计算社会学也继承了社会学对因果的关注。过去几十年中，在多学科共同努力下，因果推断从方法论到分析工具的发展都令人瞩目。从事计算社会学的研究者积极拥抱因果推断的研究进展，也在努力探索将机器学习与因果推断结合应用的方式。[⑦] 但遗憾的是，还有一些因果推断的重要进展尚未进入社会学乃至整个社会科学研究者的视野。

因果推断的研究可概括为两条路径：第一条路径是因果识别（causal identification），第二条路径是因果发现（causal discovery）。在因果识别的研究中，因果关系多被预先定义，研究者一般会事先建立关于两个变量间存在因果效应的理论假设，并以识别其中自变量对因变量的因果效应为主要研究目标。与之不同的是，在因果发现的研究中，研究者往往不预先定义因果关系，而是试图直接从观察数据（也可以是观察与实验的混合数据）中恢复生成数据的底层因果模型。另外，因果发现的主要目标通常也不仅仅针对特定的一对因变量和自变量，而是希望能对多个变量之间的因果关系给出整体性推断，得到包括多个因果关系在内的结构网络。当研究者拥有大量观测变量，但对变量间因果关系缺乏明确理论假设时，或因伦理及其他现实条件制约，无法以随机对照实验或自然实验等方式开展因果识别研究时，因果发现将大有用武之地。由于因果发现主要运用计算机科学中的搜索算法自动寻找变量间的因果关系结构，它也被称为因果结

① Nuzzo, R., “Scientific method: Statistical errors,” *Nature*, 506(7487), 2014, pp. 150–152; Hofman, J. M., Sharma, A., and D. J. Watts, “Prediction and explanation in social systems,” *Science*, 355(6324), 2017, pp. 486–488.

② 陈云松、吴晓刚、胡安宁、贺光烨、句国栋：《社会预测：基于机器学习的研究新范式》，《社会学研究》2020年第3期。

③ Hofman, J. M., D. J. Watts, S. Athey, F. Garip, T. L. Griffiths, J. Kleinberg, H. Margetts, S. Mullainathan, M. J. Salganik, S. Vazire, A. Vespignani, and T. Yarkoni, “Integrating explanation and prediction in computational social science,” *Nature*, 595(7866), 2021, pp. 181–188; Shrestha, Y. R., V. F. He, P. Puranam, and G. von Krogh, “Algorithm Supported Induction for Building Theory: How Can We Use Prediction Models to Theorize?” *Organization Science*, 32(3), 2021, pp. 856–880.

④ Lundberg, S. M., and S. -I. Lee, “A unified approach to interpreting model predictions,” *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 4768–4777; Allen, G. I., L. Gan, and L. Zheng, “Interpretable Machine Learning for Discovery: Statistical Challenges and Opportunities,” *Annual Review of Statistics and Its Application*, 11, 2024, pp. 97–121.

⑤ Verhagen, M. D., “Incorporating Machine Learning into Sociological Model-Building,” *Sociological Methodology*, 54(2), 2024, pp. 217–268.

⑥ 陈苗、陈云松：《计算扎根：定量研究的理论生产方法》，《社会学研究》2023年第4期。

⑦ 胡安宁、吴晓刚、陈云松：《处理效应异质性分析——机器学习方法带来的机遇与挑战》，《社会学研究》2021年第1期；Brand, J. E., X. Zhou, and Y. Xie, “Recent Developments in Causal Inference and Machine Learning,” *Annual Review of Sociology*, 49, 2023, pp. 81–110.

构搜索，或因果结构学习。

近年来，在科学智能（AI for Science）的潮流下，自然科学对因果发现路径给予了广泛关注。比如，有物理学研究者使用全球气候观测数据重建指标间的有向因果网络，恢复气候系统中重要的相互作用关系，并认为因果发现算法会成为未来预测气候变化的有效工具。[①] 有遗传生物学研究者利用因果发现算法尝试从观察数据寻找基因调控网络，确定基因与表型之间的因果关系。[②] 临床医学家过去曾凭借随机对照实验积累了不少关于生物标志物与复杂疾病表现间因果关系的认知，在重建这些符合“金标准”的因果结构上，因果发现算法有很好的表现。有研究利用因果发现算法，在包含阿尔茨海默症生物标志物的观察数据中，成功生成了生物标志物和阿尔兹海默症临床诊断间的因果关系模型，研究结果与实验研究中建立的因果结论高度匹配。[③]

然而，在社会科学研究中，因果发现并没有得到足够的关注。[④] 社会科学各学科过去更多关注并应用的主要是因果识别方法，相关文献中谈及的因果推断也基本等同于因果识别。目前只有国外少数经济学者和开展跨学科行为研究的学者对因果发现给予了一定关注[⑤]，对绝大多数的社会科学研究者，尤其是国内的社会学研究者而言，因果发现还是一个非常陌生的话题。

因果发现算法的起源最早可追溯到朱迪·珀尔与其同事 1987 年发表的论文。[⑥] 事实上，早期还曾有因果发现算法的经典文献发表在社会学专业期刊上。[⑦] 然而，在过去二十多年里，除少数经验研究应用[⑧]，因果发现方法很少被社会科学学者了解与学习。这背后的主要原因在于，因果发现涉及大量复杂且专业化的术语，主要采用的算法语言也与社会科学研究者熟知的统计语言有一定差异。本文希望以相对通俗的方式介绍因果发现的经典算法及其应用，主要内容将包括以下几个方面：首先，将概述当前因果推断研究的两大理论框架，并在此基础上回答“因果发现是什么”的问题，对开展因果发现研究的基础前提以及相关算法的基本原理做简要讨论。其次，将回答“因果发现有什么用”的问题，结合近五年发表的经验研究文献，总结因果发现应用于计算社会学研究的可能路径，以及在应用中可能遇到的问题及解决方案。最后，将回答“因果发现与其他方法有何关系”的问题，针对计算社会学研究者熟悉的机器学习、基于行动者建模以及近来兴起的大语言模型工具，探讨如何将它们与因果发现相结合。因果发现算法的应用可能会开辟一条不同于由机器学习算法主导的数据驱动知识生产的研究路径，这将赋予计算社会学以新的学科能力，令其在大语言模型崛起的时代具备无法被替代的价值。

一、两种传统：因果推断的理论框架

自 20 世纪 70 年代始，因果推断领域逐渐形成了两种传统，分别来自统计学家和计算机科学家，背后

---

① Nowack, P., J. Runge, V. Eyring, and J. D. Haigh, “Causal networks for climate model evaluation and constrained projections,” *Nature Communications*, 11(1), 2020, p. 1415.

② Zhang, X., X. -M. Zhao, K. He, L. Lu, Y. Cao, J. Liu, J. -K. Hao, Z. -P. Liu, and L. Chen, “Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information,” *Bioinformatics*, 28(1), 2012, pp. 98–104.

③ Shen, X., S. Ma, P. Vemuri, and G. Simon, “Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer’s Pathophysiology,” *Scientific Reports*, 10(1), 2020, p. 2975.

④ Leist, A. K., M. Klee, J. H. Kim, D. H. Rehkopf, S. P. A. Bordas, G. Muniz-Terrera, and S. Wade, “Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences,” *Science Advances*, 8(42), 2022, eabk1942; Huber, M., “An introduction to causal discovery,” *Swiss Journal of Economics and Statistics*, 160(1), 2024, p. 14.

⑤ Addo, P. M., C. Manibialoa, and F. McIsaac, “Exploring nonlinearity on the CO_2 emissions, economic production and energy use nexus: A causal discovery approach,” *Energy Reports*, 7, 2021, pp. 6196 – 6204; Sheffrin, S., and R. Zhao, “Public perceptions of the tax avoidance of corporations and the wealthy,” *Empirical Economics*, 61(1), 2021, pp. 259–277.

⑥ Rebane, G., and J. Pearl, “The Recovery of Causal Poly-Trees from Statistical Data,” *Proceedings of the Third Conference on Uncertainty in Artificial Intelligence*, 1987, pp. 222–228.

⑦ Spirtes, P., T. Richardson, C. Meek, R. Scheines, and C. Glymour, “Using Path Diagrams as a Structural Equation Modeling Tool,” *Sociological Methods & Research*, 27(2), 1998, pp. 182–225.

⑧ Glymour, B., C. Glymour, and M. Glymour, “Watching Social Science: The Debate About the Effects of Exposure to Televised Violence on Aggressive Behavior,” *American Behavioral Scientist*, 51(8), 2008, pp. 1231–1259.

对应着理解因果的两种理论框架——潜在结果模型（Potential Outcomes Model）和结构因果模型（Structural Causal Model）。目前，几乎所有的因果推断研究都建立在这两大理论框架之上。

潜在结果模型，又称鲁宾因果模型，是统计学家唐纳德·鲁宾于20世纪70年代提出的用于形式化定义因果关系并进行因果推断的理论框架。该框架最初建立在实验思维的基础之上。实验的核心是在受控条件下对被试施加某种干预 T，然后在 T 的不同状态下比较结果变量 Y 的差异。潜在结果模型认为，实验中的每个被试个体 i 存在两种潜在结果，分别是个体 i 未接受干预（$T_i=0$）状态下的潜在结果 Y_{0i}，以及接受干预（$T_i=1$）状态下的潜在结果 Y_{1i}。对于某个体 i 而言，实际观测到的结果可表示为，$Y_i=T_iY_{1i}+(1-T_i)Y_{0i}$，干预 T 的因果效应即 $Y_{1i}-Y_{0i}$。但问题在于，个体无论接受或未接受干预，Y_{0i} 和 Y_{1i} 当中都只有一个可以被实际观测到，另一个对个体而言是一种反事实状态。因此，研究者主要通过在群体层面，用接受干预组与未接受干预组结果的平均差异来计算平均的因果效应。用组别均值计算因果效应需要满足一个可忽略干预分配假定，即研究对象在不同干预状态下的分配要独立于两种潜在结果。随机对照实验能够满足这一假定，于是被视为因果推断的“金标准”。

潜在结果模型后被推广并应用在非实验的观察研究中。在过去五十年里，统计学者和计量经济学者基于该理论框架开发出很多种用于观察数据因果推断的统计分析方法，包括倾向值匹配、工具变量、差分法、断点回归等。[①] 与另一种传统的结构因果模型相比，潜在结果模型更为社会科学研究者所熟知并广泛应用[②]，特别是用于评估政策影响的研究中[③]。然而，潜在结果模型更多作为因果识别的理论基础，主要适用于有明确理论假设，并需要根据假设估计因果效应的研究，这往往要求研究者对生成数据的底层模型有一定的理解和把握。

结构因果模型，是计算机科学家朱迪·珀尔与其合作者提出的，使用有向无环图（Directed Acyclic Graph，DAG）来形式化变量之间因果关系的因果推断理论框架。结构因果模型 M 由一个三元组定义，该三元组包含两个变量集合和一个函数集合：

$$M=\langle U,V,F\rangle$$

其中，U 是外生变量集合，即不受模型中其他变量影响的变量集合；V 是内生变量集合，即受到模型中其他变量影响的变量集合；F 是一组函数的集合，记为 $\{f_1,f_2,\cdots,f_n\}$，其中每个 f_i 是将属于内生变量集合 V 中的变量 i 表示为模型中其他变量的函数。这一定义或许还比较抽象，但结构因果模型的一大优势在于，其可借用图论的语言，用更形象的方式以有向无环图 $G(M)$ 表示模型 M，由此也被称为因果图模型。图中每个节点代表 M 中的一个变量，即 U 和 V 中的变量。每条边表示对应变量之间的直接因果关系。对于属于 V 的变量 Y，如果 f_Y 的定义域中含有变量 X，那么 $G(M)$ 中会有一条从节点 X 指向节点 Y 的有向边，表示 X 是 Y 的直接原因，同时称 X 是 Y 的父代节点，Y 是 X 的子代节点。

图1 结构因果模型的有向无环图表示

图1中给出了两个示例。对图1（a）来说，内生变量集合为 $V=\{A,B\}$，外生变量集合为 $U=\{C\}$；

① 许琪：《因果推断五十年：成就、挑战与应对》，《学术月刊》2024年第11期。

② 余静文、王春超：《新“拟随机实验”方法的兴起——断点回归及其在经济学中的应用》，《经济学动态》2011年第2期；胡安宁：《倾向值匹配与因果推论：方法论述评》，《社会学研究》2012年第1期；陈云松：《逻辑、想象和诠释：工具变量在社会科学因果推断中的应用》，《社会学研究》2012年第6期。

③ 李文钊：《因果推理中的潜在结果模型：起源、逻辑与意蕴》，《公共行政评论》2018年第1期。

函数f_B由变量A和C定义，即A和C是B的直接原因，在图上表示为A和C分别指向B的有向边，即$A\rightarrow B$和$C\rightarrow B$；函数f_A由变量C定义，即C是A的直接原因，在图上表示为从C指向A的有向边，即$C\rightarrow A$。同理，对于图1（b）来说，内生变量集合为$V=\{F, G\}$，外生变量集合为$U=\{D, E\}$；函数f_G由变量F定义，即F是G的直接原因，在图上表示为$F\rightarrow G$；函数f_F由变量D和变量E定义，即D和E是F的直接原因，图上表示为$D\rightarrow F$和$E\rightarrow F$。

与潜在结果模型相比，结构因果模型更直观地表示了变量间的因果关系。然而，如何利用结构因果模型实现因果推断，在过去较长一段时间内并不被社会科学研究者所熟悉。在十多年前，国外开始有文献专门面向社会科学研究者介绍结构因果模型。[①] 近年国内也有社会学者注意到了这一缺失，专门撰文介绍结构因果模型。他们认为结构因果模型一方面有助于研究者更好掌握因果推断的框架，如理解内生性问题并选择合适的因果推断方法；另一方面也有助于研究者厘清一些模糊认知，如怎样确定混淆变量，如何选择控制变量进而优化研究设计等。[②] 然而，这些还都是将结构因果模型用于因果识别的努力，无论是改进研究设计还是辅助选取变量，都还只是聚焦于估计两个特定变量间的因果效应，限于局部因果结构的探讨。事实上，除了对因果识别的贡献之外，结构因果模型还为因果发现的研究路径奠定了理论基础，使研究者能从丰富的观察数据中挖掘因果关系、构建整体的因果结构。

二、原理概述：因果发现的假设与算法

因果发现的研究路径希望利用已有经验数据，尤其是观察数据，根据其中的信息和线索尽可能地还原出生成这些数据的底层因果模型。然而，要实现这一目标并不容易，其背后的很多道理也不显见。以珀尔、斯皮尔斯、格利莫尔等为代表的一众学者在结构因果模型的基础上，围绕因果发现的可行性及其实现路径做了大量的基础理论和算法开发工作。[③] 这里将先简要概述支撑因果发现的前提假设。

（一）作为前提的桥梁原则

一般认为，数据中体现出来的统计关联是由因果关联产生的，这为从数据中恢复因果关联提供了一定的线索。然而，从观测数据中得到的还只是概率分布信息，但概率模型与因果模型仍有着本质差异，二者并不能天然地等同。要想将二者联系起来，需要借助一些特定的“桥梁原则”——因果马尔可夫条件、*d*-分离准则和因果忠诚性假设。它们作为前提，构成了从观测数据中发现因果的基础。

第一个前提为因果马尔可夫条件。这是指在因果图中，每个变量仅由图上它的父代变量（即直接影响因素）决定，在控制父代变量的条件下，每个变量都独立于其他所有非后代变量。以图1（b）中G和D为例，这两个变量在因果图上没有直接连接，如果某观测数据是由这个因果图生成的，那么在控制G的父代变量F的条件下，G与D应相互独立，这才满足因果马尔可夫条件。用更通俗的话来说，因果马尔可夫条件意味着，如果观测数据是以某个因果图为底层模型生成的，那么因果图结构中显示的独立性会反映在观测数据上。

第二个前提是*d*-分离准则。它是用图论语言表述的一种判别变量间条件独立的准则，利用这一准则，我们可以根据因果图的结构特征，揭示变量间的条件独立性关系。具体来说，如果X和Y是有向无环图中的两个不同节点，W是该图中不包含X和Y的节点集，当且仅当X和Y之间每一条通路（这里所谓通路不考虑边的方向，且通路中不经过重复节点）都被W阻断，则X和Y是*d*-分离的（被W分离开的）。通路被阻断的情况包括：（1）该通路包含链结构$X\rightarrow Z\rightarrow Y$或分叉结构$X\leftarrow Z\rightarrow Y$，且中间节点$Z$在节点集$W$中；（2）该通路包含一个对撞结构$X\rightarrow Z\leftarrow Y$，且中间节点$Z$及其后代节点都不在$W$中。仍以图1（b）举例，

---

① Elwert, F., “Graphical Causal Models,” in *Handbook of Causal Analysis for Social Research*, S. L. Morgan (ed.), Springer Netherlands, 2013, pp. 245-273.

② 句国栋、陈云松：《图形的逻辑力量：因果图的概念及其应用》，《社会》2022年第3期。

③ Pearl, J., *Causality: Models, reasoning, and inference*, Cambridge University Press, 2000; Spirtes, P., C. N. Glymour, and R. Scheines, *Causation, prediction, and search (2nd ed)*, MIT Press, 2000.

其中 D 和 G 之间的通路只有 $D—F—G$，它包含一个链结构，节点集 $\{F\}$ 和 $\{E, F\}$ 都会阻断该通路，将 D 和 G 两节点 d-分离；D 和 E 之间的通路只有 $D—F—E$，它包含一个对撞结构，并不存在一个 F 和 G（F 的后代节点）都不在其中的节点集合，故 D 和 E 不满足 d-分离准则。

有了因果马尔可夫条件和 d-分离准则之后，只要给出任意一个因果图，我们就可以推断出观测数据中哪些变量之间应在给定条件下相互独立，也就是说，如果两个变量在因果图上被 d-分离，那么这两个变量在观测数据中应存在条件独立性。因果马尔可夫条件和 d-分离准则都是利用观测数据恢复因果结构的重要基石，然而，对于很多因果算法而言，只有这两个前提尚不足够，还需要第三个前提，即因果忠实性假设。

因果忠实性，指的是观测数据中所有条件独立关系必须是由因果结构决定的，而不是来自因果结构之外的巧合。换句话说，若两变量在因果图上没有被 d-分离，那么它们在观测数据中就不能是条件独立的。如果说因果马尔可夫条件和 d-分离准则的结合是为了从结构推断数据（结构⇒数据），因果忠实性假设保证了我们能从观测数据中的统计关系推断因果图的结构（数据⇒结构），当观测数据中两变量条件独立时，则这两个变量在因果图上被 d-分离。

这三个前提是大多数因果发现算法（尤其是后面将讨论的基于约束的算法和基于分数的算法）的理论基石，也构成了将因果模型与观测数据连接起来的“桥梁”。除这三个基础假设外，还有一些因果发现算法需要依赖的其他重要假设：如因果充分性假设，它要求用于因果发现的观测数据是完整的，不存在未被观测到的潜在混淆变量；再如，大多数因果发现算法都会有无环性假设的要求，即因果图中表示变量间因果关系的有向边不能形成闭环；还有些算法会对数据分布和数据类型有一定的要求。不过，已有相应的改进算法可以有条件地放宽这些假设的限制，实现对复杂数据更灵活的因果发现。

（二）因果发现的算法工具箱

在计算机科学家三十多年的不懈努力下，目前开发出的因果发现算法已有几十种之多。其中，经典的因果发现算法大致可以分为三类，分别是基于约束的因果发现算法、基于分数的因果发现算法以及函数因果发现算法。

1. 基于约束的因果发现算法。

基于约束的因果发现算法是利用 d-分离准则，通过检验变量之间条件独立性的方式从数据中恢复变量间的因果关系结构。最具代表性的基于约束的算法是斯皮尔斯和格利莫尔在1991年提出并以二人名字命名的PC算法（Peter-Clark Algorithm）[①]，它改进了珀尔和同事最早提出的IC算法（Inductive Causation Algorithm）。

PC算法从一个假定了所有变量两两间均有边连接的完全图开始，通过不断去掉那些不太可能存在因果关系的边，最后得到有可能表示变量间因果结构的图。这就像是一位雕塑师，从一块完整的材料开始，逐渐修剪掉上面无用的部分。所谓不太可能存在因果关系的边，就是那些相互独立，或在一定条件下相互独立的变量之间的连边。更具体来讲，PC算法包括两个主要步骤。第一步是搭建骨架，即逐对检验变量间是否独立或条件独立，若是，则删除图上的对应边。在完成骨架搭建后，会得到一个无向图，此时还不能确定变量间因果关系的方向，因此还要进行第二步，即确定方向。PC算法此时直接利用 d-分离准则，分两步确定图中边的方向：首先，对上述无向图中所有类似 $X—Y—Z$ 的变量三元组（“—”代表变量间存在无向边，其中 X 和 Z 之间不存在直接连边），验证 X 和 Z 是否会在给定 Y 的条件下相互独立。这里又分为两种情况：第一，如果发现在控制了变量 Y 后，X 和 Z 有统计关联，即条件独立性不成立，说明 Y 是一个对撞变量，则确定方向为 $X→Y←Z$；第二，如果发现在控制了变量 Y 后，X 和 Z 相互独立，说明 Y 不是一个对撞变量，但此时还不能确定变量间的因果关系方向。这一步骤被称为 V-结构定向。其次，在完成上

① Spirtes, P., and C. Glymour, “An Algorithm for Fast Recovery of Sparse Causal Graphs,” *Social Science Computer Review*, 9(1), 1991, pp. 62–72.

述步骤后得到的图上，继续应用一些规则，确定图中剩余无向边的方向。例如对类似 $X\rightarrow Y\text{—}Z$ 的变量三元组（Y 和 Z 之间的方向未被确定，且 X 和 Z 之间不存在直接连边），将 Y 与 Z 之间的方向都确定为 $Y\rightarrow Z$。这是因为，如果将其确定为 $Y\leftarrow Z$，Y 会成为对撞变量，这与已知给定 Y 条件下 X 和 Z 独立相矛盾。图 2 以具体示例演示了 PC 算法的工作原理。

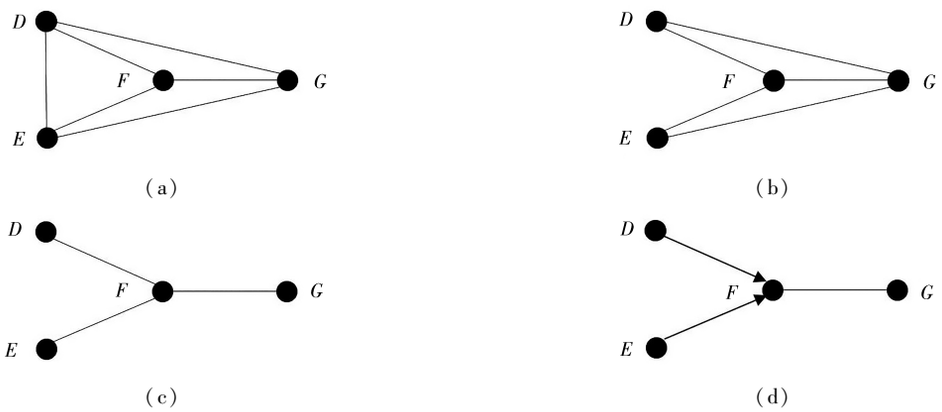

图 2　PC 算法工作原理示例

注：以图 1（b）所示因果图为例，说明 PC 算法的工作步骤。首先，从四个节点两两相互连接的完全图（a）开始，先在不控制其他变量的条件下，逐对检验两两变量间是否独立。结果发现，只有 D 与 E 相互独立，于是删除 $D\text{—}E$ 边，得到（b）。其次，逐对检验变量间是否在控制其他变量（或者变量集合）的条件下独立。由于 D 与 G 和 E 与 G 之间的相关性均在控制了 F 之后消失，于是删除 $D\text{—}G$ 和 $E\text{—}G$ 边，得到（c）。至此，完成了骨架搭建步骤，但得到的仍是无向图。基于此，利用 d-分离准则，针对 $D\text{—}F\text{—}E$，检验在控制 F 后，D 和 E 是否有统计关联，结果发现关联性存在，F 是对撞变量，于是确定方向为 $D\rightarrow F\leftarrow E$（d）。最后，根据 $X\rightarrow Y\text{—}Z$ 规则确定 $F\rightarrow G$，得到图 1（b）所示的因果图。

在图 2 示例中，所有边的方向都得到了确定。然而，在实际应用 PC 算法时，很多时候最终得到的因果图中未必每条边都能确定方向。这是因为，一些结构具有相同的 d-分离性质，在进行 V-结构定向时无法对它们加以区分，比如 $X\rightarrow Y\rightarrow Z$ 和 $X\leftarrow Y\leftarrow Z$，这些无法区分的结构被称为马尔可夫等价类（Markov Equivalence Class）。PC 算法输出的结果一般是能够表示马尔可夫等价类的部分有向无环图，而不是确定的有向无环图。部分有向无环图中包含的无向边表示两变量间的因果方向无法确定，如 $X\text{—}Y$，既可能是 $X\rightarrow Y$，也可能是 $X\leftarrow Y$。

在 PC 算法中，条件独立性检验是关键环节，它直接决定了因果发现的结果。根据不同的数据类型，条件独立性检验的方法也有所不同：连续变量之间一般考察偏相关系数，离散变量之间则运用皮尔逊卡方检验或者似然比卡方检验，对于非线性相关的情况可利用互信息度量。条件独立性检验的显著性水平 α 是算法需要设定的重要超参数，α 设定得越高，检验越宽松，有可能导致纳入虚假的因果边；α 设定得越低，检验越保守，有可能遗漏真实的因果边，导致生成的因果图不完整。

PC 算法后续还发展出了很多改进版本，以应对一些潜在挑战。比如，PC-Stable 改进了原始算法对条件独立性检验顺序敏感的问题。再比如，标准的 PC 算法无法处理同时包含连续型变量和离散型变量的数据，目前也开发出了相应的改进算法。[①] 另外，PC 算法结果的正确性还取决于因果充分性假设是否满足，即是否存在未被观测的混淆变量。如果无法满足因果充分性假设，可能会导致遗漏关键因果边或引入虚假的因果边。快速因果推断（Fast Causal Inference，FCI）算法是 PC 算法的另一种扩展，放宽了因果充分性假设，即便有混淆变量未被观测到，FCI 算法发现的因果结构结果也是接近正确的[②]，其输出的结果是部

① Cui, R., P. Groot, and T. Heskes, "Copula PC Algorithm for Causal Discovery from Mixed Data," in *Machine Learning and Knowledge Discovery in Databases*, P. Frasconi, N. Landwehr, G. Manco, and J. Vreeken (eds.), Springer International Publishing, 2016, pp. 377–392; Tsagris, M., G. Borboudakis, V. Lagani, and I. Tsamardinos, "Constraint-based causal discovery with mixed data," *International Journal of Data Science and Analytics*, 6(1), 2018, pp. 19–30.

② Glymour, C., K. Zhang, and P. Spirtes, "Review of Causal Discovery Methods Based on Graphical Models," *Frontiers in Genetics*, 10, 2019.

分祖先图（Partial Ancestral Graph，PAG），它比部分有向无环图包含了更多的信息。对此，后文会予以详细介绍。

2. 基于分数的因果发现算法。

贪婪等价搜索（Greedy Equivalence Search，GES）算法是另一种经典的因果发现算法。与PC算法类似，GES算法本质上也是根据变量间的条件独立性还原因果结构图。但二者实现路径截然不同：PC算法如同雕塑师，从完全图中“剔除”被d-分离的边以恢复因果图的骨架；GES算法更像是一位在白纸上进行创作的画家，从空图开始，通过逐步“添笔”和“擦除”有向边来构建因果图。GES算法在确定是否应该生成边时，并不依靠直接检验变量间的条件独立性，而是通过贝叶斯信息准则（BIC）等传统上用来评估模型拟合的分数指标，间接评估图结构符合d-分离准则的程度，并以此作为增减边的依据。因此，GES被称为基于分数的算法。在每步迭代中，算法首先以贪婪策略添加能最大程度提升得分的边，当添加边无法进一步提高分数时，再尝试通过删除已有边进一步优化模型的拟合效果。这一做法与数据挖掘中向前与向后结合的逐步回归方法类似，都采用了启发式搜索策略。

GES算法与PC算法有很多相似之处。首先，GES算法输出的结果大多也是用部分有向无环图表示的马尔可夫等价类，其中同样可能存在无法确定因果方向的边。其次，GES算法也需要满足因果充分性假设，但它可以在一定条件下放松因果忠实性假设。GES算法也有不少优化版本，快速贪婪等价搜索（Fast GES，FGES）是被经常提及的一种，其优势在于计算速度快，适用于小样本数据且结果更准确。[①] 贪婪快速因果推断（Greedy Fast Causal Inference，GFCI）算法融合了GES与FCI算法的思想，放宽了因果充分性假设，并与FCI算法类似，GFCI也以部分祖先图作为输出结果。[②]

3. 函数因果发现算法。

使用基于约束的算法和基于分数的算法时，许多因果关系的方向无法被唯一确定，输出结果往往是马尔可夫等价类。相比之下，函数因果发现算法能克服这一局限，输出因果关系方向明确的有向无环图。这一优势的实现依赖于函数因果发现算法能识别数据中残留的某些“不对称痕迹”，从而推断因果关系的方向。

具体来说，假定X与Y之间存在真实的因果关系$X\to Y$。我们可以通过以下两种方式拟合观测数据：（1）以X为自变量，Y为因变量进行回归（模型设定的因果方向正确）；（2）反过来，以Y为自变量，X为因变量进行回归（模型设定的因果方向不正确）。如果我们发现这两种回归模型的残差与自变量之间表现出不同的依赖性——即在（1）的模型设定下，残差项与自变量（X）独立；而在（2）的模型设定下，残差项与自变量（Y）不独立——那么我们便可根据（1）确定因果关系的方向为$X\to Y$。这种残差与原因变量无关的特性，正是数据中残留的“不对称痕迹”。

然而遗憾的是，这种判定方法并非任何时候都适用。要想用这种方式确定因果关系方向，需对变量间关系的函数形式和数据分布属性做额外假定。例如，当两个变量服从正态分布且存在线性相关关系时，便无法通过这种方式确定因果方向，因为在线性—高斯条件下，上述两种模型设定都会得到残差项与自变量独立的结果。

如果变量间的关系符合非线性函数形式，数据中便会留下“不对称痕迹”。加性噪声模型（Additive Noise Model，ANM）算法便是基于这种非线性关联中的“痕迹”来判断因果方向。[③] 然而，由于变量间存在非线性关联，传统的线性相关度量方法不再适用，因此，ANM算法采用希尔伯特—施密特独立性准则（Hilbert-Schmidt Independence Criterion，HSIC）来判断残差与自变量之间是否独立。对于线性函数形式，

① Malinsky, D., and D. Danks, “Causal discovery algorithms: A practical guide,” *Philosophy Compass*, 13(1), 2018, e12470.

② Ogarrio, J. M., P. Spirtes, and J. Ramsey, “A Hybrid Causal Search Algorithm for Latent Variable Models,” *JMLR workshop and conference proceedings*, 52, 2016, pp. 368-379.

③ Hoyer, P. O., D. Janzing, J. M. Mooij, J. Peters, B. Schölkopf, “Nonlinear causal discovery with additive noise models,” *Advances in Neural Information Processing Systems*, 21, 2008.

因果方向的判定需进一步假定数据分布是非高斯的，即自变量和残差项最多只能有一个服从正态分布。线性非高斯模型（Linear Non-Gaussian Model，LiNGAM）就是在这一假定基础上对连续型数据进行因果发现的算法。

函数因果发现算法与此前讨论的基于约束的算法和基于分数的算法不同，后二者在寻找因果结构时都主要依据观测变量间的条件独立性，而函数因果发现算法则主要依据噪声（即模型残差）独立于原因变量的条件。函数因果发现算法尽管对数据分布有一定要求，但也绕过了一些假设，比如 ANM 和 LiNGAM 都不依赖于因果忠诚性假设。此外，单独使用基于约束或基于分数的算法都只能对变量之间是否存在因果关系给出定性判断，而以 LiNGAM 为代表的算法不仅能给出定性判断，还能同时估计变量间因果关系的强度，这也是函数因果发现算法相较于前两种算法的优势所在。

三、应用路径：因果发现的计算社会学实践

（一）区分影响结果的直接与间接原因

在很多情况下，同一个结果可能是由多种原因共同导致的，但这些原因在因果结构中所处的位置往往不同：有些原因是导致结果的直接原因，另外一些并不直接作用于结果，而是通过其他因素间接产生影响。直接原因被认为是近端机制，相对应地，间接原因是远端机制。然而，对于如何区分近端与远端，过去的社会科学研究更多是凭借理论演绎，甚或直觉的方式确定。应用机器学习的数据驱动研究能够基于训练得到的模型衡量预测变量的重要性，但即便筛选出的重要预测变量可以被识别为目标变量的原因，它们在因果结构中的具体位置（即近端或远端）也仍然无法被有效区分。基于输出的完整因果结构图，因果发现算法可以帮助研究者方便地区分直接与间接影响因素，作为一种识别近端和远端机制的高效工具。例如，昆塔纳利用 PC 算法和 FGES 算法挖掘了影响儿童学业成就因素的有向无环图。在分析中，他纳入了此前研究中涉及的重要影响因素，包括人口属性、儿童心理、家庭背景、学校和邻里环境因素等 32 个变量，最终筛选出 8 个直接影响语文成绩的变量和 7 个直接影响数学成绩的变量，对影响结果的近端机制进行了有效识别。①

在众多因果发现算法中，不变因果预测（Invariant Causal Prediction，ICP）是一种专门用于寻找“近端机制”的算法。该算法旨在针对一个特定的目标变量（Y），从一系列变量中找到影响 Y 的直接原因。其核心思想是：假设存在两个不同的环境（$E=0$ 和 $E=1$，当然环境数量可以多于两个，这里为简化讨论仅以两个为例）②，在控制一组候选的原因变量 S 后，如果目标变量 Y 的条件均值在这两个环境中保持不变（即 Y 与 E 条件独立），那么 S 中就包含对 Y 有直接影响的变量。通过遍历数据中所有可能的候选变量组合，筛选出能满足上述不变性检验的所有 S，最后将这些 S 取交集，即可得到直接影响 Y 的变量（或变量集合）。③ 胡贝尔在一个美国残障青年训练项目的实验数据集上展示了 ICP 算法的应用。该实验通过随机分配的方式让样本中的残障青年加入某个康复训练项目，这原本希望用来识别参与康复训练对残障青年几年后健康状况的因果效应。事实上，没能被分配进项目的残障青年中同样也会有人参与训练，但进入项目会大大增加他们参与训练的可能性。该研究将随机分配作为环境变量 E，并应用 ICP 算法寻找其他影响残障青年健康状况的直接原因。此外，该算法还能够计算这些直接影响因素的效应值及其对应的置信区间，为因果推断提供更精确的量化支持。④

---

① Quintana, R., “The Structure of Academic Achievement: Searching for Proximal Mechanisms Using Causal Discovery Algorithms,” *Sociological Methods & Research*, 52(1), 2023, pp. 85-134.

② ICP 算法用到的环境变量应满足以下两个条件：一是环境变量 E 不能是目标变量 Y 的后代变量；二是环境变量 E 只能间接影响目标变量 Y，而不能是 Y 的直接原因，如果 E 是 Y 的直接原因，会导致 ICP 算法失效。

③ Peters, J., P. Bühlmann, and N. Meinshausen, “Causal Inference by using Invariant Prediction: Identification and Confidence Intervals,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(5), 2016, pp. 947-1012.

④ Huber, M., “An introduction to causal discovery,” *Swiss Journal of Economics and Statistics*, 160(1), 2024, p. 14.

（二）提示混淆变量的存在及潜在影响

大多数因果发现算法通常依赖因果充分性假设，即数据中未遗漏任何潜在的重要变量。然而，当该假设不被满足时，因果发现的结果可能会引入混淆偏误。FCI 和 GFCI 算法突破了因果充分性假设的限制，能够在存在未观测混淆变量的条件下进行因果发现。通过输出部分祖先图，这两种算法可以提示混淆变量的存在并揭示其在因果结构中的可能位置。

与部分有向无环图类似，部分祖先图也是一种用来表示马尔可夫等价类的图模型，但它比部分有向无环图包含更多的信息。部分祖先图中的边不仅限于有向边和无向边两种类型，还包括其他形式的边，能更全面地呈现变量之间可能存在的关系。如图3所示，图3（a）中的双向箭头表示 A 和 B 之间的关系至少受到一个混淆变量的影响，图3（b）和（c）中用○表示该处的方向是不清楚的，既可以有箭头，也可以没有箭头。因此，图3（b）表示 A 和 B 之间的关系既可能是 $A\rightarrow B$，也可能是 $A\leftrightarrow B$，后一种情况下 A 和 B 之间至少受到一个混淆变量影响。但无论怎样，B 一定不会是 A 的原因。同理，图3（c）中 A 和 B 的关系既可能是 $A\rightarrow B$，也可能是 $B\rightarrow A$，还可能是 $A\leftrightarrow B$，在这种情况下，并不存在一个变量集合能够 d-分离 A 和 B。

（a） （b） （c）

图3 部分祖先图示例

《美国社会学评论》曾经发表过一篇讨论第三世界国家的外国投资渗透率与其国内政治排斥之间关系的研究[①]，图4是斯皮尔斯等人应用 FCI 算法对该研究的原始数据进行再分析后得到的部分祖先图[②]。该图显示，外国投资渗透率与国家经济发展水平之间由双向箭头（↔）相连，说明二者至少受到一个共同原因的影响；同样，外国投资渗透率与公民自由程度之间也是如此。此外，“国内政治排斥○→国家经济发展水平”说明二者间可能存在因果关系，即政治排斥可能是国家经济发展水平的原因，但也可能是这二者间的关系受到未观测到的混淆变量的影响。类似方法也被其他一些经验研究用于提示混淆变量的存在。[③] 另外值得注意的是，斯皮尔斯等人发现应用 FCI 算法得到的结果还在很大程度上颠覆了原研究的认知。原研究以国内政治排斥为因变量，以其他三个变量作为自变量，建立回归模型进行分析，结果显示外国投资渗透率和公民自由程度的回归系数估计值为正，而国家经济发展水平的回归系数估计值为负。然而，因果发现算法的分析结果显示，外国投资渗透率与国内政治排斥之间并不存在直接的因果关系，同时公民自由程度也不太可能是国内政治排斥的原因。这一发现揭示了传统回归模型在因果推断中的局限性以及因果发现算法的重要作用。除独立应用 FCI 算法外，格利莫尔等还建议可以将 FCI 算法与其他因果发现算法提供的信息结合起来判断混淆变量是否存在。[④]

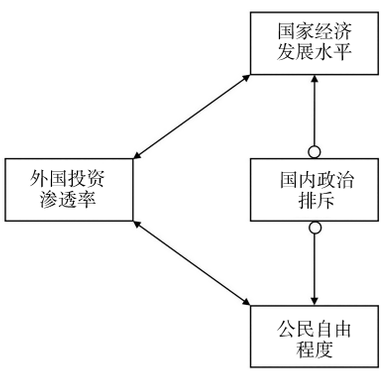

图4 斯皮尔斯等应用 FCI 算法得到的部分祖先图

（三）挖掘变量间的复杂因果关系结构

变量间的因果关系在现实中往往错综复杂，以传统的回归模型作为统计工具难以实现对复杂因果结构

① Timberlake, M., and K. R. Williams, “Dependence, Political Exclusion, and Government Repression: Some Cross-National Evidence,” *American Sociological Review*, 49(1), 1984, pp. 141–146.

② Spirtes, P., C. N. Glymour, and R. Scheines, *Causation, prediction, and search (2nd ed)*, MIT Press, 2000.

③ Andrews, K. S., M. I. Ohannessian, and E. Zheleva, “See Me and Believe Me: Causality and Intersectionality in Testimonial Injustice in Healthcare,” 2024, No. arXiv: 2410.01227; Glymour, B., C. Glymour, and M. Glymour, “Watching Social Science: The Debate About the Effects of Exposure to Televised Violence on Aggressive Behavior,” *American Behavioral Scientist*, 51(8), 2008, pp. 1231–1259.

④ Glymour, C., K. Zhang, and P. Spirtes, “Review of Causal Discovery Methods Based on Graphical Models,” *Frontiers in Genetics*, 10, 2019.

的把握。回归模型擅于分析一组解释变量与一个被解释变量之间的关系，但其估计的每个解释变量的效应仅反映其单独的贡献，而对于各解释变量间的相互作用，单个回归模型无能为力。相比之下，因果发现算法的一大优势在于其推断是整体性的，能够从数据中提取变量间复杂的因果关系结构。

鉴于因果发现算法可以构建变量间关系的整体结构，有学者提出可以在因果发现结果的基础上结合运用社会学者熟悉的网络分析方法，对复杂因果关系的结构性特征实现更深入的挖掘。[①] 比如，基于利用生成的因果网络图，研究者不仅能区分变量的影响属于近端机制还是远端机制，还可以利用网络中心性度量的方法来量化节点的重要性，从而评估各个变量在因果网络中的关键性。比如，某个因素在因果网络中有较高的出度，说明受该因素直接影响者较多，在造成影响后果的多寡方面，该因素在整体系统中是更关键的；类似地，介数中心性（Betweenness Centrality）可以用于衡量某个变量在因果网络中成为中介因素的程度，介数中心性越大，该因素更可能成为系统中的中介因素。此外，运用网络社区发现算法对因果图中的节点进行群组识别，能够揭示出内部相互强化、对外协同作用的变量集群，也有助于挖掘变量间的复杂互动关系和集体效应。

这种结合网络分析的因果发现研究被认为更契合社会与行为科学对结构性解释的追求，是一种考察人类行为的生态学方法。对完整因果结构进行挖掘分析在潜在结果框架下是不容易实现的，随机对照实验虽被誉为因果推断的“金标准”，但其引入随机分配的干预相当于切断了干预变量与因果结构其他部分的连接，因此只能孤立地评估干预变量对结果变量的影响。

（四）因果发现算法用于识别测量模型

社会科学研究中有不少用作社会分类的变量，如性别。它们看似明确，代表了个体的固有特征，但很多社会学者从建构主义而非本质主义的视角出发，主张应将其理解为具有多维性和动态性的复杂概念。另一典型是美国社会学尤其关注的种族变量，持建构主义观点的研究者认为，种族不应被简单地视为一种已预先确定且固定不变的分类标准，而应深入剖析其构成要素与多维特性。然而，在社会科学研究使用的绝大多数数据中，对类似性别和种族等特征，几乎都是采用预先确定分类的方式予以测量的。昆塔纳提出了一套使用因果发现算法识别分类变量建构关系的方法。[②]

马尔可夫毯（Markov Blanket，Mb）是图模型中的一个重要概念，它是针对有向无环图上的特定节点而定义的。对于有向无环图上的特定节点 X，它的马尔可夫毯 $Mb(X)$ 是包含 X 的父节点、子节点以及子节点的其他父节点的最小节点集合，使得在给定 $Mb(X)$ 的条件下，X 与图中所有其他节点 d-分离。如图 5 所示，其中灰色节点的集合即是节点 X 的马尔可夫毯。在机器学习领域，搜索马尔可夫毯被用作一种特征选择方法。[③] 与传统的包裹法、过滤法和嵌入法等基于相关性的特征选择方法不同，利用马尔可夫毯筛选最能表征目标变量 X 的变量集合，被认为是一种更具稳定性和可解释性的因果特征选择方法。但与特征选择不同的是，当测量 X 的潜在构成因素时，研究的重点在于马尔可夫毯中 X 的父节点（如图 5 中边框加粗的灰色节点所示）。昆塔纳采用以传统方式测量的分类变量（如社会调查数据中的种族）作为 X 的代理变量，并使用 FGES-MB 算法[④]搜索 X 的马尔可夫毯，从而确定其父节点作为分类变量的潜在构成因素。应用这种方法还可以进一步考察种族测量模型在不同时期的动态变化，这是预先确立标准的传

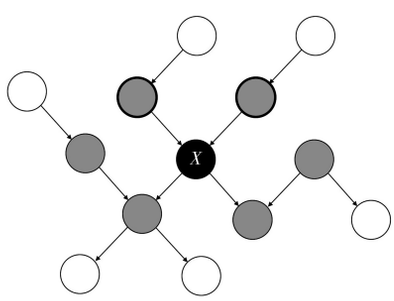

图 5　节点 X 的马尔可夫毯

① Quintana, R., “The ecology of human behavior: A network perspective,” *Methodological Innovations*, 15(1), 2022, pp. 42-61.

② Quintana, R., “What race and gender stand for: Using Markov blankets to identify constitutive and mediating relationships,” *Journal of Computational Social Science*, 5(1), 2022, pp. 751-779.

③ Aliferis, C. F., A. Statnikov, I. Tsamardinos, S. Mani, and X. D. Koutsoukos, “Local Causal and Markov Blanket Induction for Causal Discovery and Feature Selection for Classification Part I: Algorithms and Empirical Evaluation,” *Journal of Machine Learning Research*, 11, 2010, pp. 171-234.

④ FGES-MB 算法是 FGES 的变体，其特点在于无需构建完整的因果结构即可得到特定变量的马尔可夫毯。

统测量方法无法实现的。除性别与种族外，社会科学中的许多重要概念（如健康、贫困、越轨等）也可以采用类似方法，对其潜在构成因素进行测量和分析。

（五）应用因果发现算法面临的问题及解决路径

因果发现算法在应用于实际研究时会面临一些问题，这些问题大多具有普遍性，并不局限于社会科学研究。以下是对这些问题的简要归纳以及对可能解决路径的讨论。

第一，对于算法发现的因果关系，多大程度上能确定它们是可信的？各类因果发现算法的有效性均建立在一定假设基础之上，而不同算法对假设条件的违背表现出了不同的适应性。对特定算法来说，期望所有假设条件都能被满足可能并不现实，尤其是像因果充分性假设（即不存在未观测的混淆变量）这样的强假设，在实际数据中往往难以满足。在假设条件几乎不可避免地会被违背的情况下，有文献建议可以探索尽可能多的因果发现算法，而非依赖单一算法的结果。[①] 如果特定的因果边在应用不同类型的因果发现算法时均被找到，则其可信度更高。此外，自助法重抽样（Bootstrapping）也常被用来考察由算法发现的因果关系的可信性。[②] 研究者可在重抽样样本上反复运行算法，通过估计因果边出现的概率来验证因果关系的稳定性，对出现概率较低的边应持谨慎态度，如昆塔纳在其经验研究中只保留出现概率不低于50%的因果边。[③]

第二，对于发现的整体因果结构，又当如何评估算法性能，并选择更合适的因果模型？在因果发现研究中，使用不同类型的算法对同一观测数据进行分析时，可能会得到不同的因果结构。并且，大多数因果发现算法都涉及需要调整的超参数，比如PC算法中条件独立性检验的显著性水平α，不同的超参数选择也会导致不同的因果结构输出。因此，如何评估算法性能并选择更合适的因果模型成为一个关键问题。检验算法性能，理想方式是将算法发现的因果结构与真实因果结构进行比较，但在实际研究中，真实因果结构通常是未知的——如果已知，因果发现研究本身也就失去了意义。借用经典机器学习算法的分类方式，因果发现当属于非监督学习算法。一些研究者选择采用类似检验模型样本内拟合的方式，具体来说，是用结构方程模型拟合算法发现的因果图，然后借助结构方程模型中常用的模型拟合优度指标评估因果发现算法的性能。[④] 也有研究者开发了样本外因果调优算法，希望借助机器学习中常用的检验样本外预测能力的方法，如K折交叉验证等来寻找更优的模型。[⑤]

第三，大多数因果发现算法得出的是对因果关系的定性判断，能否在此基础上探索因果效应的强度？将结构方程模型与因果发现算法结合使用是可行路径之一。结构方程模型过去一直被期望用于因果分析，尽管它与结构因果模型一样，都使用图模型刻画因果结构[⑥]，但它作为一种验证性方法，变量间的路径图需要依靠理论知识事先建构。将结构方程模型与因果发现算法结合使用，不仅可用于验证因果结构并对各种模型的性能进行比较，还可用于估计因果结构中变量间的因果效应强度。乔汉等在一项关于人们出行方式选择决策的研究中，演示了如何将结构方程模型与因果发现算法结合用于挖掘因果结构，并估计了其中

---

① Malinsky, D., and D. Danks, "Causal discovery algorithms: A practical guide," *Philosophy Compass*, 13(1), 2018, e12470.

② Glymour, C., K. Zhang, and P. Spirtes, "Review of Causal Discovery Methods Based on Graphical Models," *Frontiers in Genetics*, 10, 2019; Schmidt, A. C., et al., "Searching for explanations: Testing social scientific methods in synthetic ground-truthed worlds," *Computational and Mathematical Organization Theory*, 29(1), 2023, pp. 156-187.

③ Quintana, R., "The Structure of Academic Achievement: Searching for Proximal Mechanisms Using Causal Discovery Algorithms," *Sociological Methods & Research*, 52(1), 2023, pp. 85-134; Quintana, R., "From single attitudes to belief systems: Examining the centrality of STEM attitudes using belief network analysis," *International Journal of Educational Research*, 119, 2023, 102179.

④ Ding, C. S., "Bayesian Network for Discovering the Potential Causal Structure in Observational Data," in *Dependent Data in Social Sciences Research: Forms, Issues, and Methods of Analysis*, M. Stemmler, W. Wiedermann, and F. L. Huang (eds.), Springer International Publishing, 2024, pp. 259-286; Chauhan, R. S., C. Riis, S. Adhikari, S. Derrible, E. Zheleva, C. F. Choudhury, and F. C. Pereira, "Determining causality in travel mode choice," *Travel Behaviour and Society*, 36, 2024, 100789.

⑤ Biza, K., I. Tsamardinos, and S. Triantafillou, "Tuning Causal Discovery Algorithms," *International Conference on Probabilistic Graphical Models*, 2020, pp. 17-28.

⑥ Pearl, J., "Graphs, Causality, and Structural Equation Models," *Sociological Methods & Research*, 27(2), 1998, pp. 226-284.

各直接因果效应的大小。他们在文中总结了相应的研究步骤，为类似研究提供了建议。[①]

四、方法关联：因果发现与其他计算方法

（一）因果发现与机器学习

因果发现与因果识别作为因果推断的两种研究路径，有着十分显著的差异。除本文开篇曾提及的不同外，二者用于实现目标的核心技术手段也有很大区别：因果识别的重点在于对因果效应的估计和推断，其核心技术手段仍是传统统计分析的显著性检验方法——先设定因果效应不存在的零假设，然后试图证明零假设成立的概率太小导致其无法被接受，据此说明因果效应很可能存在；与因果识别不同的是，因果发现的重点在于找到最合适的模型，其核心技术手段是迭代与搜索等计算方法。不过，二者尽管存在上述不同，它们最大的相同点是都属于生成建模，即布莱曼“两种建模文化”中的数据建模范畴[②]，都假设观测数据由刻画了某些底层机制的模型生成，生成数据的底层模型是不可或缺的。在因果识别中，生成数据的底层模型是由研究者依据理论知识或假设事先设定的；而在因果发现中，生成数据的底层模型并不为研究者事先知晓，需要利用数据和算法，从众多可能的底层模型中寻找到最有可能的那一个。从这个意义上来说，可将前者视为理论驱动的方法，将后者视为数据驱动的方法。

机器学习一直是计算社会学中数据挖掘路径所运用的主要工具。从数据驱动的角度看，因果发现的确与机器学习有类似之处。但实际上，二者有着实质的不同。一般认为，机器学习算法属于预测建模[③]，布莱曼在其“两种建模文化”中将之称为算法建模。这种建模方式根本不关心数据是如何生成的，也不需要对生成数据的底层模型做任何假设，其主要关注预测模型的泛化能力，即模型在训练样本外的预测表现。对机器学习算法而言，只要模型在新数据上有很好的预测效果，泛化能力强，生成数据的底层模型往往并不重要。从这个意义上说，机器学习算法只是完成了数据拟合，而因果发现算法则是通过揭示数据生成的机制，实现对数据的理解。[④]

尽管因果发现算法与机器学习算法存在实质区别，但学者们并没有放弃将二者进行结合的努力。此前，经济学者在将机器学习应用于因果识别的方法创新方面做出了重要贡献（如双重机器学习）。[⑤] 对于因果发现与机器学习的结合，目前的主要贡献更多由计算机科学家推动。[⑥] 已有社会科学研究开始关注因果发现结合机器学习预测的效果，如有关于出行方式选择决策的研究中对比了因果发现与可解释机器学习分析的结果，发现被因果发现算法识别为直接原因的变量，在预测模型中同样具有最高的 SHAP 值。[⑦] 昆塔纳在研究儿童学业成就的影响因素时，仅将算法发现的近端机制变量纳入机器学习模型预测儿童学业成就。经 K 折交叉验证检验，该简化模型与包含全部变量模型的预测效果不相上下。[⑧]

（二）因果发现与基于行动者建模

与机器学习不同，基于行动者建模（Agent-based Modeling，ABM）是一种理论驱动的研究方法。该方

---

① Chauhan, R. S., C. Riis, S. Adhikari, S. Derrible, E. Zheleva, C. F. Choudhury, and F. C. Pereira, “Determining causality in travel mode choice,” *Travel Behaviour and Society*, 36, 2024, 100789.

② Breiman, L., “Statistical Modeling: The Two Cultures,” *Statistical Science*, 16(3), 2001, pp. 199-231.

③ Molina, M., and F. Garip, “Machine Learning for Sociology,” *Annual Review of Sociology*, 45, 2019, pp. 27-45; 陈云松、吴晓刚、胡安宁、贺光烨、句国栋：《社会预测：基于机器学习的研究新范式》，《社会学研究》2020 年第 3 期。

④ Pearl, J., “Radical empiricism and machine learning research,” *Journal of Causal Inference*, 9(1), 2021, pp. 78-82.

⑤ Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins, “Double/debiased machine learning for treatment and structural parameters,” *The Econometrics Journal*, 21(1), 2018, pp. C1-C68; 郭峰、陶旭辉：《机器学习与社会科学中的因果关系：一个文献综述》，《经济学（季刊）》2023 年第 1 期。

⑥ Dai, H., R. Ding, Y. Jiang, S. Han, and D. Zhang, “ML4C: Seeing Causality Through Latent Vicinity,” 2023, No. arXiv: 2110. 00637.

⑦ Chauhan, R. S., U. Sutradhar, A. Rozhkov, and S. Derrible, “Causation versus Prediction: Comparing Causal Discovery and Inference with Artificial Neural Networks in Travel Mode Choice Modeling,” 2023, No. arXiv: 2307. 15262.

⑧ Quintana, R., “The Structure of Academic Achievement: Searching for Proximal Mechanisms Using Causal Discovery Algorithms,” *Sociological Methods & Research*, 52(1), 2023, pp. 85-134.

法通过计算机模拟个体行动者的行动及相互作用，从而揭示复杂社会现象的涌现机制。ABM在应对复杂系统中的非线性特征、动态演化过程以及多元主体间的复杂互动等方面具有独特优势，受到了计算社会学研究者的广泛重视[①]，被视为与数据挖掘相区别的另一条研究路径。值得注意的是，ABM本身也可作为一种有效的因果分析工具：研究者利用计算机模拟程序创造不同的“平行世界”，在其之上实施“理论调参”——调整个体行动或互动的规则或环境参数，观察并分析系统后果的变化规律。

目前，已有工程学者开始致力于将因果发现与ABM相结合，以深化对复杂现象的理解，同时也用于帮助改进设计，或辅助政策制定和评估。詹森与合作者提出了两种方法路径：其一是运用因果发现方法对ABM中的涌现现象进行分析[②]，其二是基于因果发现进行ABM设计。[③] 第一种路径将ABM模拟中生成的数据作为因果发现算法的输入，以期深入揭示涌现现象的内在机制。詹森等人不仅提供了具体的分析流程与算法建议，还以社会博弈论领域的经典问题——埃尔法罗酒吧问题（El Farol bar problem）为例，展示了如何对涌现属性开展因果结构分析。第二种路径则强调根据因果发现算法得出的因果结构来指导ABM设计，不过詹森等人特别指出，由于不同算法及超参数设置可能导致差异化的因果图输出，因此在此过程中要仔细比较不同因果发现算法的结果，并与专家判断相结合。另外，还有学者提出了将上述两种方法进行结合的迭代分析框架：先用因果发现算法分析ABM模拟的结果，再根据发现的因果结构推动ABM设计的优化。这种迭代分析框架不仅增强了模型的解释能力，还能够为政策的制定和评估提供更可靠的工具支持。[④] 这些方法论层面的探索都值得社会学研究者借鉴，它们提供了新的思路，也为复杂社会系统的研究开辟了新的方向。

（三）因果发现与大语言模型

大语言模型的兴起无疑是具有颠覆性的。但大语言模型还是一种基于概率的预测模型，有知名数学家将其喻为“大型猜测机器”，在这一点上其与传统机器学习模型并无本质不同，其强大的预测能力也并非源于对因果的理解。而因果模型将可能导向一种透明性、可解释性和跨环境可迁移性均更强的推理机器。[⑤] 从技术的底层原理来看，因果发现算法与大语言模型有着本质不同。然而，这并不妨碍应用大语言模型能够增强因果发现。过去两年中已有不少研究开始积极探索如何将大语言模型应用于因果发现。

有研究者从推理模式的角度比较了纯粹数据驱动的因果发现和大语言模型增强的因果发现。[⑥] 尽管比较是否体现了二者的本质差异仍有待商榷，但大语言模型的应用的确可能在突破数据局限、确定因果方向、提供专家知识等方面增强因果发现算法。[⑦] 特别是在专家知识的获取与整合方面，各类因果发现算法都会因专家知识的输入而获得性能提升，而大语言模型恰好可以应用预训练时从海量文献数据中学习到的领域知识，扮演已掌握特定领域因果知识的专家角色。研究者可以通过先向大语言模型输入明确的变量定义，再以结构化提问的方式获取大语言模型对变量间关系的判断，进而将这些知识输入因果发现算法。当

① 乔天宇、邱泽奇：《复杂性研究与拓展社会学边界的机会》，《社会学研究》2020年第2期；吕鹏：《智能体仿真模拟：推进行动与结构互构研究》，《社会学研究》2024年第4期。

② Janssen, S., A. Sharpanskykh, R. Curran, and K. Langendoen, “Using causal discovery to analyze emergence in agent-based models,” *Simulation Modelling Practice and Theory*, 96, 2019, 101940.

③ Janssen, S., A. Sharpanskykh, and S. S. Mohammadi Ziabari, “Using Causal Discovery to Design Agent-Based Models,” in *Multi-Agent-Based Simulation XXII*, K. H. Van Dam and N. Verstaevel(eds.), Springer International Publishing, 2022, pp. 15-28.

④ Chang, S., T. Kato, Y. Koyanagi, K. Uemura, and K. Maruhashi, “An Iterative Analysis Method Using Causal Discovery Algorithms to Enhance ABM as a Policy Tool,” *2023 Winter Simulation Conference (WSC)*, 2023, pp. 138-149.

⑤ 朱迪亚·珀尔、麦肯齐·达纳：《为什么：关于因果关系的新科学》，北京：中信出版集团，2019年。

⑥ Kıcıman, E., R. Ness, A. Sharma, and C. Tan, “Causal Reasoning and Large Language Models: Opening a New Frontier for Causality,” 2024, No. arXiv: 2305. 00050.

⑦ Cohrs, K.-H., G. Varando, E. Diaz, V. Sitokonstantinou, and G. Camps-Valls, “Large Language Models for Constrained-Based Causal Discovery,” 2024, No. arXiv: 2406. 07378.

然，需要指出的是，大语言模型也可能充当“不完美专家”的角色[①]，它虽然能够增强数据驱动的因果发现，但也可能因“幻觉”而提供错误或不准确的信息。因此，将大语言模型应用于因果发现时，如何有效开展提示工程，采用何种提问策略，以及如何校准模型输出的不确定性等，都是研究者需要重点解决的问题。此外，目前的大语言模型多是在通用语料数据库上训练得到的，若想让其更好地辅助专业领域的因果发现，有必要构建高质量的专业领域数据集，并在此基础上训练垂直大模型。尽管目前应用大语言模型增强因果发现尚处于尝试阶段，但其展现出的潜力已引起广泛关注，有待研究者持续跟进和深入探索。

五、总结

面对日益复杂的社会现象和不断提升的数据维度，社会学研究既迎来了新的机遇，也面临着前所未有的挑战。在过去十余年中，计算社会学蓬勃发展，成为了社会学把握机遇、应对挑战的路径之一。但大语言模型兴起后，其无比优异的表现让人们看到人工智能算法在挖掘大数据方面的无限潜能，数据驱动的计算社会学路径可能被替代的想象也随之生发。在这一背景下，计算社会学何为，尤其是数据驱动的计算社会学路径是否还有继续下去的必要，成为了值得探讨的问题。

本文认为，计算社会学的价值不仅在于应用机器学习算法挖掘社会模式和规律，更重要的是，它还继承了社会学寻求因果解释的努力，从事计算社会学的研究者一直积极拥抱因果推断的研究进展。受到自然科学与社会科学各学科共同关注的因果推断，在潜在结果模型和结构因果模型两种理论传统下，形成了因果识别和因果发现两种研究路径。社会学研究者过去更多聚焦于因果识别方法的应用和改良，而对因果发现路径关注不足，既缺乏对其理论基础的系统性考察，也未能充分关注其方法应用的最新进展。

发现因果的算法可以突破传统机器学习预测无法揭示因果关系的局限，为我们更深入理解和解释复杂社会现象、探索因果关系提供新的研究工具。可以预见，因果发现算法的应用将赋予计算社会学以新的学科能力。即使是在预测算法高度发达的大语言模型时代，对因果推断全面、持续的关注会让计算社会学具备无法被替代的价值，这对推动“文科智能”（AI for Humanity and Social Science）的发展以及数智时代人文社会科学知识生产范式变革都将具有深远意义。

不同于理论驱动的因果识别路径，因果发现主要由数据驱动，并不依赖于已有理论认知或预先设定的假设，而是试图直接从观测数据中发现因果关系。不同于因果识别往往局限于少数因果路径，因果发现的目标是搜索给定变量间所有可能的因果关系，力图构建完整的因果结构。当然，在某些研究情境下，因果发现方法也可用于检验特定因果关系假设。此外，因果识别设计多限于检验可实施操纵的干预措施（如治疗方案、社会计划、经济政策等）的因果效应，而因果发现允许模型中包括像年龄这种刻画固有属性的不可操纵变量[②]，拓宽了因果推断的研究视域。

本文在回顾了因果推断两种理论传统的基础上，从算法原理、应用路径和方法关联三个方面，系统介绍了因果发现的经典算法并探讨了它们在计算社会学中可能的应用方向。在算法原理方面，本文在介绍因果马尔可夫条件、*d*-分离准则和因果忠实性假设三个“桥梁原则”和其他相关假设的基础上，概述了 PC、FCI、GES、LiNGAM 等因果发现经典算法的原理。需要补充的是，目前已有多种软件工具支持这些因果发现算法的实现。卡耐基梅隆大学研究团队开发的 Tetrad 程序是专门用于实现因果发现算法的软件工具。此外，主流编程语言中的多个专用软件包（如 Python 语言中的 gCastle 和 causal-learn 以及 R 语言中的 pcalg 和 bnlearn 等）也提供了便捷的算法实现平台。在应用路径方面，本文总结了因果发现算法在区分影响结果的直接与间接原因、提示混淆变量的存在及潜在影响、挖掘变量间的复杂因果关系结构、识别测量模型等

① Long, S., A. Piché, V. Zantedeschi, T. Schuster, and A. Drouin, “Causal Discovery with Language Models as Imperfect Experts,” 2023, No. arXiv: 2307. 02390.

② Leist, A. K., M. Klee, J. H. Kim, D. H. Rehkopf, S. P. A. Bordas, G. Muniz-Terrera, and S. Wade, “Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences,” *Science Advances*, 8(42), 2022, eabk1942; Malinsky, D., and D. Danks, “Causal discovery algorithms: A practical guide,” *Philosophy Compass*, 13(1), 2018, e12470.

方面的应用实践，以及因果发现算法在应用中面临的问题和解决路径。在方法关联方面，本文讨论了因果发现与机器学习、基于行动者建模、大语言模型等其他计算社会学方法间的关联性以及可能的结合应用方式。

当然，因果发现方法也有其局限性。首先，多数因果发现算法都无法直接估计因果效应强度，其推测的因果边也可能是虚假的因果关联，需寻求随机对照实验或结合其他因果识别方法予以进一步验证。其次，算法生成的因果结构通常无法覆盖所有潜在的因果细节，这也是为什么很多研究者将其作为一种探索性方法，而非得出确定性结论的工具。[①] 最后，方法应用存在一定门槛，因果发现算法数量众多，不同算法往往具有不同的使用条件、对数据类型及分布有不同要求、对假设前提具有不同的适应性等。本文涉及的各种经典算法主要针对横截面数据，从纵向数据中发现因果结构还会面临其他挑战。

因果发现算法的应用还对高质量数据提出了要求。样本规模充足且信息覆盖全面的数据集将会提升因果发现算法的性能表现。应用因果发现的研究者不仅要能充分理解算法，更要深入把握数据，需关注算法对数据类型及分布的要求，同时也要重视数据预处理环节。

恰当应用的因果发现算法一定能够助力社会学对因果关系的探索，它们提供的启发性见解可以作为进一步发展理论假设的依据。与可解释机器学习方法类似，因果发现同样具有“打通从经验观察到理论生产的逆向路径”的能力[②]，是数据驱动知识生产的重要研究工具。总之，过去三十多年中因果发现算法的发展为社会学研究提供了丰富的算法工具箱。这些进展可能推动学科迈入数智时代因果分析的新阶段，值得计算社会学研究者给予更多的关注。

〔本文系国家社会科学基金青年项目“数字生态视角下数据要素交易治理的组织体系研究”（23CSH066）、“‘虚拟—现实’融合背景下青少年价值极化的预防机制研究”（23CSH006）的阶段性成果。赵一璋为本文通讯作者。〕

（责任编辑：朱 颖）

Causal Discovery Algorithms: A Neglected Toolbox for Computational Sociology

QIAO Tianyu, ZHAO Yizhang

Abstract: Sociologists have long focused on methods for causal identification and their applications, but have paid relatively little attention to advances in causal discovery. Building on a review of the two theoretical traditions of causal inference, this paper provides a systematic introduction to classical causal discovery algorithms, examining their underlying principles, practical applications, and potential integration with other computational approaches. In an era marked by the rise of large language models, the adoption of causal discovery algorithms offers new disciplinary capabilities for computational sociology, fostering data-driven knowledge production. These algorithms provide a rich toolkit for gaining deeper insights into complex social phenomena and for exploring causal relationships. Moreover, they hold profound implications for advancing a paradigm shift in knowledge production within the humanities and social sciences in the age of digital intelligence.

Key words: causal inference, causal discovery, algorithm tools, data-driven, computational sociology

① Malinsky, D., and D. Danks, “Causal discovery algorithms: A practical guide,” *Philosophy Compass*, 13(1), 2018, e12470; Quintana, R., “The Structure of Academic Achievement: Searching for Proximal Mechanisms Using Causal Discovery Algorithms,” *Sociological Methods & Research*, 52(1), 2023, pp. 85-134.

② 陈茜、陈云松：《计算扎根：定量研究的理论生产方法》，《社会学研究》2023年第4期。