

基于特征选择与瞬时条件独立性检验的两阶段时间序列因果推断方法

甘晓鹏, 达朝究*

西北民族大学数学科学学院, 甘肃 兰州

收稿日期: 2026年7月25日; 录用日期: 2026年8月17日; 发布日期: 2026年8月28日

摘要

从观测时间序列数据中识别并量化因果关系, 是理解复杂动力系统的核心科学问题。然而, 由于实际观测数据往往具有高维、非线性且样本量有限等特征, 传统数据驱动方法常受特征冗余干扰, 难以获得可靠的因果推断结果。为此, 本文提出一种基于全局冗余最小化的自动加权特征选择框架(AGRM)与瞬时条件独立性(MCI)检验的两阶段因果推断方法。在第一阶段, 首先利用信息增益(IG)计算变量间的特征得分, 随后引入AGRM从全局视角剔除原始时间序列中的无关和冗余变量, 从而精准提取目标变量的最相关特征集合。在第二阶段, 进行MCI检验以推断变量间的因果关系, 从而刻画出时间序列的因果网络结构。最后, 在一个合成数据集上进行实验, 并引入真阳性率、假阳性率以及运行时间, 作为方法性能评价指标。结果表明, 所提方法能够有效刻画变量间的因果关系, 为时间序列的因果关系推断提供了一种兼顾冗余抑制与因果检验能力的有效思路。

关键词

时间序列, 因果关系分析, 特征选择, 因果网络学习方法

A Two-Stage Method for Time Series Causal Inference Based on Feature Selection and Momentary Conditional Independence Testing

Xiaopeng Gan, Chaojiu Da*

School of Mathematical Sciences, Northwest Minzu University, Lanzhou Gansu

Received: July 25, 2026; accepted: August 17, 2026; published: August 28, 2026

*通讯作者。

文章引用: 甘晓鹏, 达朝究. 基于特征选择与瞬时条件独立性检验的两阶段时间序列因果推断方法[J]. 应用数学进展, 2026, 15(8): 459-472. DOI: 10.12677/aam.2026.158367

Abstract

Identifying and quantifying causal relationships from observed time series data is a core scientific problem for understanding complex dynamical systems. However, real-world observational data are often characterized by high dimensionality, nonlinearity, and limited sample sizes, which make traditional data-driven methods vulnerable to interference from redundant features and hinder the acquisition of reliable causal inference results. To address this issue, this paper proposes a two-stage causal inference method that integrates an Automatic Global Redundancy Minimization-based feature selection framework (AGRM) with Momentary Conditional Independence (MCI) testing. In the first stage, Information Gain (IG) is first used to calculate feature scores among variables, and AGRM is then introduced to remove irrelevant and redundant variables from the original time series from a global perspective, thereby accurately extracting the most relevant feature set for the target variable. In the second stage, MCI testing is performed to infer causal relationships among variables, thus characterizing the causal network structure of the time series. Finally, experiments are conducted on a synthetic dataset, where the true positive rate, false positive rate and computational time are adopted as performance evaluation metrics. The results show that the proposed method can effectively characterize causal relationships among variables, providing an effective approach for time series causal inference that balances redundancy suppression and causal testing capability.

Keywords

Time Series, Causality Analysis, Feature Selection, Causal Network Learning Method

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

复杂动力系统广泛存在于气候变化[1]、水文循环[2]、神经活动[3]和生理调控[4]等诸多领域,深入理解系统内各变量之间的作用机制,对于揭示其运行规律、解释关键现象并提升预测与决策能力具有重要意义。传统上,对这类系统机理的认识往往依赖于控制实验、干预实验或高成本数值模拟[5],但这类方法实施成本高、周期长,并且在医学[6]、神经科学[7]和社会经济[8]等领域还常受伦理约束与现实条件限制。近年来,随着观测技术、传感技术、遥感技术以及数值模拟能力的持续发展,多变量时间序列数据在自然科学与社会科学中日趋丰富,为从观测数据中反演系统内部作用关系提供了新的可能[9][10]。正是在这一背景下,因果推断方法逐渐受到广泛关注[11][12]。相较于只能刻画统计相关性的传统分析方法,因果推断旨在从观测数据中区分真实驱动、间接传递以及共同驱动造成的伪相关[11],从而为复杂系统机理解释和因果关系推断提供更可靠的依据[10]。

根据研究[13],现有时间序列因果推断方法主要包括格兰杰因果、非线性状态空间方法、结构因果模型框架以及基于条件独立性检验的因果网络学习算法。考虑到复杂动力系统往往表现出高维、强时序依赖和复杂变量耦合关系[14],基于条件独立性检验的因果网络学习算法由于在伪相关剔除和因果关系推断方面具有优势[15],因而逐渐成为时间序列因果关系推断的重要研究方向[16]。其中,Runge等人提出的PCMCI算法已成为高维、自相关且非线性时间序列的代表性因果推断方法[17]。该方法首先通过条件选择阶段为每个目标变量筛选候选父节点集合,再据此进行MCI[17]检验,从而提高因果检测能力,减少无关变量引起的效应量稀释,并在控制假阳性的同时提升结果可靠性。随后,这一思路又被扩展到发

现滞后与同时因果关系的 PCMCI⁺ [18]以及面向多数据集的 J-PCMCI⁺ [19]。表明 PCMCI 已形成较为重要的方法谱系并持续在复杂动力系统领域得到应用[20]-[22]。

尽管 PCMCI 在高维、非线性且具有自相关特征的时间序列因果推断中表现出较强优势,但其条件选择阶段是使用 PC 算法[23]进行的,PC 算法本质上是一种基于条件独立性检验结果的自适应条件集搜索过程[24]。已有研究表明,样本条件下的 PC 类算法存在顺序依赖,边删除过程中产生的统计检验误差会进一步影响后续邻接集与分离集的构造,从而影响最终因果推断结果[25][26];同时,在强自相关背景下,传统 PC 类算法及其扩展方法的条件独立性检验结果还可能出现召回率下降以及假阳性增加等问题[27]。进一步地,PC 阶段的目标主要在于剔除无关变量,而并没有考虑变量之间的冗余性,因此其输出结果仍可能包含由变量之间的冗余引起的虚假连接。若不考虑候选父节点集合的冗余程度直接将其传入 MCI 阶段,便可能增加条件集的冗余程度从而影响因果推断的可靠性。

鉴于传入 MCI 阶段的候选父节点集合会直接影响最终因果推断结果,如何在条件选择阶段兼顾信息保留与冗余抑制,成为值得进一步研究的问题。已有研究表明,在故障诊断、设备状态监测和能耗预测等工程任务中,特征选择能够从高维变量中提炼低维、紧凑且更具判别性的输入子集,从而提升后续模型性能[28]-[30]。从两阶段时间序列因果推断的角度看,条件选择与特征选择在功能目标上具有一致性,二者都旨在围绕目标变量构造出适用于后续研究的低维变量集合。PCMCI 及其扩展方法表明,前一阶段所构造的候选父节点集合质量,将直接影响后续 MCI 检验的检测能力与结果可靠性[17]。因而,特征选择本质上可视为对后续 MCI 输入条件集的优化。然而,若要将特征选择真正用于提升 MCI 阶段的输入质量,仅仅筛除无关变量仍然是不够的,还需要进一步关注候选父节点集合内部的冗余结构。

现有许多特征选择方法通常先依据某一评价准则对各候选特征与目标特征之间进行评分,再按照得分高低选取前 k 个特征[31]。此类方法实现简便、计算效率高,但其评价依据主要反映单个特征与目标变量之间的关系,而忽略了特征之间的冗余关系[32]。因而,当多个候选特征彼此高度相关、表征的信息来源相近时,它们往往会因具有相似的重要性得分而被同时保留下来[33]。这将导致有限的特征预算可能被若干携带重复信息的变量占据,从而削弱所选最相关特征集合对目标变量的表征能力[33],并进一步影响后续因果推断的稳定性与可靠性。由此可见,在特征选择阶段,仅关注单个特征的重要性仍然是不充分的,还需要从特征集合整体出发,同时兼顾信息相关性与冗余抑制。

为此,Nie 等人提出了基于全局冗余最小化的自动加权特征选择框架(AGRM) [34]。该方法将已有特征选择方法得到的原始特征得分作为输入,在保持精炼后得分与原始得分一致性的同时,引入冗余矩阵从全局视角抑制携带重复信息的特征的重复保留,从而筛选出更具代表性且冗余程度较低的最相关特征集合。进一步地,AGRM 通过自加权参数自动平衡冗余项与一致性项,无需额外调节超参数;同时,其冗余矩阵由余弦相似度的平方构造,可保证目标函数的凸性,从而获得全局最优解,克服了贪心搜索策略易陷入局部最优的不足。总体而言,相较于仅依据局部排序结果直接截取特征的方法,AGRM 更强调从整体结构上筛选低冗余、高代表性和强互补性的最相关特征集合,因此能够获得更好的判别性能。

鉴于当前的研究现状,本文提出一种面向高维时间序列的两阶段因果推断方法 AGRMIG-MCI。该方法在总体框架上由特征选择阶段与 MCI 检验阶段构成:首先,利用信息增益(IG) [35]获得各候选特征相对于目标变量的原始特征得分;在此基础上,引入 AGRM 框架,对原始特征得分进行进一步精炼,从而为每个目标变量筛选出低冗余、强代表性的最相关特征集合;随后,将所获得的最相关特征集合输入 MCI 检验阶段,以推断变量之间的因果关系并构建时间序列因果网络。

2. 预备知识

目前,特征选择方法大致可以分为以下几类:基于相似性的方法、基于信息论的方法、基于稀疏学

习的方法和基于统计的方法[36]。由于基于信息论的方法不仅可以刻画线性相关关系,还能在一定程度上反映非线性依赖结构,因此许多特征选择方法基于信息论建立[37]。其中,信息增益(IG)是相对基础且常用的方法。鉴于本文方法首先需要获得候选滞后变量相对于目标变量的原始特征得分,因此IG被选为本文的特征选择方法,并在本节中简要描述。

此外,本文所采用的AGRM框架属于带约束的优化问题,因此我们选择了在大规模约束优化和特征选择等问题中应用广泛的增广拉格朗日乘数(ALM)法求解AGRM框架。

2.1. 信息增益

IG是一种广泛用于度量特征之间相关性的方法,其结构简单、易于理解,是目前常用的特征选择方法之一[38]。该方法通过度量变量间的关联程度,对候选变量的重要性进行评分。计算公式如下:

$$J_{IG}(X;Y)=H(X)-H(X|Y) \quad (1)$$

其中, $H(\cdot)$ 表示信息熵,表示变量含有信息量的大小。 $H(X)$ 和 $H(X|Y)$ 分别表示变量 X 的信息熵和变量 X 基于变量 Y 的条件熵[39]。

$$H(X)=-\sum_{x_i \in X} p(x_i) \log(p(x_i)) \quad (2)$$

$$H(X|Y)=-\sum_{y_i \in Y} p(y_i) \sum_{x_i \in X} p(x_i|y_i) \log(p(x_i|y_i)) \quad (3)$$

2.2. 增广拉格朗日乘数法

ALM法是求解约束优化问题的一类经典方法,与普通拉格朗日乘数法相比,ALM在目标函数中进一步加入二次惩罚项,使优化过程不仅能够利用拉格朗日乘子刻画约束条件,同时还能够通过惩罚项增强对约束违背程度的控制。由于其在数值稳定性和收敛性方面具有较好表现,ALM已被广泛应用于大规模约束优化和特征选择等任务中[40][41]。ALM用于求解如下优化问题:

$$\min f(X) \text{ s.t. } Tr(\varphi(X))=0 \quad (4)$$

其中, $f(X)$ 是目标函数, X 是待优化的变量, $Tr(\cdot)$ 是迹运算, $\varphi(X)$ 表示约束函数。公式(4)的增广拉格朗日函数为:

$$\mathcal{L}(X, \Lambda, \mu)=f(X)+Tr(\Lambda^T \varphi(X))+\frac{\mu}{2}\|\varphi(X)\|_F^2 \quad (5)$$

其中, Λ 是拉格朗日乘子, μ 是二次惩罚参数, $\|\cdot\|_F^2$ 是Frobenius范数的平方。公式(5)可进一步改写为:

$$\mathcal{L}(X, \Lambda, \mu)=f(X)+\frac{\mu}{2}\left\|\varphi(X)+\frac{\Lambda}{\mu}\right\|_F^2 \quad (6)$$

文献[42]-[44]讨论并证明了ALM的收敛性。关于ALM法求解AGRM框架的更详细推导可以在文献[34]中找到。

3. 研究方法

3.1. 因果网络框架

为了理解所提出的模型,图1给出了一个示意图,其中描述了两阶段因果方法的过程。图1(a)是FullCI[17]检验的示意图。对于一个 N 维的多元时间序列,FullCI根据公式(7)测试 $X_{t-\tau}^i$ 和 X_t^j 之间的因果关系

$$\text{FullCI: } X_{t-\tau}^i \perp\!\!\!\perp X_t^j | X_t^- \setminus X_{t-\tau}^i \quad (7)$$

其中, X_t^- 表示包含 $N^* \tau_{\max}$ 个变量的历史公共数据集, $X_t^- \setminus X_{t-\tau}^i$ 表示除去测试变量 $X_{t-\tau}^i$ 后的条件集。从公式(7)可以看出, FullCI 在给定包含 $N^* \tau_{\max} - 1$ 个变量的 $X_t^- \setminus X_{t-\tau}^i$ 条件下检验变量之间是否存在因果关系, 这将导致计算量大且准确性差。此外, 该方法在处理复杂的因果网络时往往表现出较差的性能, 这在实验部分将详细描述。

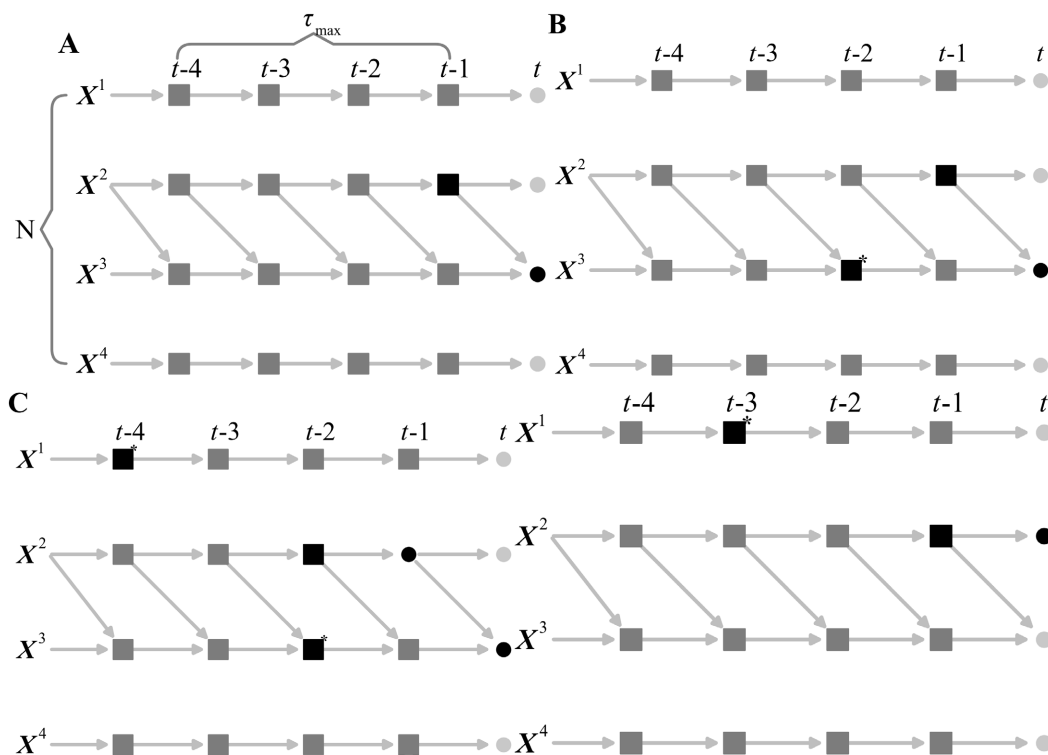


Figure 1. Schematic diagram of FullCI and the two-stage causal inference method

图 1. FullCI 与两阶段因果推断方法的示意图

本文采用的方法由两个阶段组成, 首先进行如图 1(b)所示的特征选择阶段, 该阶段可以通过基于评分的特征选择方法或基于条件独立性检验完成。在完成第一阶段后, 可得到各变量在当前时刻 t 对应的最相关特征集合。如图 1(b)所示, 黑色方块分别表示 X_t^2 和 X_t^3 的最相关特征集合。同时图中标记为*的部分表示被误选入集合的特征, 这可能与特征之间存在较强耦合关系有关。图 1(c)是 MCI 检验阶段。基于特征选择阶段的结果, MCI 使用低维条件集来消除虚假关系并构建因果网络。对于测试 $X_t^2 \rightarrow X_t^3$, 目标变量 X_t^3 的最相关特征集合 $\hat{P}(X_t^3)$ 足以测试条件独立性, 但 MCI 还考虑了测试变量 X_t^2 的最相关特征集合 $\hat{P}(X_t^2)$ 以考虑自相关的影响, 这种方式已被证明在提高性能方面是有效的[17]。

3.2. 全局冗余最小化自动加权特征选择框架

AGRM 框架首先通过一般特征选择方法(例如 IG、拉普拉斯得分、最大方差、稀疏正则化等, 在本文选择 IG)获得原始特征得分 s , 随后引入冗余矩阵 $A \in \mathbb{R}^{d \times d}$ (d 是特征的数量)度量所有特征的冗余性, 并在保持与原始特征得分 s 一致性的前提下, 结合候选特征之间的冗余关系对特征进行重评分, 进而得到新的特征得分 z [34]。理论上, AGRM 既需要抑制候选变量之间的冗余关系, 又需要保证优化后的特征得分 z 与原始特征得分 s 之间具有较高一致性。基于此, AGRM 框架的目标函数如下:

$$\begin{aligned} \min \lambda^2 \mathbf{z}^T \mathbf{A} \mathbf{z} - \lambda \mathbf{z}^T \mathbf{s} \\ \text{s.t. } \mathbf{z}^T \mathbf{1} = 1, z \geq 0 \end{aligned} \quad (8)$$

其中, $\mathbf{1} = (1, 1, \dots, 1)^T \in \mathbb{R}^{d \times 1}$, λ 是自加权参数(自动平衡第一项 $\mathbf{z}^T \mathbf{A} \mathbf{z}$ 和第二项 $\mathbf{z}^T \mathbf{s}$), 冗余矩阵 $\mathbf{A}$ 的计算基于余弦相似度的平方:

$$A_{i,j} = (\mathbf{B}_{i,j})^2 = \left(\frac{\mathbf{f}_i^T \mathbf{f}_j}{\|\mathbf{f}_i\| \|\mathbf{f}_j\|} \right)^2 \quad (9)$$

其中, $\|\mathbf{f}_i\|_2$ 表示 l_2 范数, $\|\mathbf{f}_i\|_2 = \sqrt{\mathbf{f}_i^T \mathbf{f}_i}$, $\mathbf{f}_i \in \mathbb{R}^{n \times 1}$ 和 $\mathbf{f}_j \in \mathbb{R}^{n \times 1}$ 分别表示原始数据集中第 i 和第 j 个中心化特征, 可以通过公式(10)计算:

$$\begin{cases} \mathbf{f}_i = \mathbf{H}_n \mathbf{X}^{(i)} \\ \mathbf{f}_j = \mathbf{H}_n \mathbf{X}^{(j)} \end{cases} \quad (10)$$

其中, $\mathbf{H}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}^T$ 表示中心化矩阵, $\mathbf{I}_n$ 是单位矩阵, $\mathbf{X}^{(i)}$ 和 $\mathbf{X}^{(j)}$ 表示原始数据集 $\mathbf{X}$ 中的第 i 和第 j 个特征。

冗余矩阵 $\mathbf{A}$ 也可以通过 $\mathbf{A} = \mathbf{B} \circ \mathbf{B}$ 计算, 其中 $\circ$ 表示 Hadamard 积[45], 而 $\mathbf{B}$ 可以表示为:

$$\mathbf{B} = \mathbf{D} \mathbf{F}^T \mathbf{F} \mathbf{D} = (\mathbf{F} \mathbf{D})^T \mathbf{F} \mathbf{D} \quad (11)$$

其中, $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_d] \in \mathbb{R}^{n \times d}$, $\mathbf{D} \in \mathbb{R}^{d \times d}$ 是对角矩阵, 元素为 $\frac{1}{\|\mathbf{f}_i\|}$, $i = 1, 2, \dots, d$ 。根据式(11)可知, $\mathbf{B}$

是半正定矩阵, 因此 $\mathbf{A}$ 也是一个半正定矩阵。因此公式(8)是一个凸优化问题[46], 这使得我们可以获得公式(8)的最优解。在本研究中, 目标函数通过 ALM 求解, ALM 在第 2.2 节中介绍。

3.3. 瞬时条件独立性检验

在特征选择阶段之后, 我们获得了每个目标变量 X_t^j ($i = 1, 2, \dots, d$) 的最相关变量, 这些变量组成了 X_t^j 的最相关特征集合 $\hat{P}(X_t^j)$, 其中索引 t 表示当前时间。然后, 使用 MCI 检验来推断变量之间的因果关系, 并构建因果网络。根据马尔可夫理论[17], 通常只需选择目标特征的最相关特征集合作为条件集来进行条件独立性检验。然而, MCI 还将测试变量的父集 $(\hat{P}_{PX}(X_{t-\tau}^i))$ 添加到条件集中, 如公式(12)所示:

$$\text{MCI: } X_t^j | \hat{P}(X_t^j) \setminus X_{t-\tau}^i, \hat{P}_{PX}(X_{t-\tau}^i) \quad (12)$$

其中, PX 表示变量 X^i 的最相关集合中包含变量的最大数量, 但是一般不做限制[17]。MCI 的这种检验方式已被证明可以考虑自相关的影响, 从而降低假阳性率[18]。

3.4. 总体因果网络方法

本文提出的方法 AGRMIG-MCI 方法采用两阶段结构, 首先通过 AGRM 框架进行特征选择, 随后利用 MCI 检验推断变量间的因果关系。

第一阶段: 基于 AGRM 框架的特征选择

步骤 1: 对研究的时间序列数据 $\mathbf{X}$ 进行中心化处理, 然后扩展以获得历史公共数据集 $\mathbf{X}^-$ 如图 1 所示。

步骤 2: 使用特征选择方法 IG, 计算历史公共数据集 $\mathbf{X}^-$ 中每一个变量和目标变量 X_t^j ($i = 1, 2, \dots, d$) 之间的原始特征得分 $\mathbf{s}$ 。

步骤 3: 使用步骤 2 获得的原始特征得分 $\mathbf{s}$ 构建 AGRM 框架, 如公式(8)所示。通过 ALM 对 AGRM 框架进行求解, 得到精炼后的特征得分 $\mathbf{z}$ 。

步骤 4: 通过选择 $\mathbf{z}$ 中得分最高的 k 个特征, 得到目标变量 X_t^j 的最相关特征集合 $\hat{P}(X_t^j), j=1, 2, \dots, d$ 。

第二阶段: MCI 检验阶段

步骤 1: 根据第一阶段的结果, 目标变量 $X_t^j (j=1, 2, \dots, d)$ 的最相关特征集合 $\hat{P}(X_t^j)$ 被得到。当度量变量 $X_{t-\tau}^i (i=1, 2, \dots, d)$, $1 \leq \tau \leq \tau_{\max}$ 和 X_t^j 之间的因果关系时, 如果 $X_{t-\tau}^i$ 存在于 $\hat{P}(X_t^j)$, 那么将 $X_{t-\tau}^i$ 从 $\hat{P}(X_t^j)$ 中移除。

步骤 2: 测试变量 $X_{t-\tau}^i$ 的最相关特征集合 $\hat{P}_{PX}(X_{t-\tau}^i)$ 可以通过将 $\hat{P}(X_t^j)$ 移动 τ 个时间步得到。

步骤 3: 将测试变量和目标变量的最相关特征集合代入公式(12)得到 MCI 值和 p_{value} 。

步骤 4: 通过比较 p_{value} 与显著性阈值 α 来推断变量间的因果关系。

4. 实验及结果

为验证 AGRMIG-MCI 的有效性, 本节从作用机制和统计性能两个方面开展实验。首先, 通过因果结构已知的非线性仿真系统, 分析候选父节点集合中的冗余变量对 MCI 检验的影响; 随后, 基于 Runge 等人提出的高维时间序列因果发现合成数据实验框架[17], 在不同变量规模下比较各方法的真阳性率、假阳性率以及运行时间, 并通过消融实验评估 AGRM、特征截断及初始特征评分方法的作用。

4.1. 候选父节点集合冗余对 MCI 阶段的影响

本小节采用一种非线性的九变量四阶系统 VAR₄(9)进行仿真实验, 其生成方程如下:

$$\begin{aligned} x_{1,t} &= 0.5x_{1,t-1} + u_{1,t} \\ x_{2,t} &= 0.3x_{2,t-1} + 0.5x_{1,t-2} - 0.3x_{1,t-2}^2 + u_{2,t} \\ x_{3,t} &= 0.4x_{3,t-1} - 0.5x_{2,t-3} + u_{3,t} \\ x_{4,t} &= 0.2x_{4,t-1} + 0.4x_{3,t-4} + 0.3x_{3,t-4}^2 + u_{4,t} \\ x_{5,t} &= 0.6x_{5,t-1} + u_{5,t} \\ x_{6,t} &= 0.3x_{6,t-1} - 0.4x_{5,t-3} + 0.2x_{5,t-3}^2 + u_{6,t} \\ x_{7,t} &= 0.5x_{7,t-1} + 0.4x_{6,t-2} + u_{7,t} \\ x_{8,t} &= 0.2x_{8,t-1} + 0.4x_{7,t-1} - 0.2x_{7,t-1}^2 - 0.3x_{4,t-2} + u_{8,t} \\ x_{9,t} &= 0.3x_{9,t-1} - 0.4x_{8,t-4} + 0.3x_{8,t-4}^2 + u_{9,t} \end{aligned} \quad (13)$$

其中, $u_{i,t} (i=1, 2, \dots, 9)$ 是零均值、单位方差的高斯噪声, 一共生成 2000 组样本。在本实验中, $\tau_{\max}$ 设置为 4, 采用 CMI 进行条件独立性检验, 这里根据专家经验[47]设置 CMI $k_{\text{CMI}} = 0.1n$, 最近邻 $k_{\text{perm}} = 5$ 以及替代数量 $B = 500$ 来近似零分布。其中, n 表示样本数量。

使用 PCMCI、FullCI 和 AGRMIG-MCI 对式(13)所定义系统进行因果推断。图 2 给出了不同方法在非线性的 VAR₄(9)系统上的因果推断结果, 图中的有向边表示识别出的因果关系, 颜色条对应 CMI 统计量的大小。从图 2(a)中可看出 PCMCI 将 $x_{2,t-1} \rightarrow x_{8,t}$ 识别为真实的因果链接, 而该连接并不存在于系统的真实因果图中。其主要原因在于, PCMCI 虽然在 PC 阶段成功将真实父节点 $x_{8,t-1}$ 、 $x_{7,t-1}$ 、 $x_{4,t-2}$ 纳入了 $\hat{P}(x_{8,t})$, 但在有限样本条件下, 受条件独立性检验误差及候选集合逐步更新的影响, $x_{2,t-4}$ 、 $x_{8,t-3}$ 等冗余变量也被保留在 $\hat{P}(x_{8,t})$ 中, 导致传入 MCI 阶段的候选父节点集合存在冗余变量, 从而影响了 PCMCI 对真实因果边的准确识别。而 FullCI 算法不仅将 $x_{1,t-1} \rightarrow x_{2,t}$ 、 $x_{1,t-3} \rightarrow x_{2,t}$ 误判为真实边, 同时漏检了 $x_{7,t-1} \rightarrow x_{8,t}$ 。

AGRMIG-MCI 在该仿真系统上推断出了全部的因果链接, 且未出现虚假连接。其原因在于, 该方法

通过在原始特征得分的基础上从全局视角对候选特征之间的冗余关系进行统一建模与重评分, 从而抑制高冗余的特征并保留具有更强代表性的最相关特征。在此基础上, 传入 MCI 阶段的条件集更加紧凑且冗余度更低, 从而有助于提高真实因果边的检出能力并抑制伪因果链接的保留。

仿真实验表明, 候选父节点集合既需要具有较高的真实父节点覆盖率, 也需要尽量减少冗余变量。真实父节点覆盖率不足会造成重要因果信息缺失, 而冗余变量过多则会增大 MCI 检验的条件集规模并降低统计功效。二者都会影响 MCI 阶段的检验结果, 进而影响 PCMCI 的整体因果推断性能。

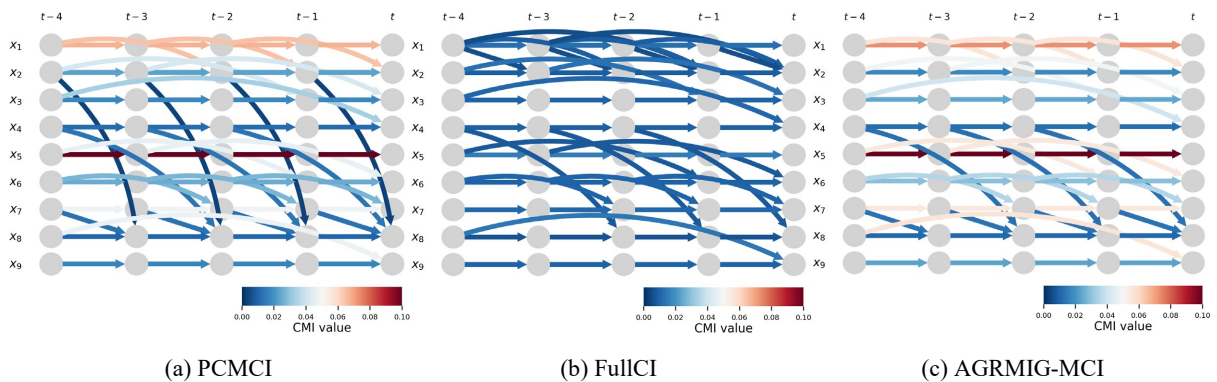


Figure 2. Causal inference results of different methods on the nonlinear VAR4(9) system

图 2. 不同方法在非线性 VAR4(9)系统上的因果推断结果

4.2. 数值实验

4.2.1. 模型设置

为了评估所提方法的有效性, 我们使用一个模拟真实世界数据特性的模型。对于给定的变量数 N 、跨变量链接数 L 和耦合强度 c , 随机生成 10 个因果耦合拓扑。每个因果耦合拓扑随机生成 $L = N$ 条跨变量滞后因果链接。对于 $i, j \in \{1, \dots, N\}$, 随机选择 L 条链接 $i \rightarrow j$, 其中 $i \neq j$, 并按照下式生成时间序列

$$X_t^j = a_j X_{t-1}^j + c \sum_i f_i(X_{t-\tau_i}^i) + \eta_t^j \quad (14)$$

其中, $j \in \{1, \dots, N\}$, a_j 表示自相关系数从集合 $\{0, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95\}$ 中随机抽取, $\eta^j \sim \mathcal{N}(0, 1)$ 为独立同分布高斯噪声, $f_i(x) = x$, τ_i 表示跨变量链接的时间滞后从集合 $\{1, 2\}$ 中均匀随机抽取, 耦合强度 c 的绝对值为 0.287, 其符号以相同概率取正或负。据此每个耦合拓扑生成 100 条时间序列, 每个时间序列的样本量均为 150。

对于每个随机因果拓扑, 独立生成 100 组长度为 150 的多变量时间序列。因此, 在每一种变量规模 N 下, 共获得 10 个随机因果拓扑和 1000 组独立时间序列, 用于统计不同方法的因果推断性能。条件独立性检验采用 ParCorr, 最大时间滞后设为 $\tau_{\max} = 5$ 。数据生成模型中的真实跨变量因果滞后仅取 1 或 2, 而最大时间滞后设置为 5, 使各方法还需要从更大的候选滞后范围中排除不存在的因果连接, 从而考察其识别真实滞后和抑制虚假连接的能力。

4.2.2. 对比方法与参数设置

为全面评估所提方法的因果推断性能, 并分析 AGRM 模块、特征截断策略以及初始特征评分方法对实验结果的影响, 本文选取以下六种方法进行比较: FullCI、PCMCI、IG-MCI、AGRMIG-MCI ($k = 5$)、AGRMIG-MCI (不限制 k) 和 AGRMCorr-MCI ($k = 5$)。其中, FullCI、PCMCI、IG-MCI 和 AGRMIG-MCI 已在前文中进行了介绍, 本节主要对另外两种对比方法进行说明。

AGRMIG-MCI(不限制 k)与本文所提 AGRMIG-MCI 采用相同的 IG 初始评分、AGRM 全局冗余优化和 MCI 检验流程, 区别在于该方法不对第一阶段构造的最相关特征集合设置固定的数量上限, 即不按照预先指定的 k 值对 AGRM 重评分结果进行硬截断。设置该方法的目的在于分析固定 k 截断对候选父节点集合规模以及因果推断结果的影响。

AGRMCorr-MCI ($k = 5$)通过将候选滞后变量与目标变量之间的绝对 Pearson 相关系数作为原始特征得分, 并利用 AGRM 框架从全局角度对原始得分进行优化, 最后选择优化后得分最高的 5 个候选特征构造最相关特征集合并执行 MCI 检验。该方法与 AGRMIG-MCI 的主要区别在于初始特征评分方式不同, 用于考察 AGRM 框架对初始评分准则的依赖程度, 以及所提两阶段框架在不同特征相关性度量下的适应性。

各方法参数设置如下: PCMCI 在 PC 阶段的显著性阈值设置为 $\alpha_{PC} = 0.20$, 其 MCI 阶段的显著性阈值设置为 0.05。IG-MCI、两种 AGRMIG-MCI 以及 AGRMCorr-MCI 的 MCI 阶段显著性阈值均设置为 0.05。FullCI 不包含独立的特征选择和 MCI 阶段, 其条件独立性检验显著性阈值设置为 0.05。为保证消融实验中特征预算一致, IG-MCI、AGRMIG-MCI ($k = 5$)和 AGRMCorr-MCI ($k = 5$)均选择得分最高的 5 个候选滞后变量。

所有方法的运行时间均在同一云端容器实例上进行测试。实验环境配置为 16vCPU Intel Xeon Platinum 8470Q、80 GB 内存、Ubuntu 20.04 操作系统和 Python 3.8, 以避免计算环境差异对运行时间比较造成影响。

4.2.3. 性能评价指标

本文采用真阳性率(True Positive Rate, TPR)、假阳性率(False Positive Rate, FPR)和运行时间(Runtime)评价不同方法的因果推断性能。真阳性率定义为:

$$TPR = \frac{TP}{TP + FN} \quad (15)$$

其中, TP 表示被正确识别的真实因果边数量, FN 表示未被识别的真实因果边数量。TPR 越高, 说明方法恢复真实因果结构的能力越强。假阳性率定义为:

$$FPR = \frac{FP}{FP + TN} \quad (16)$$

其中, FP 表示被错误识别为因果关系的非真实边数量, TN 表示被正确排除的非真实边数量。FPR 越低, 说明方法抑制虚假因果连接的能力越强。

Runtime 表示单次因果网络推断所需的实际运行时间。本文统计 10 个随机因果拓扑、每个拓扑 100 次独立实验所得 TPR、FPR 和 Runtime 的均值与标准差, 并以“均值 $\pm$ 标准差”的形式报告结果。

4.3. 实验结果与分析

图 3 给出了六种方法在不同变量规模下的真阳性率、假阳性率和运行时间。横坐标为时间序列变量数 N , 图 3(a)、图 3(b)和图 3(c)的纵坐标分别为 TPR、FPR 和 Runtime。图中的实心标记表示各指标的均值, 误差棒表示相应的标准差, 用于反映实验结果的波动程度。图中不同颜色和标记分别对应不同方法: 红色圆形表示 AGRMCorr-MCI ($k = 5$), 浅红色方形表示 AGRMIG-MCI ($k = 5$), 浅蓝色上三角形表示 AGRMIG-MCI(不限制 k), 蓝色菱形表示 IG-MCI, 深蓝色下三角形表示 PCMCI, 绿色五边形表示 FullCI。对于因条件集维数过高而无法完成有效估计的情况, 图中将相应的 TPR、FPR 和 Runtime 记为 0, 该数值仅表示方法未获得有效因果推断结果。

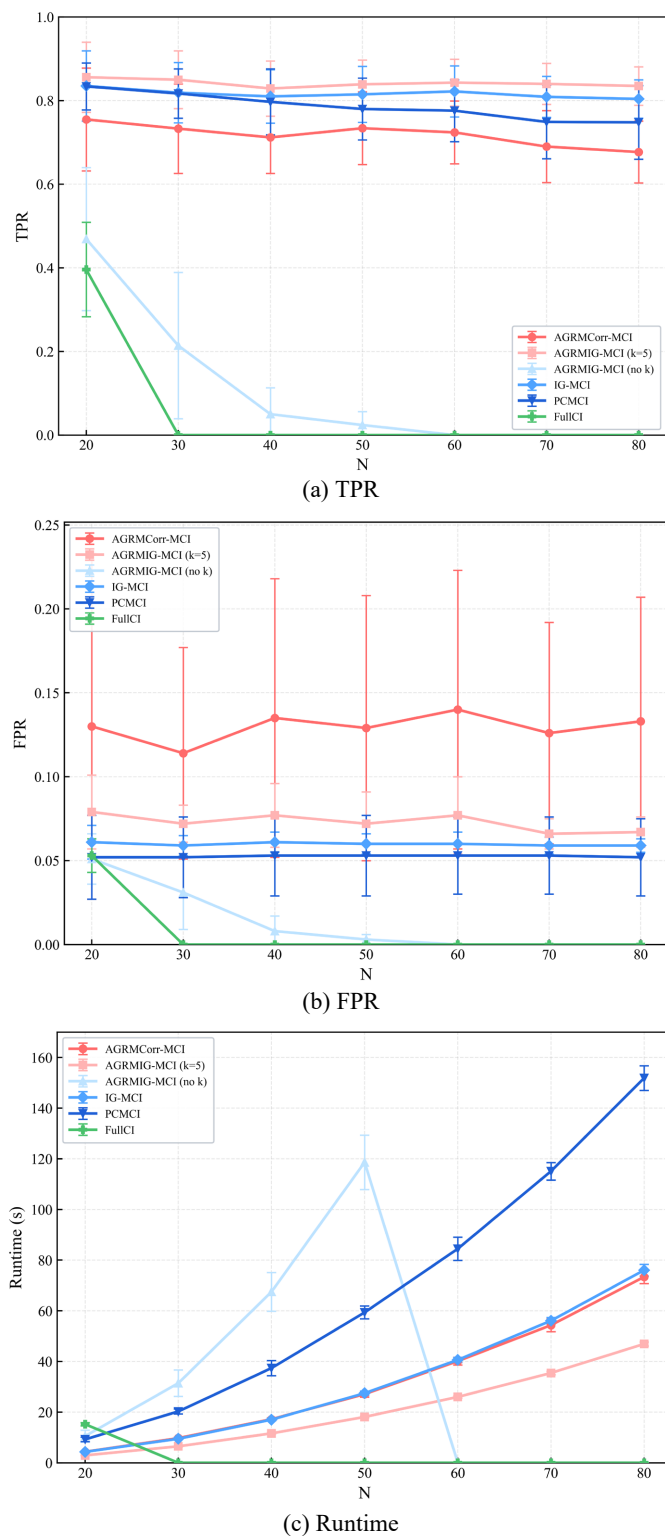


Figure 3. Causal inference performance of different methods under different numbers of variables
图 3. 不同方法在不同变量规模下的因果推断性能

从图 3(a)可以看出, 随着变量规模增加, PCMCI 的 TPR 呈现较为明显的下降趋势。当 $N = 80$ 时,

其 TPR 仅为 0.748。图 3(c)表明 PCMCI 的运行时间随变量数增加而快速增长。这是因为 PCMCI 需要在 PC 阶段执行多轮条件独立性检验, 随着候选滞后变量数量增加, 其搜索空间和计算开销也随之扩大。尽管如此, PCMCI 的 FPR 始终保持在较低水平, 说明其仍具有较好的虚假因果连接控制能力。

相比之下, AGRMIG-MCI ($k=5$)在各变量规模下的 TPR 均高于 PCMCI, 提升幅度约为 2.0~8.5 个百分点, 且其误差棒整体较短, 表明该方法在不同随机因果拓扑下具有较好的稳定性。从图 3(b)可以看出, 该方法的 FPR 始终处于较低水平, 未因 TPR 的提高而明显增加。在运行时间方面, AGRMIG-MCI ($k=5$)的增长速度明显低于 PCMCI。因此, AGRMIG-MCI ($k=5$)能够在提高真实因果边检出率的同时, 将假阳性率和计算开销控制在较低水平。

FullCI 在检验候选因果边时, 需要将最大时间滞后范围内的大量历史变量同时纳入条件集。当 $N=30$ 时, 条件集维数已超出有限样本下 ParCorr 能够有效估计的范围, 导致 FullCI 无法继续完成因果推断, 相应指标被记为 0。这说明 FullCI 对条件集维数较为敏感, 在样本量固定而变量规模持续增加的情况下, 其适用性受到明显限制。

AGRMIG-MCI (不限制 k)首先利用 AGRM 对原始特征得分进行优化, 并剔除优化后得分为 0 的候选滞后变量, 因此其在较低变量规模下的性能优于 FullCI。由图 3 可知, 该方法在 $N=30$ 、 $N=40$ 时仍能完成有效因果推断, 说明 AGRM 的重评分过程能够在一定程度上压缩传入 MCI 阶段的条件集。然而, 由于该方法未对最终保留的特征数量设置上限, 随着变量规模增加, 大量与目标变量仅弱相关, 或其相关性主要由自相关、间接关联等因素引起的候选变量仍可能具有较小的非零得分, 并被传入 MCI 阶段。故当 $N=60$ 时, 条件集维数同样过高, 导致方法无法完成有效估计。

上述结果表明, 仅依靠 AGRM 重评分仍不足以保证条件集始终保持低维, 还需要通过固定 k 对进入 MCI 阶段的特征数量进行约束。因此, AGRMIG-MCI (不限制 k)虽然比 FullCI 具有更好的高维适应性, 但其性能仍明显低于 AGRMIG-MCI ($k=5$)。其中, AGRM 用于抑制候选变量之间的冗余并优化特征排序, k 截断则用于控制条件集规模。二者结合后, 既能够提高候选父节点集合的质量, 又能够避免条件集维数随变量规模持续增长。本文将固定特征预算设置为 $k=5$, 并在 IG-MCI、AGRMIG-MCI 和 AGRMCorr-MCI 中采用相同设置, 以保证消融实验和对比实验的公平性。在实际应用中, k 应结合样本量、时间尺度和具体研究问题进行确定。例如, 当研究对象为以日为时间分辨率、以周为分析尺度的气候系统时, 可将 k 设置为 7, 以覆盖一个完整周尺度内的候选滞后信息。

AGRMCorr-MCI ($k=5$)在各变量规模下均能够完成因果推断, 说明固定特征数量能够有效避免高维条件集导致的估计失效。然而, 其 TPR 整体低于 AGRMIG-MCI ($k=5$), FPR 则明显较高, 且误差棒相对较长。这主要是因为 Pearson 相关系数只能反映候选滞后变量与目标变量之间的线性相关程度, 容易受到自相关、间接关系和共同驱动等因素的影响, 使部分非真实父节点获得较高的初始得分。虽然 AGRM 能够进一步抑制特征之间的冗余, 但其优化结果仍会受到初始特征评分质量的影响。因此, AGRMCorr-MCI 的整体性能逊于 AGRMIG-MCI。

IG-MCI 与 AGRMIG-MCI ($k=5$)采用相同的 IG 初始评分、特征数量和 MCI 检验流程, 二者的区别仅在于 IG-MCI 未使用 AGRM 对原始特征得分进行冗余优化。从图 3 可以看出, IG-MCI 的 TPR 在各变量规模下均低于 AGRMIG-MCI ($k=5$), 并且随着变量数增加下降更为明显。两种方法的 FPR 均保持在较低水平, 但 AGRMIG-MCI 在提高 TPR 的同时并未明显增加 FPR。该消融实验表明, 仅根据 IG 得分直接选择前 5 个候选变量时, 有限的特征预算可能被多个携带相似信息的冗余变量占用; 加入 AGRM 后, 可以从全局角度调整候选特征得分, 优先保留代表性和互补性更强的变量, 从而提高传入 MCI 阶段的候选父节点集合质量。

综合以上结果, 候选父节点集合不仅需要具有较高的真实父节点覆盖率, 还需要尽量减少冗余变量

并控制条件集规模。真实父节点覆盖不足、冗余变量过多或条件集维数过高, 都会影响后续 MCI 检验的统计功效和计算效率, 进而降低最终的因果推断性能。AGRMIG-MCI ($k = 5$) 通过 IG 初始评分、AGRM 冗余优化和固定 k 截断, 在真实信息保留、冗余抑制和条件集规模控制之间取得了较好的平衡, 因此在不同变量规模下表现出更好的综合性能。

5. 结束语

本文提出了一种基于特征选择和瞬时条件独立性检验的两阶段因果推断方法 AGRMIG-MCI。该方法首先利用信息增益对候选滞后变量进行初始评分, 然后通过 AGRM 从全局角度抑制冗余特征、优化最相关特征集合质量, 最后结合 MCI 检验实现因果关系推断。与传统两阶段方法相比, AGRMIG-MCI 进一步压缩了冗余条件集, 从而为后续因果推断提供了更优的输入。

实验结果表明, 所提方法在非线性 VAR₄(9) 系统和不同变量规模的合成时间序列数据集上均表现出较好的因果推断性能。在非线性 VAR₄(9) 系统中, AGRMIG-MCI 能够准确地恢复真实因果结构; 在高维数值实验中, 该方法在保持较高真阳性率和较低假阳性率的同时, 具有较好的计算效率和稳定性。总体而言, 所提方法通过优化传入 MCI 阶段的候选父节点集合, 提高了复杂时间序列因果推断的准确性和可靠性。

尽管该方法在上述实验中表现良好, 但仍然存在一些需要改进的问题。首先, 根据时间顺序(原因先于结果), 该方法只能推断滞后变量与当前变量之间的因果关系。然而, 当因果变化频率高于采样频率时, 同时因果关系将无法被发现。其次, 由于算法的假设, 该方法要求数据满足平稳性。但是, 由于数据预处理后会丢失部分信息, 非平稳模型仍有必要进行更深入的研究。因此, 未来应进一步研究同时因果和非平稳因果推断方法。

基金项目

西北民族大学生态系统恢复建模理论及其应用创新团队项目(10018632)。

参考文献

- [1] Klose, A.K., Donges, J.F., Feudel, U. and Winkelmann, R. (2024) Rate-Induced Tipping Cascades Arising from Interactions between the Greenland Ice Sheet and the Atlantic Meridional Overturning Circulation. *Earth System Dynamics*, **15**, 635-652. <https://doi.org/10.5194/esd-15-635-2024>
- [2] Sun, J., Yang, K., He, X., Wang, G., Wang, Y., Yu, Y., et al. (2025) Causal Pathways Underlying Global Soil Moisture-Precipitation Coupling. *Nature Communications*, **16**, Article No. 8935. <https://doi.org/10.1038/s41467-025-63999-7>
- [3] Langdon, C. and Engel, T.A. (2025) Latent Circuit Inference from Heterogeneous Neural Responses during Cognitive Tasks. *Nature Neuroscience*, **28**, 665-675. <https://doi.org/10.1038/s41593-025-01869-7>
- [4] Bachar, M. (2025) Sensitivity Analysis for a Delay Mathematical Model: The Glucose-Insulin Model. *Frontiers in Applied Mathematics and Statistics*, **11**, Article 1562636. <https://doi.org/10.3389/fams.2025.1562636>
- [5] Zhao, J., Gan, Z., Huang, R., Guan, C., Shi, J. and Leng, S. (2024) Detecting Dynamical Causality via Intervened Reservoir Computing. *Communications Physics*, **7**, Article No. 232. <https://doi.org/10.1038/s42005-024-01730-6>
- [6] Leite, A., Kislaya, I., Machado, A., Aguiar, P., Nunes, B. and Dias, C.M. (2025) Use of Quasi-Experimental Studies to Evaluate Causal Effects of Public Health Interventions in Portugal: A Scoping Review. *BMC Medical Research Methodology*, **25**, Article No. 263. <https://doi.org/10.1186/s12874-025-02701-3>
- [7] Haag, L., Starke, G., Ploner, M. and Ienca, M. (2025) Ethical Gaps in Closed-Loop Neurotechnology: A Scoping Review. *NPJ Digital Medicine*, **8**, Article No. 510. <https://doi.org/10.1038/s41746-025-01908-4>
- [8] Gopinathan, U., Elgersma, I., Dalsbø, T., Bjørnbæk, M. and Fretheim, A. (2025) Strengthening Research Preparedness for Crises: Lessons from Norwegian Government Agencies in Using Randomized Trials and Quasi-Experimental Methods to Evaluate Public Policy Interventions. *Health Research Policy and Systems*, **23**, Article No. 8. <https://doi.org/10.1186/s12961-024-01271-y>

- [9] Vance, T.C., Huang, T. and Butler, K.A. (2024) Big Data in Earth Science: Emerging Practice and Promise. *Science*, **383**, eadh9607. <https://doi.org/10.1126/science.adh9607>
- [10] Camps-Valls, G., Gerhardus, A., Ninad, U., Varando, G., Martius, G., Balaguer-Ballester, E., *et al.* (2023) Discovering Causal Relations and Equations from Data. *Physics Reports*, **1044**, 1-68. <https://doi.org/10.1016/j.physrep.2023.10.005>
- [11] Runge, J., Gerhardus, A., Varando, G., Eyring, V. and Camps-Valls, G. (2023) Causal Inference for Time Series. *Nature Reviews Earth & Environment*, **4**, 487-505. <https://doi.org/10.1038/s43017-023-00431-y>
- [12] Bailey, D.H., Jung, A.J., Beltz, A.M., Eronen, M.I., Gische, C., Hamaker, E.L., *et al.* (2024) Causal Inference on Human Behaviour. *Nature Human Behaviour*, **8**, 1448-1459. <https://doi.org/10.1038/s41562-024-01939-z>
- [13] Gong, C., Zhang, C., Yao, D., Bi, J., Li, W. and Xu, Y. (2024) Causal Discovery from Temporal Data: An Overview and New Perspectives. *ACM Computing Surveys*, **57**, 1-38. <https://doi.org/10.1145/3705297>
- [14] Cai, S., Peng, H., Liu, R. and Chen, P. (2025) Causal-Oriented Representation Learning for Time-Series Forecasting Based on the Spatiotemporal Information Transformation. *Communications Physics*, **8**, Article No. 242. <https://doi.org/10.1038/s42005-025-02170-6>
- [15] Mokhtarian, E., Elahi, S., Akbari, S., *et al.* (2025) Recursive Causal Discovery. *Journal of Machine Learning Research*, **26**, 1-65.
- [16] Allione, M., Del Totto, V. and Laio, A. (2025) Linear Scaling Causal Discovery from High-Dimensional Time Series by Dynamical Community Detection. *Physical Review Letters*, **135**, Article ID: 047401. <https://doi.org/10.1103/kd73-93cg>
- [17] Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. and Sejdinovic, D. (2019) Detecting and Quantifying Causal Associations in Large Nonlinear Time Series Datasets. *Science Advances*, **5**, eaau4996. <https://doi.org/10.1126/sciadv.aau4996>
- [18] Runge, J. (2020) Discovering Contemporaneous and Lagged Causal Relations in Autocorrelated Nonlinear Time Series Datasets. *Conference on Uncertainty in Artificial Intelligence. PMLR*, 3-6 August 2020, 1388-1397.
- [19] Günther, W., Ninad, U. and Runge, J. (2023) Causal Discovery for Time Series from Multiple Datasets with Latent Contexts. *Uncertainty in Artificial Intelligence. PMLR*, Pittsburgh, 31 July-4 August 2023, 766-776.
- [20] Gao, J., Li, C., Zhou, X., Yi, Y., Wang, X. and Liu, Q. (2025) Climate Change and Human Activities Amplify Runoff Variability Risks in Lower Reaches of Large Rivers. *Communications Earth & Environment*, **6**, Article No. 794. <https://doi.org/10.1038/s43247-025-02759-3>
- [21] Tian, Y., Giaquinto, D., Di Capua, G., Claassen, J.N., Ali, J., Li, H., *et al.* (2024) Historical Changes in the Causal Effect Networks of Compound Hot and Dry Extremes in Central Europe. *Communications Earth & Environment*, **5**, Article No. 764. <https://doi.org/10.1038/s43247-024-01934-2>
- [22] Gamella, J.L., Peters, J. and Bühlmann, P. (2025) Causal Chambers as a Real-World Physical Testbed for AI Methodology. *Nature Machine Intelligence*, **7**, 107-118. <https://doi.org/10.1038/s42256-024-00964-x>
- [23] Spirtes, P. and Glymour, C. (1991) An Algorithm for Fast Recovery of Sparse Causal Graphs. *Social Science Computer Review*, **9**, 62-72. <https://doi.org/10.1177/089443939100900106>
- [24] Kalisch, M. and Bühlmann, P. (2007) Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm. *Journal of Machine Learning Research*, **8**, 613-636.
- [25] Bang, C., Witte, J., Foraita, R., *et al.* (2026) Improving Finite Sample Performance of Causal Discovery by Exploiting Temporal Structure. *Statistical Methods in Medical Research*, **35**, 867-886.
- [26] Chang, T., Guo, Z. and Malinsky, D. (2026) Post-Selection Inference for Causal Effects after Causal Discovery. *Biometrika*, **113**, asaf073. <https://doi.org/10.1093/biomet/asaf073>
- [27] Nathaniel, J., Roesch, C., Buch, J., DeSantis, D., Rupe, A., Lamb, K.D., *et al.* (2025) Deep Koopman Operators for Causal Discovery. *Communications Physics*, **8**, Article No. 513. <https://doi.org/10.1038/s42005-025-02426-1>
- [28] Rao, S., Zou, G., Yang, S. and Barmada, S. (2024) A Feature Selection and Ensemble Learning Based Methodology for Transformer Fault Diagnosis. *Applied Soft Computing*, **150**, Article ID: 111072. <https://doi.org/10.1016/j.asoc.2023.111072>
- [29] Ullah, S., Siddique, M.F. and Kim, J. (2025) Multi-Sensor Observer-Based Residual Learning with Auto-Permutation Feature Importance for Fault Diagnosis of Multistage Centrifugal Pumps under Variable Pressures. *Scientific Reports*, **15**, Article No. 45735. <https://doi.org/10.1038/s41598-025-32726-z>
- [30] Eskandari, H., Saadatmand, H., Ramzan, M. and Mousapour, M. (2024) Innovative Framework for Accurate and Transparent Forecasting of Energy Consumption: A Fusion of Feature Selection and Interpretable Machine Learning. *Applied Energy*, **366**, Article ID: 123314. <https://doi.org/10.1016/j.apenergy.2024.123314>
- [31] Akman, D.V., Malekipirbazari, M., Yenice, Z.D., Yeo, A., Adhikari, N., Wong, Y.K., *et al.* (2023) K-Best Feature Selection and Ranking via Stochastic Approximation. *Expert Systems with Applications*, **213**, Article ID: 118864. <https://doi.org/10.1016/j.eswa.2022.118864>

- [32] Li, J., Zhang, F. and Ma, J. (2025) A New Representation in Genetic Programming with Hybrid Feature Ranking Criterion for High-Dimensional Feature Selection. *Complex & Intelligent Systems*, **11**, Article No. 185. <https://doi.org/10.1007/s40747-025-01784-1>
- [33] Han, Q., Hu, L. and Gao, W. (2024) Feature Relevance and Redundancy Coefficients for Multi-View Multi-Label Feature Selection. *Information Sciences*, **652**, Article ID: 119747. <https://doi.org/10.1016/j.ins.2023.119747>
- [34] Nie, F., Yang, S., Zhang, R. and Li, X. (2019) A General Framework for Auto-Weighted Feature Selection via Global Redundancy Minimization. *IEEE Transactions on Image Processing*, **28**, 2428-2438. <https://doi.org/10.1109/tip.2018.2886761>
- [35] Quinlan, J.R. (1986) Induction of Decision Trees. *Machine Learning*, **1**, 81-106. <https://doi.org/10.1023/a:1022643204877>
- [36] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., et al. (2017) Feature Selection: A Data Perspective. *ACM Computing Surveys*, **50**, 1-45. <https://doi.org/10.1145/3136625>
- [37] Gao, W., Hao, P., Wu, Y. and Zhang, P. (2023) A Unified Low-Order Information-Theoretic Feature Selection Framework for Multi-Label Learning. *Pattern Recognition*, **134**, Article ID: 109111. <https://doi.org/10.1016/j.patcog.2022.109111>
- [38] Qu, K., Xu, J., Hou, Q., Qu, K. and Sun, Y. (2023) Feature Selection Using Information Gain and Decision Information in Neighborhood Decision System. *Applied Soft Computing*, **136**, Article ID: 110100. <https://doi.org/10.1016/j.asoc.2023.110100>
- [39] Mielniczuk, J. (2022) Information Theoretic Methods for Variable Selection—A Review. *Entropy*, **24**, Article No. 1079. <https://doi.org/10.3390/e24081079>
- [40] Boschi, T., Testa, L., Chiaromonte, F. and Reimherr, M. (2025) Fasten: An Efficient Adaptive Method for Feature Selection and Estimation in High-Dimensional Functional Regressions. *Journal of Computational and Graphical Statistics*, **34**, 567-579. <https://doi.org/10.1080/10618600.2024.2407464>
- [41] Zhang, P., Xiu, N. and Qi, H.D. (2025) Sparse SVM with Hard-Margin Loss: A Newton-Augmented Lagrangian Method in Reduced Dimensions. *Journal of Machine Learning Research*, **26**, 1-55.
- [42] Bertsekas, D.P. (1982) Constrained Optimization and Lagrange Multiplier Methods. Academic Press, 383-392.
- [43] Lin, Z., Liu, R. and Su, Z. (2011) Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation. In: *Advances in Neural Information Processing Systems 24*, Curran Associates Inc., 612-620.
- [44] Curtis, F.E., Jiang, H. and Robinson, D.P. (2015) An Adaptive Augmented Lagrangian Method for Large-Scale Constrained Optimization. *Mathematical Programming*, **152**, 201-245. <https://doi.org/10.1007/s10107-014-0784-y>
- [45] Wallis, J.S. (1976) On the Existence of Hadamard Matrices. *Journal of Combinatorial Theory, Series A*, **21**, 188-195. [https://doi.org/10.1016/0097-3165\(76\)90062-5](https://doi.org/10.1016/0097-3165(76)90062-5)
- [46] Boyd, V., Vandenberghe, L. and Foybusovich, L. (2006) Convex Optimization. *IEEE Transactions on Automatic Control*, **51**, 1859-1859.
- [47] Runge, J. (2018) Conditional Independence Testing Based on a Nearest-Neighbor Estimator of Conditional Mutual Information. *International Conference on Artificial Intelligence and Statistics*. PMLR, Playa Blanca, 9-11 April 2018, 938-947.