arXiv:2303.02833v1 [cs.LG] 6 Mar 2023

eCDANs：针对自相关和非平稳数据的高效时间因果发现（学生摘要）

Muhammad Hasan Ferdous、Uzma Hasan、Md Osman Gani

C因果人工智能实验室，信息系统系，马里兰大学巴尔的摩县分校 h.ferdous@umbc.edu、uzmahasan@umbc.edu、mogani@umbc.edu

摘要

传统的时间因果发现（CD）方法存在高维性问题，无法识别滞后的因果关系，并且往往忽略关系中的动态变化。本研究提出一种面向自相关和非平稳时间序列数据的新型基于约束的因果发现方法（eCDANs），能够检测滞后和同时因果关系及其随时间的变化。eCDANs 在进行条件独立性（CI）检验时优化条件集，以应对高维性；同时引入替代变量来表示时间依赖性，从而识别因果关系的变化。合成数据和现实世界数据上的实验表明，eCDANs 能够识别时间影响，且优于基线方法。

引言

目前已有许多成熟的方法用于估计时间序列数据背后的因果机制。然而，当时间序列数据具有非平稳性和自相关性时，大多数方法都会失效。为发现自相关数据中的因果关系，大多数基于约束的方法会对变量进行常规条件独立性检验，这些变量可能包括整个过去以及所有同时变量，其中某些条件变量并不相关。这会导致高维性、检测能力降低以及结果不正确（Runge 2020）。近期的研究采用连续优化来处理高维性（Pamfil 等人（2020）；Sun 等人（2021））。然而，这类方法可能存在多个极小值，无法处理数据重新缩放，而且返回的边可能并不代表因果关系（Kaiser 和 Sipos 2022）。此外，变量的季节性和周期性会产生时间影响，从而改变其分布。这种时间影响体现在变化模块 中，并可由替代变量表示（Zhang 等人 2017）。大多数算法忽略了时间序列数据的这一重要组成部分。

为应对这些挑战，我们提出了一种用于从自相关且非平稳数据中高效进行因果发现（eCDANs）的算法。我们提出的方法通过执行PC-stable条件独立性检验来处理高维问题，并利用检验结果执行瞬时条件独立性（MCI）检验。这使我们能够确定同期变量的邻接集（骨架图）。使用MCI检验可确保条件集中排除无关变量。随后，我们利用骨架图和替代变量识别变化模块。该方法利用时间流向、变化模块和定向规则来识别因果方向。

版权所有 © 2023，人工智能促进协会（www.aaai.org）。保留所有权利。

方法论：eCDANs

我们提出的方法 eCDANs 分五个步骤发现因果结构。下面将介绍这些步骤。

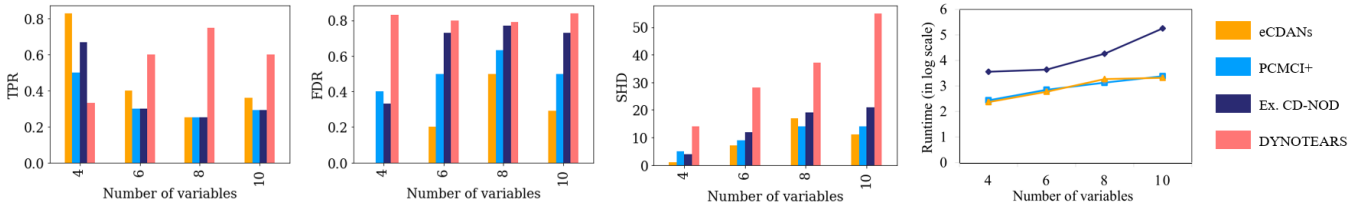
步骤 1 – 检测初始邻接集：令 X_t 表示同时变量，X_t^j 表示时间 t 的 j^{th} 观测值，$X_{t-\tau}^i$ 表示滞后 τ 的 i^{th} 观测值。eCDANs 首先针对所有 i（$i = 1$、2、.....、m），对 X_t^j 和 $X_{t-\tau}^i$ 进行迭代 PC1检验，并为每个 X_t^j 推导出滞后邻接集 L_a（X_t^j）的超集。随后执行 MCI检验，为每个 X_t^j 得到邻接集 Adj（X_t^j）。邻接集中的变量按效应量（检验统计量）降序排列存储。

步骤 2 – 构建无向图：在此阶段，eCDANs 使用滞后邻接关系 L_a（X_t）、同时邻接关系 C_a（X_t）以及替代变量 C，构建部分完备无向图 G。

步骤 3 – 变化模块检测：为检测变化模块，e CDANs 首先针对所有 $j \in$（1，m）对 X_t^j 和 C 进行无条件独立性检验；然后根据效应量从高到低，持续将其他变量从（L_a（X_t^j）$\cup C_a$（X_t^j））中加入条件集。通过减少冗余的条件检验，这种方法优化了条件集和搜索过程。如果 X_t^j 与 C 相互独立，则移除它们之间的边。

步骤 4 – 同时邻接关系识别：在检测能够发现某些虚假边的 C_a（X_t）时，我们在步骤*–1* 中忽略了 C。为消除这些边，eCDANs 在 X_t^i 和 X_t^j 之间执行条件独立性检验；如果在（L_a（X_t^i）$\cup C_a$（X_t^i）$\cup L_a$（X_t^j）$\cup C_a$（X_t^j）$\cup C$）条件下二者相互独立，则移除 X_t^i 与 X_t^j 之间的边。该步骤生成一个包含同时边、滞后边以及同时变量与 C 之间边的因果骨架。

步骤 5——恢复因果方向：基于因果关系遵循时间流向的假设，



(a) TPR (b) FDR (c) SHD (d) Runtime (in sec.)

图 1：不同算法的性能指标

对于所有$\tau = 1$，2、……、 $\tau_{\max}$，算法将 $(X^i_{t-\tau}, X^j_t)$ 定向为 $(X^i_{t-\tau} \to X^j_t)$。假设替代变量导致分布偏移，eCDANs 将 $\left(C, X^j_t\right)$ 定向为 $C \to X^j_t$。对于形式为 $\left(C - X^i_t - X^j_t\right)$的三元组，eCDANs 会回溯 C 与 X^j_t 之间条件独立性检验的条件集。如果该条件集不包含 X^i_t，则将该三元组定向为 $C \to X^i_t \leftarrow X^j_t$；否则，定向为 $C \to X^i_t \to X^j_t$。当 X^i_t 和 X^j_t 都与 C相邻时，eCDANs 使用扩展 HSIC 对 X^i_t 与 X^j_t 之间的边进行定向。eCDANs 还利用因果模块的独立变化来确定因果方向（Huang 等人（2020））。

评估

我们利用滞后父节点对条件集进行了部分优化，并在以下两类数据上，将 eCDANs 的性能与三种基线方法进行了比较：(1) 合成数据集（4、6、8 和 10 个变量）；(2) 一个临床数据集（Gani 等人（2023））数据库，其中包含 12 个时间序列变量1。下面讨论实验结果。

合成数据集上的结果： 图 1 展示了 eCDANs、PCMCI+（Runge 2020）、扩展 CD-NOD（Huang 等人（2020））和 DYNOTEARS（Pamfil 等人（2020））在真正率（TPR）、错误发现率（FDR）和结构汉明距离（SHD）方面的性能。就所有指标而言，eCDANs 的表现始终优于 PCMCI+ 和扩展 CD-NOD。DYNOTEARS 的 TPR 最高，但它无法识别真正的因果关系，并产生非常高的 FDR 和 SHD。因此，我们未在后续分析中考虑 DYNOTEARS。

真实世界数据集上的结果：由于缺少真实因果图，我们无法比较在真实世界数据集上的性能。相反，我们将结果与（Gani 等人（2023）；Bikak 等人（2020））提出的非时间因果图进行了比较，发现 eCDANs 得出的图最接近这些非时间研究的结果。

运行时间：我们在（图 1）中展示了 eCDANs、PCMCI+ 和扩展 CD-NOD 的运行时间。实验结果表明，与其他方法相比，eCDANs 效率更高，且始终具有最短的运行时间。在 8 个变量的设置下，其运行时间略长于 PCMCI+。

未来工作

这项工作仍在进行中，我们正在努力寻找同期邻接关系以及滞后邻接关系，以进一步优化条件独立性检验。这将进一步提升所提出算法的性能并缩短运行时间。我们还计划扩展 eCDANs，以便在时间序列数据集包含潜在混杂因素和变化模块时，识别部分祖先图。

致谢

本研究部分由美国国家科学基金会（项目编号 2118285）和 UMBC START 的拨款资助。

参考文献

Bikak, M.; Kethireddy, S.; Gani, M. O.; and Adibuzza- man, M. 2020. 利用专家增强知识构建结构因果模型，以估计氧疗的效果。*CHEST*, 158(4): A636.

Gani, M. O.; Kethireddy, S.; Adib, R.; Hasan, U.; Griffin, P.; and Adibuzzaman, M. 2023. 利用专家增强知识构建结构因果模型，以估计氧疗对 ICU 死亡率的影响。《医学人工智能》, 137: 102493. Huang, B.; Zhang, K.; Zhang, J.; Ramsey, J. D.; Sanchez- Romero, R.; Glymour, C.; and Sch¨olkopf, B. 2020. 从异质性/非平稳数据中进行因果发现。《机器学习研究杂志》, 21(89): 1–53. Kaiser, M.; and Sipos, M. 2022. 处理维度量时 NOTEARS 不适用于因果图发现。《神经处理快报》, 1–9. Pamfil, R.; Sriwattanaworachai, N.; Desai, S.; Pilgerstor- fer, P.; Georgatzis, K.; Beaumont, P.; and Aragam, B. 2020. Dynotears：从时间序列数据中进行结构学习。发表于人工智能与统计国际会议, 1595–1605. PMLR. Runge, J. 2020. 在自相关非线性时间序列数据集中发现同期与滞后因果关系。发表于人工智能不确定性会议, 1388–1397. PMLR. Sun, X.; Liu, G.; Poupart, P.; and Schulte, O. 2021. NTS-NOTEARS：结合时间序列数据与先验知识学习非参数时间有向无环图。*arXiv* 预印本 *arXiv:2109.04286*. Zhang, K.; Huang, B.; Zhang, J.; Glymour, C.; and Sch¨olkopf, B. 2017. 从非平稳/异质性数据中进行因果发现：骨架估计与方向判定。发表于 *IJCAI*：会议论文集, 卷 2017, 1347. 美国国立卫生研究院公共访问。