

基于观测数据的时间序列因果推断综述

曾泽凡¹, 陈思雅¹, 龙 洗², 金 光¹

1. 国防科技大学系统工程学院, 湖南 长沙 410073;

2. 国防科技大学空天科学与工程学院, 湖南 长沙 410073

摘要: 数据存储量的扩大和计算能力的提升, 为基于观测数据推断时间序列的因果关系开辟了新途径。在时间序列因果推断的基本性质和研究现状的基础上, 系统梳理了 5 种基于观测数据的时间序列因果推断方法, 即 Granger 因果分析方法、基于信息论的方法、因果网络结构学习算法、基于结构因果模型的方法和基于非线性状态空间模型的方法。然后, 根据不同应用场景的数据特点, 结合方法的功能和适配性, 对基于观测数据的时间序列因果推断方法在经济金融、医疗和生物学、地球系统科学和其他工程领域的典型应用进行了简要介绍。最后, 结合时间序列因果推断的重难点问题, 比较 5 种方法的优缺点, 分析下一步研究重点, 展望未来的研究方向。

关键词: 时间序列; 因果推断; Granger 因果分析; 信息熵; 贝叶斯网络; 结构因果模型; 非线性状态空间模型

中图分类号:

文献标识码: A

doi:10.11959/j.issn.2096-0271.2022059

Overview of Observational Data-Based Time Series Causal Inference

ZENG Zefan¹, CHEN Siya¹, LONG Xi², JIN Guang¹

1. College of Systems Engineering, National University of Defense Technology, Changsha 410073,
China

2. College of Aerospace Science and Engineering, National University of Defense Technology,
Changsha 410073, China

Abstract: With the increase of data storage and the improvement of computing power, using observational data to infer time series causality has become a novel approach. Based on the properties and research status of time series causal inference, five observational data-based

收稿日期: 2022-03-09

通信作者:

基金项目:

Foundation Items:

methods are induced, including Granger causal analysis, information theory-based method, causal network structure learning algorithm, structural causal model-based method and method based on nonlinear state-space model. Then we briefly introduced typical applications in economics and finance, medical science and biology, earth system science and other engineering fields. Further, we compared the advantages and disadvantages and analyzed the ways for improvement of the five methods according to the focus and difficulties of time series causal inference. Finally, we looked into the future research directions.

Key words: time series, causal inference, Granger causal analysis, information entropy, Bayesian network, structural causal model, nonlinear state space model

1 引言

时间序列 (Time Series, TS) 又称动态数据, 是指将同一统计指标的数值按其发生的时间先后顺序排列而成的数列, 它反映了随机变量随时间不断变化的趋势^[1]。随着数据采集和处理技术的发展, 时间序列数据的可用性迅速增长, 且愈加呈现出大规模、多变量、强交互、噪声复杂等特点^[2]。大量高维的数据对时间序列数据挖掘与分析提出了更高的要求, 如何从序列中提取信息、发现规律, 建立更加准确、合理的模型, 为时间序列的预测、分类和异常检测等提供可靠的决策支持, 已经成为国内外时间序列研究的热门课题。

时间序列往往来自真实的系统, 系统的内部机制导致变量之间存在复杂的关联。相关关系是多变量系统中最常用的关联关系评价指标, 相关性分析通过统计分析、机器学习等方法, 发掘和分析变量之间的相关性^[3]。然而, 相关关系具有对称性, 只能反映变量在统计学意义上的同步变化趋势, 在分析具有时延影响和非对称、间接关联的时间序列时, 特别是应用于系统的机理分析时, 存在较大的局限性^[4]。进入 21 世纪后, 因果关系研究受到广泛关注。因果关系是指: 如果系统中 A 变量的变化必然 B 变量的相应变化, 但 B 变量的变化未必会引起 A 变量的变化, 则称 A 是 B 的“因”, B 是 A 的“果”。与相关关系相比, 因果关系可以发现时间序列的驱动因素, 在揭示事物发生机制和指导干预行为等方面有着相关关系不能替代的重要作用。

因果推断^[5]的目标是发现变量 (事物) 背后的因果关系。图灵奖得奖者 Judea Pearl^[6]把因果关系分成三个层面, 第一层是“关联”, 第二层是“干预”, 第三层是“反事实推理”。他认为当前的研究只处于第一层, 是“弱人工智能”, 要实现“强人工智能”还需要干预和

反事实推理，即因果推断。随机控制实验和计算机模拟实验是因果推断的传统方法，但由于实验技术局限、昂贵耗时、伦理限制等原因，这两种方法在实践中的应用范围十分有限，且效果不理想^[5]。数据挖掘技术的发展和计算机性能的提升为基于观测数据推断因果关系提供了可能。相比传统方法，基于观测数据的方法通过数据直接推断因果关系，不需要专家知识、繁琐的假设和复杂的实验，更加便捷实用。近年来，潜在因果模型^[7]和因果表示学习^[8]等方法在非时序数据的因果关系学习中表现出了良好的性能，在机器学习等领域得到的广泛运用，但这些方法多数适用于与时间无关的多维随机变量，因此本文不对这些方法作相关介绍。

随着基于观测数据的时间序列因果推断应用逐渐增多，国内外学者、高校和研究机构对相关理论、算法和应用研究的重视程度也不断增加。国内作相关研究较多的院校机构包括浙江大学、南京航空航天大学 and 广东工业大学等。特别是广东工业大学的郝志峰教授和蔡瑞初教授团队，在因果推断领域做了大量的研究工作，提出了很多先进的方法^[19,31,37,63,65]。国外开展时间序列因果推断相关研究的单位主要包括加利福尼亚大学、德国航空航天中心、德国马克斯普朗克智能系统研究所和克拉克森大学等，贡献突出的知名学者包括诺贝尔经济学奖获得者 Clive Granger, Judea Pearl, 以及加利福尼亚大学的 Schreiber 教授、德国航空航天中心的 Runge 教授和德国马克斯普朗克智能系统研究所 Sugihara 教授等。

目前，基于观测数据的时间序列因果推断的研究内容相对分散、不成体系，且很多研究工作仅仅围绕其具体领域的个性化需求，忽略了理论和应用的普适性。此外，高维序列的计算复杂度和推断精度、非平稳序列的虚假因果、隐变量的识别、因果强度的衡量等问题也制约了因果推断的可靠性。本文从时间序列因果推断的基本性质和研究现状出发，根据不同方法的理论依据和判定标准，将基于观测数据的时间序列因果推断方法归纳为 5 种，即 Granger 因果分析方法、基于信息论的方法、因果网络结构学习算法、基于结构因果模型的方法和基于非线性状态空间模型的方法。然后，结合不同应用场景数据特点，对这些方法在几个领域中的典型应用进行介绍。最后，根据时间序列因果推断的重难点，对 5 种方法的优缺点进行对比，分析下步研究重点，展望未来研究方向，力求为基于观测数据的时间序列因果推断的理论研究和实践应用提供可行的借鉴思路。

表 1 5 种基于观测数据的时间序列因果推断方法简介

方法名称	模型/指标	推断标准
Granger 因果分析方法	向量自回归模型	回归模型的预测误差
基于信息论的方法	信息熵	变量的不确定度（概率密度）
因果网络结构学习算法	因果贝叶斯网络	d-分离原则，条件独立性，评价函数等
基于结构因果模型的方法	结构因果模型	残差与变量序列的独立性

2 基于观测数据的时间序列因果推断方法

本节对 5 种基于观测数据的时间序列因果推断方法的基本原理进行概述,着重分析相关方法的建模思路以及同类别、不同分支方法之间的区别与联系。表 1 给出了这 5 种方法的基本介绍。

2.1 Granger 因果分析方法

最早的基于观测数据的时间序列因果推断方法是计量经济学家Granger^[9]在 1969 年提出的,称为Granger因果关系分析方法。Granger因果分析是一种定性的时间序列因果推断方法。该方法基于Wiener^[10]所提出的预测理论,其基本思想是:对于时间序列变量 X 和 Y ,如果 X 和 Y 共同建立的向量自回归 (Vector Auto-Regressive, VAR)模型比变量 Y 单独建立的VAR模型有更高的预测精度 (更小的预测误差),则称 X 为 Y 的Granger原因。Granger因果关系分析具有一定的可解释性,但只能适用于两变量、平稳、线性的时间序列^[11]。为了克服其缺点,学者们提出了很多变体,使其能够拓展到多变量、非线性和非平稳的情形中。

对于多元时间序列,Arize^[12]提出了多元Granger因果分析方法,该方法在VAR模型中添加了条件变量集,建立条件VAR模型,然后通过比较模型的预测误差来判定因果关系,添加条件变量集可以去除无关变量的影响。Gweke^[13]提出了条件Granger模型,其基本原理与多元Granger因果分析一致,在多元Granger模型的基础上利用矩阵表示来简化计算,并提出了一种更可靠的基于 χ^2 检验的因果判定方法,这两种方法为后续学者们研究高维序列的因果推断奠定了基础。

传统的Granger模型是线性模型,只能发现时间序列中的线性因果关系,而真实复杂系统中时间序列的因果关系往往是非线性的。为解决这一问题,许多拓展方法应运而生。Craig等^[14]通过估算时间序列概率分布和最大范数,提出一种基于非参数检验的Granger因果分析方法,该方法首先建立Granger因果模型,然后通过计算向量在不同时延下的最大范数来检验非线性因果关系。但该方法依赖于分布假设,可能会造成较多虚假因果。Ancona等^[15]利用机器学习中的径向基 (Radial Basis Function, RBF) 函数和Granger模型来实现两变量时间序列的非线性因果推断。该方法采用RBF将输入空间的特征向量映射到更高维的空间中,然后建立Granger因果模型,比Craig提出的方法具有更强的可靠性和鲁棒性。Liao等^[16]又提出了核Granger因果 (Kernel Granger Causality, KGC) 分析方法,该方法通过选择合适的特征

空间执行核函数，有效发现非线性因果关系。相比RBF-Granger方法，KGC可以根据需要选择不同的线性核、多项式核、Sigmoid核等，灵活性和适应性更强，精度也更高。

传统Granger因果分析方法在非平稳序列中会出现大量虚假因果，因此需要首先进行平稳性检验。针对此问题，Zhou等^[17]等利用改进的Hodrick-Prescott滤波器提取趋势分量，然后针对不同分量训练具有外生变量的自回归综合移动平均(Auto Regressive Integrate Moving Average, ARIMA)模型，并将其Granger因果分析相结合，采用F检验来判定因果关系。通过ARIMA模型的差分处理，实现了非平稳信号的因果分析。Schaeck等^[18]引入了稳健的时变广义部分定向相干函数，并将其与Granger因果模型相结合，提出了一种基于时变自回归(Temporal VAR, TVAR)参数的估计器，该方法使用分段线性时变移动平均模型来近似TVAR过程的残差，进而挖掘非线性因果关系的信息，比ARIMA具有更强的鲁棒性。Chen等^[19]提出了一种基于自适应模式非平稳Hawkes过程的Granger因果分析方法(Granger-Causal non-stationary Hawkes Process, GC-nsHP)，该方法使用动态规划算法将非平稳长过程划分为多个平稳子过程，然后迭代自适应划分非平稳过程，最后使用期望最大化算法来学习不同模式的因果关系。该方法可以有效解决点过程中的序列非平稳性，相较前两种方法拓展性更强。

高维时间序列的因果推断计算量大，且容易出现虚假因果，是时间序列因果推断的难点之一。为解决此问题，Schindler^[20]提出了Lasso-Granger因果模型，该方法在Granger因果模型中添加正则惩罚项，可以实现变量筛选，有效降低高维序列中的计算复杂度。Yang等^[21]采用群Lasso-Granger模型，利用不同的径向基函数集来近似每对节点之间的非线性相互作用，建立多元非线性Granger模型，然后利用L2 正则化将稀疏性整合到分组变量选择中，相比Lasso-Granger模型，该方法有着更高的推断精度。Bahadori和Liu^[22]提出了Copula-Granger模型，该模型结合了Lasso-Granger模型与潜在因果模型，利用“d-分离”规则排除混杂因素的影响，最后利用Granger“非超常分布”来识别因果关系。该方法相比传统的Lasso-Granger模型，在减少虚假因果和降低高维序列的计算复杂度方面表现突出。

2.2 基于信息论的方法

20 世纪 70 年代以来，随着信息熵理论的发展，通过传递熵^[23]等信息度量指标判定因果关系的方法逐渐兴起。相比Granger因果分析方法，基于信息论的方法可以衡量因果关系的强度，且在高维时间序列的因果推断中有更高的精度。这类方法基于“不确定性”（概率分布）的概念，认为如果变量 X 可以减少变量 Y 的不确定性，那么则认为 X 是 Y 的“因”， Y 是 X 的“果”^[24]。信息熵 $H(X) = -\sum_{x \in X} p(x) \log p(x)$ （ $p(x)$ 代表 X 的概率密度函数）是信息

论的基本指标，是系统混乱度的度量，也是不确定性的一种量化。

互信息（Mutual Information, MI）和传递熵（Transfer Entropy, TE）及其变体是基于信息论的因果推断方法的主要指标。MI是一个变量 X 中包含的关于另一个变量 Y 的信息量，也可以理解为 X 由于已知 Y 而减少的不确定性，即 $MI(X;Y)=H(X)-H(X|Y)$ 。MI可以量化两个变量之间的非线性依赖关系^[25]，它是一个非负量，如果MI显著不等于 0，则认为两个变量之间存在因果关系，反之亦然。条件互信息（Conditional MI, CMI）通过计算条件概率，可以将互信息拓展到多变量情形。但是，由于互信息具有对称性，因此在确定因果关系后还需要其他方法进一步定向。

TE是一种度量因果关系的非对称指标，它可以解释为如果 X 和 Y 的历史信息所决定的 X 的不确定度，小于单独通过 X 的历史信息所决定的 X 的不确定度，那么 Y 就是 X 的原因，即 $TE_{Y \rightarrow X}(t) = H(X_t | \mathbf{X}_{t-1}^k) - H(X_t | \mathbf{X}_{t-1}^k, \mathbf{Y}_{t-1}^l)$ 。如果 $TE_{Y \rightarrow X} > 0$ ，那么就认为存在 $Y \rightarrow X$ 的因果关系。类似地，通过添加条件变量，可以定义条件传递熵（Conditional TE, CTE），将其拓展到多元时间序列中。TE可以推断线性和非线性的因果关系，并且能够衡量因果关系的强弱。但TE在应用于非平稳序列因果推断时精度较低，高维序列中的变量选择也制约了其计算效率。为解决这些问题，学者们提出了很多TE的改进指标和方法。

对于非平稳序列因果推断问题，将时间序列符号化是最常用的手段。Staniek和Lehnertz^[26]提出了符号传递熵（Symbolic TE, STE）。STE通过重新排列时间序列的振幅值来定义符号，将输入变量转化为带有时延和嵌入维度的秩向量，然后依据秩向量的概率分布计算TE。STE利用数据的排序降低序列非平稳性（异质性）带来的影响，具有计算速度快和鲁棒性强的特点。Kugiumtzis^[27]提出了基于秩向量的偏传递熵，基于秩向量的偏传递熵的实质是STE在多变量条件下的拓展，通过添加条件变量集来排除无关变量的干扰。Rashidi等^[28]提出了符号动态归一化传递熵（Symbolic Dynamic-based Normalized Direct TE, SDNDTE）并将其应用于复杂系统根源变量定位中。SDNDTE首先利用归一化操作排除变量自信息对因果推断造成的影响，然后将时间序列转化为基于频率计数的状态转移矩阵，通过符号发射矩阵的估计代替传统多维核概率密度函数拟合方法。SDNDTE能不仅能推断多元非平稳序列中的因果关系，还能有效降低计算复杂度、减少虚假因果的出现。此外，Zeng等^[29]提出了归一化有效传递熵（Normalized Effective TE, NETE），NETE在保持数据概率分布不变的情况，将传递熵减去随机打乱数据后得到的随机传递熵，以消除时间序列的噪声和非平稳性造成的影响。

基于信息论的方法在应用于高维序列因果推断时存在计算复杂度大和冗余、虚假因果多

的问题。添加条件集、归一化操作和使用迭代算法是解决该问题的主要途径。Sun等^[30]提出了因果熵（Causation Entropy, CE）。CE在TE的基础上，添加目标变量的已知原因变量为条件集，然后用一种迭代方法评估和修正CE。CE有效克服了TE无法识别间接因果关系的缺点，减少了冗余因果。郝志峰等^[31]提出了归一化因果熵（Normalized CE, NCE），NCE在CE的基础上，将其除以目标（结果）变量的自信息，通过这种归一化处理改进了传统方法量纲不统一且冗余较多的问题，比CE的精度更高。Lizier^[32]提出了一种基于传递熵的多元有效源选择（Multivariate Effective Source Selection, MESS）算法。该方法首先进行信息分解，将变量的分解为自身的历史信息、源变量集（原因）提供的信息和不确定信息，然后通过贪婪迭代算法，逐步添加源变量，最后再进行显著性检验删除冗余。MESS可以在尽量减少虚假因果的前提下高效地实现高维序列的因果推断。为检验因果关系的显著性，去除虚假因果，Chen等^[33]在MESS的基础上提出了一种改进多元传递熵（Improved Multivariate TE, IMTE）方法，该方法通过时间序列随机置换和传递熵的均值检验，删除不显著的因果关系，确定因果关系的方向。

基于信息论的方法与Granger因果分析方法的原理和判断依据均不相同，但Barnett^[34]证明了对于高斯变量，Granger因果关系和传递熵推断得到的结果是完全等价的。

2.3 因果网络结构学习算法

1988年，Pearl^[35]提出了贝叶斯网络，为不确定知识表达和推理提供了新的有效理论模型。贝叶斯网络（Bayesian Network, BN）^[36]又称信念网络，是一种概率图模型，由结点及结点间的有向边构成。因果贝叶斯网络（Causal BN, CBN）^[37]是因果网络结构学习算法的基础，它在BN的基础上进一步规定有向边代表因果关系，出度节点代表原因变量，入度节点代表结果变量。从BN到CBN的拓展，通常需要满足忠诚性假设和因果马尔科夫条件^[38]等。此外，这些算法通常遵循莱辛巴赫共因原则^[5]，即如果变量是相关的，那么它们要么存在因果关系，要么是由共同的驱动因素驱动的。

因果网络结构学习算法效率较高，可以处理高维时间序列因果推断问题，也可以发现非线性的因果关系，已被开发用于时间序列因果网络的重构，这类方法也被称为基于约束的方法^[37]。它们通常从一个空图或完全连通图开始，以独立性或条件独立性为统计标准删除或添加边，逐步搜索建立因果网络架构。最经典的因果网络结构学习算法是PC（Peter-Clark）^[39]算法和IC（Inductive Causation）^[40]算法。PC算法包括“因果网络骨架学习”和“方向学习”两个阶段：首先建立一个完全无向图，然后借助d-分离规则^[41]和条件独立性检验迭代的删除边，得到部分有向无环图，最后再根据部分有向无环图定向规则^[37]进行定向。除了带

有时延的因果关系之外，PC算法还可以用于同时期因果关系的推断。IC算法的流程与PC算法相反，它根据部分有向图的结构特点，从一个空的图开始，利用条件独立性信息逐步迭代添加边，最后定义一种重构算法来确定方向。

PC和IC算法虽然适用于高维时间序列，但其鲁棒性较差，某一步的推断错误可能就会使推断结果发生很大改变。此外，由于马尔科夫等价类^[5]的存在，导致其定向比较困难，数据非平稳性也会导致其推断精度下降。为了克服PC和IC算法的缺点，学者们研发了多种因果网络结构学习算法。Chickering^[42]提出了一种启发式算法——贪婪等价搜索（Greedy Equivalence Search, GES）。GES与IC算法的原理相似，首先建立一个空图，然后定义分数函数来评价特定图结构的可能性，确定等价类空间，最后在这个搜索空间内迭代地选择、添加和删除边。由于缩小了搜索空间，GES相比PC和IC算法都有更低的时间复杂度，但GES也会导致得到局部最优解而非全局最优解。Tsamardinos等^[43]提出了最大最小爬山算法（Max-Min Hill Climbing, MMHC），这种方法首先利用最大最小父子算法^[44]学习因果无向图，然后用贪婪贝叶斯评分爬山搜索方法对无向图进行定向。相比GES，MMHC在搜索环节采用了随机重启爬山搜索和禁忌搜索策略，可以有效防止陷入局部最优解。Huang等^[45]提出了一种用于非平稳和异构时间序列因果推断的CD-NOD（Constraint-based causal Discovery from Nonstationary/heterogeneous Data）方法。该方法利用变量代理和条件独立性检验重构因果网络框架，然后利用潜在因果模型中隐含的数据分布进行定向。相比此前的方法，CO-NOD是非参数的，对数据分布没有严格限制，且不依赖于时间窗口分割，还能在带有混杂因素的异质数据中识别因果关系。

传统的因果网络结构学习算法依赖于因果充分性假设^[5]，即假设系统中不存在未观测的共同驱动因素，但真实系统中往往存在未观测变量（隐变量）。Spirtes等^[46]针对隐变量的问题，提出一种快速因果推断（Fast Causal Inference, FCI）算法。FCI以祖先图模型为基础，首先建立完全连通图，然后提出基于概率依赖性和独立性的定向方法，通过拓展搜索空间来输出因果信息，可以克服隐变量（混淆偏差）和选择偏差的影响。Colombo等^[47]提出了RFCI（Really FCI）算法，RFCI避免了对不必要的d-分离子集进行条件独立性检验，只对邻接集的子集施加条件。RFCI不仅可以允许任意多个隐变量，还能解决FCI在高维情况下计算复杂度过大的问题。Ogarrio等^[48]提出了GFCI（Greedy FCI）和GRFCI（Greedy Really FCI）算法，针对FCI算法对样本需求量大的问题，调整了d-分离集合和定向的方法，能在小样本的情形下发现因果关系，确保渐进准确性。Kummerfeld等^[49]基于多协方差子矩阵的排序约束和条件独立性检验，提出了一种FTFC（Find Two Factor Clusters）算法，该算法可以处理隐

变量、测量误差和非线性循环关系，具有更强的鲁棒性。

大规模序列是指采样时间长、维度高的序列。在大规模序列中，条件独立性信息的获取困难且耗时，这导致其因果推断存在计算复杂度大、精度低的问题。Runge教授对高维大规模时间序列因果网络推断的理论和实践作出了较大贡献，特别是对未观测变量、数据采样、测量误差、动态噪声、自相关和显著性检验等问题做了深入细致的研究^[50]。他在2019年提出了PCMCI算法^[51]，通过条件选择(PC)和瞬时条件独立性(MCI)检验，解决了自相关高维和非线性时间序列数据的特殊挑战，实现了大规模时间序列因果网络重构。2020年，Runge^[52]又提出了PCMCI+算法，通过分离存在时延和同期的条件集，以及修改单个条件独立性测试的条件集，可以在高召回率的前提下有效减少虚假因果，降低时间复杂度。

2.4 基于结构因果模型的方法

2006年，Shimizu等^[53]提出了结构因果模型(Structural Causal Model, SCM)。SCM通过映射函数来关联原因变量和结果变量。假设时间序列变量 X 为原因， Y 为结果，且 X 为 Y 的直接原因，那么可以建立关于 X 和 Y 的结构因果模型为： $X := E_X; Y := f_Y(X, E_Y)$ 。 E_X 和 E_Y 分别代表变量 X 和 Y 的噪声， f_Y 代表因果函数，且 E_X 与 E_Y 相互独立。基于SCM的因果推断方法从数据产生的因果机制出发，利用因果函数模型来确定因果关系和识别因果方向^[54]。

最早的基于结构因果模型的方法是Shimizu等^[53]提出的线性非高斯无环模型(Linear Non-Gaussian Acyclic Model, LiNGAM)。LiNGAM假设数据生成过程为线性，系统中无未观测的共因，且误差服从非高斯分布，然后根据因果机制的不对称性(或机制独立性原则)^[5]和独立成分分析(ICA)^[55]，判定因果方向。首先分别建立 X 对 Y 的线性回归模型和 Y 对 X 的线性回归模型，如果残差 r_t^X 与 Y_t 相互独立，而残差 r_t^Y 与 X_t 不相互独立，就可以认为 X 是 Y 的原因。基于ICA的LiNGAM具有局部收敛的缺陷，在多变量情况下难以适用。Shimizu等^[56]又提出了非参数的DirectLiNGAM框架。DirectLiNGAM通过识别外生变量来估计因果次序，控制在多变量情况下因果推断的假阳性。随着样本量的增大，DirectLiNGAM可以保证在固定的小步数内渐进收敛到正确的解。

LiNGAM方法依赖于对噪声分布的高斯假设，这给模型的应用造成了很大的限制。Hoyer等^[57]提出了加性噪声模型(Additive Noise Model, ANM)方法，ANM假设噪声以可加项的形式作用于因果变量之间，它不需要对噪声的非高斯假设，且不局限于线性回归。此外，ANM判定因果方向的方法除了残差独立性检验，还可以利用后续独立检验^[58]或类熵分数^[59]

等方法。Zhang和Hyvärinen^[60]提出了后非线性因果模型（Post Non-Linear Causal Model, PNLCM）。PNLCM在LiNGAM的基础上，对原因变量和噪声项进行两次非线性变换，建立复合的非线性模型。通过向量化和矩阵计算，LiNGAM，ANM和PNLCM都很容易拓展到多个变量的情形。此外，通过数据平移可以实现有时延的时间序列因果推断。

对于含隐变量的时间序列，Tashiro和Shimizu^[61]在LiNGAM的基础上提出了ParcelLiNGAM算法，通过检验估计回归残差与外生变量的独立性来学习因果关系的阶数、找到包含未被隐变量所影响的变量子集，进而发现隐变量，该方法对混淆偏差具有鲁棒性。Hoyer等^[62]基于LiNGAM模型，提出适用于含隐变量系统的lvLiNGAM（latent variable LiNGAM）框架。lvLiNGAM基于祖先图，将与观测变量有高连接强度集的潜在变量组合为单个隐变量，然后通过调整连接强度将隐变量标准化为零均值和单位方差。Cai等^[63]通过引入非高斯假设，提出了一种隐变量因果结构学习算法LSTC（Learn the Structure of latent variables based on Triad Constraints）。LSTC提出了一种“伪残差”，通过残差与原始数据的条件独立性检验，发现可观测变量与隐变量之间的因果方向，相比LiNGAM方法，LSTC能实现可观测变量较少情况下的隐变量识别。

此外，部分学者们还研究了信息论与结构因果模型相结合的方法。Janzing等^[64]提出了信息几何因果推断（Information Geometric Causal Inference, IGCI）方法。IGCI将信息熵与ANM相结合，利用信息空间的正交性来定义独立性，进而判定因果方向。Xie等^[65]提出了一种基于熵的ETPIA算法，该算法利用信息熵可以衡量任意分布数据背后信息量的特性，将LiNGAM拓展到具有高斯噪声的数据中，还能防止测量误差导致的因果方向误判。

2.5 基于非线性状态空间模型的方法

2012年，Sugihara等^[66]提出了收敛交叉映射(Convergent Cross Mapping, CCM)方法，该方法基于Takens定理^[67]，通过状态空间重构推断因果关系，引起了国内外研究学者的关注。基于非线性状态空间模型的方法早在20世纪初就已被应用于神经科学中，近年来更是在生态、气象科学等多个领域得到广泛运用，在大型复杂系统中应用最多。

状态空间模型是描述和揭示系统内部关系和运动规律的有效形式。基于非线性状态空间模型的时间序列因果推断方法假设交互作用发生在一个潜在的动力系统中，然后基于Takens定理和非线性状态空间重构来推断因果关系。Takens定理可以用于重构时间序列中的动力学信息。它证明在满足某些条件时，从一个吸引子到重构空间的映射是一一对应的，只要找到合理的嵌入维数，就能实现相空间中轨道的重构，并保持其原来的微分结构不变。

基于非线性状态空间模型的因果推断方法中最经典的是非线性相互依赖度量

(Non-Linear Interdependence Measure, NLIM) 和收敛交叉映射。Arnold等^[68]提出了非线性相互依赖度量,首次将非线性状态空间模型用于发现具有“非对称性”的相互依赖,即因果关系。对于两个时序变量 \mathbf{X} 和 \mathbf{Y} ,通过在多维相空间中嵌入时延,得到相空间向量 \mathbf{x} 和 \mathbf{y} ,分别计算在空间 \mathbf{x} 中和 \mathbf{y} 映射到 \mathbf{x} 的空间中的 k 个近邻点的欧氏距离。然后根据距离度量计算全局相互依赖度量 $S(\mathbf{X}|\mathbf{Y})$ 。如果 $S(\mathbf{X}|\mathbf{Y})$ 显著大于 0,那么认为存在 \mathbf{X} 到 \mathbf{Y} 的因果关系。同理可得到 $S(\mathbf{Y}|\mathbf{X})$,通过比较 $S(\mathbf{X}|\mathbf{Y})$ 和 $S(\mathbf{Y}|\mathbf{X})$ 的大小,可以判定在相互依赖关系中哪个变量更加“活跃”。在文献^[68]中还提出了另一种全局相互依赖度量 H ,它是平均均方欧氏距离与条件均方欧氏距离的几何均值,对弱因果关系更敏感。此后,研究学者们陆续提出了 N ^[69]、 M ^[70]和 L ^[71]等改进度量指标,这些指标具有更高的准确性和更强的鲁棒性。

根据Takens定理,如果变量 X 可以通过变量 Y 的时延嵌入重构系统来预测,那么 X 和 Y 之间就存在因果效应。CCM是一种利用Takens定理和影子流形^[66]来推断非线性系统因果关系的方法。CCM推断因果关系的基本依据是交叉映射原理,即用 X 构造影子流形 M_X ,用 Y 构造影子流形 M_Y ,如果存在 $X \rightarrow Y$ 因果关系,那么用 M_Y 中某点的邻居点就能识别 M_X 中对应点的邻居点。CCM具有收敛性,即如果 X 和 Y 存在因果关系的话,那么使用的时间序列越长(样本量越大),所得到的交叉映射估计误差就会越小。

除了NLIM和CCM方法之外,还有部分学者提出了其他基于非线性状态空间模型的方法,这些方法多数是与基于信息论方法的结合。比如,Mao等^[72]提出一种基于相空间时延重构和传递熵相结合的因果推断方法PSDR (Phase State Delay Reconstruction)-TE。该方法将多元时间序列通过时延重构转化为矩阵形式,然后根据矩阵的多元分布函数计算标准化方向性传递熵。PSDR-TE可以克服传递熵只能推断单个变量之间因果关系的缺点,有效发现非标量多元时间序列之间的因果关系。Krug等^[73]基于相互作用单元的非线性动力学,提出了一种利用元胞非线性网络 (Cellular Nonlinear Network, CNN) 估计时间序列非线性相互依赖关系的方法。该方法首先计算NLIM,然后利用CNN逼近求解NLIM,测量不对称的非线性相互依赖关系。Krakovská和Jakubík^[74]利用重构状态空间预测原理,提出了基于交叉预测 (Cross Prediction, CP) 和混合预测 (Predictability Improvement, PI) 的因果关系推断方法。其中CP方法通过自预测和交叉预测的平均绝对误差判定因果关系,PI则通过两种混合优化预测方法判定因果关系。CP和PI能适用于多变量系统,还能用于因果推断结果的敏感性分析。

尽管上述 5 种方法的原理和依据不同,但其本质也有不少相通之处。比如Granger因果分析、基于SCM的方法和CCM都是基于时间序列的趋势预测,基于信息论的方法和因果网络结构学习算法的都是基于数据的统计概率。

3 应用现状

基于观测数据的时间序列因果推断已经在经济金融、生物医疗、地球系统科学和工业工程等众多领域得到了广泛应用，在时间序列的机理分析、预测、分类和异常检测等方面发挥着重要的作用。本节，针对不同应用领域的数据特点以及 5 种方法在不同应用场景下的适配性和功能，对相关领域的典型应用进行介绍。

3.1 经济金融领域

时间序列是经济和金融领域最常见的数据类型。受到多种因素的综合影响，经济金融时间序列具有信噪比低、易分离、非平稳和非线性显著的特点^[76]，且其维数往往不高。因此，Granger因果分析和基于信息论的方法在该领域中应用最为广泛。Chimobi^[77]将Granger因果分析和增强单位根（ADF）检验相结合，研究分析了金融发展、贸易开放和经济增长之间的因果关系；Fiedor^[78]提出一种结合传递熵的拓展Granger因果关系，并将其应用于构建具有特定时间滞后的股票因果网络，以调查市场的短期因果关系；Sebri等^[79]基于Granger因果分析，提出了协整和向量误差修正模型的边界检验方法，发现了经济增长与可再生能源消费之间存在双向Granger因果关系，解释了可再生能源对金砖国家经济增长的拉动作用；Judge^[80]提出一种基于因果熵的二进制网络行为重构方法，并将其用于研究开放动态微观经济系统中自适应智能行为和自组织均衡寻求行为之间的关系；Liu等^[81]利用条件熵检测反馈驱动的交易和反映市场回报流的“自因果性”，并使用传递熵识别新闻情绪和市场回报的信息流相关的交易活动。

3.2 医学和生物学领域

在医学和生物学领域，病理分析、疾病诱发机制研究和受试者行为分析，动植物的结构、功能、发生和发展规律研究，以及脑科学和神经科学领域各类信号的分析，都涉及时间序列因果关系。这些领域的时序变量具有实际物理含义，维数较高且部分变量难以观测。此外，为了对不同变量之间影响展开比较分析，往往需要测量因果关系的强度^[82]。因此，基于信息论的方法和因果网络结构学习算法在该领域中应用相对较多。Walter等^[83]提出了信息因果熵-复杂度的方法，用以研究壁厚和心室压力的变化与缺血性壁行为的系统动力学，为心室壁供氧后心肌行为的恢复提供了客观指标。Stephan等^[84]基于贝叶斯网络模型提出一种动态因果建模的方法，该方法能够提供神经生物学上可解释数量的后验估计，如神经元群体间突触连接的有效强度和它们的环境依赖性调节等。Lee等^[85]使用传递熵和核密度评估因果关系，实现了观测量少且存在异常值的定向耦合变化检测，并将其应用于呼吸化学反射系统中。

Faes等^[86]利用Granger因果分析与传递熵相结合的方法, 识别与显著信息传递相关的时间滞后集, 从而量化短期生理调节的重要机制和疾病潜伏期的变化。Valenza等^[87]提出一种瞬时点过程传递熵 (ipTE), ipTE可以同步跟踪高分辨率非线性耦合中的时变变化, 用于研究心血管和呼吸动力学的非线性相互作用。

3.3 地球系统科学 (气象、海洋、地理、生态) 领域

研究地球系统科学中时间序列的因果关系, 对探寻其运行规律, 揭示现象背后的机制原理有着非常重要的意义。地球系统科学的研究对象是大规模的复杂动力系统, 其中包含着形式多样的因果关系。这些关系往往呈现出非平稳性、非线性、耦合性等特点^[54]。因此, Granger因果分析、基于信息论的方法和基于非线性状态空间模型的方法在该领域使用较多, 且常以相结合的方式得以应用。Suay等^[88]提出基于一种ANM和灵敏度依赖性估计准则的因果推断方法, 并将其应用于碳循环问题等 28 个地球科学因果推理问题中。Silva等^[89]利用调整最大滞后窗口的Granger因果分析方法, 评估季节性降水与厄尔尼诺-南方涛动相关的海面温度模式的响应。Sugihara^[66]将CCM方法用于生态系统的因果关系推断, 准确检测出太平洋沙丁鱼着陆、北鳀鱼着陆和加利福尼亚州斯克里普斯码头和新港码头测量的海表面温度之间的关系。Reyes等^[90]利用频域Granger因果分析和统计显著检验法, 证明了太平洋年代际振荡异常等区域变动是凤尾鱼繁殖和身体状况指标的主要驱动因素。Li和Convertino^[91]提出一种互信息和传递熵的最优信息流生态系统模型, 通过从时间序列中提取复杂生态系统的预测因果网络, 提供广泛的生态信息。Oh等^[92]提出一种相对符号传递熵来研究南大洋的南极绕极波大规模气候现象, 并解释了厄尔尼诺-南方涛动的特征。

3.4 其他工程领域

除了上述领域外, 基于观测数据的时间序列因果推断在机械、制造、网络、航天、工业等领域也有较多应用, 特别是在原理分析、根因诊断、异常检测等方面发挥了重要作用。这些工程数据容量大、采样时间短、交互性强, 且往往呈现出低价值密度^[93]。基于信息论的方法由于能够衡量因果关系强度, 在这些领域中应用最为广泛, 其他方法也在部分领域中得到运用。Yu等^[94]利用传递熵对工业报警数据中的二进制报警序列进行因果关系检测。Shi等^[95]利用传递熵方法, 有效解决了 4 种网络物理系统攻击检测问题。Zeng等^[96]利用基于归一化修正传递熵和改进的因果网络结构学习算法, 将遥测参数因果关系与注意力机制的LSTM相结合, 实现了低误报率的航天器遥测数据异常检测。Yao等^[97]采用Granger因果分析和最小生成树对工业网络进行建模, 从能源消耗的角度研究工业耦合机制。Yoshida等^[98]使用因果贝叶斯网络集成机器人的听觉、视觉和运动信息, 估计主动运动对噪声鲁棒性语音活动检测

性能的影响。Tian等^[99]提出一种基于CCM的重大工业事故主因告警和根因追踪方法，利用CCM识别变量之间的因果方向和间接因果关系，作为后续告警根因追踪的依据。

4 比较分析与展望

5种基于观测数据的时间序列因果推断方法，由于理论依据、推断过程、判定指标等的不同，导致其优缺点各不相同。本节，结合时间序列因果推断的不同情形和重难点问题，对这5类方法的优势、不足进行比较，分析下一步研究重点。最后，展望基于观测数据的时间序列因果推断的未来研究方向。

4.1 比较分析

Granger因果分析方法。Granger因果分析的优点是容易理解、可解释性强，适用于多变量、非平稳、非线性时间序列。其不足在于：要求误差满足正态分布假设；易受噪声干扰，鲁棒性不强；是一种定性方法，难以衡量因果关系强弱；高维情形计算复杂度高；一般要求全部变量可观测。今后的研究重点应解决噪声依赖、计算复杂度、隐变量识别和因果强度衡量等问题。

基于信息论的方法。基于信息论的方法能够判定因果关系方向、衡量因果关系强弱，适用于非线性、非平稳、多变量时间序列，在离散数据因果关系推断时较其他方法更优越。其不足在于：对高维序列的可靠性较低；存在未观测变量时会推断出很多虚假因果；因果关系显著性判定方法比较缺乏^[75]。未来研究重点应针对在高维数据的熵估计、多变量条件概率的计算、因果关系显著性判定和隐变量结构识别等方面。

因果网络结构学习算法。此类方法是非参数方法，可以发现有时延或同时期的因果关系，也可以用于含有隐变量的时间序列中，在高维数据因果推断中计算效率高，能在某种程度上评估不确定性，还能将专家知识融入因果推断。主要不足是：鲁棒性比较差，某步推断出现问题可能会大大影响最终结果；无法衡量因果关系强度；高维、非线性、非平稳数据获取的条件独立性信息具有不确定性。下步应针对迭代搜索算法优化和鲁棒性提升进行研究，探索更加准确有效的非线性、非平稳、高维数据的条件独立性检验方法。

基于SCM的方法。此类方法的形式简单，参数较少，计算方便，在检测同时期因果和含有隐变量的因果关系方面效果较好。但目前此类方法在时间序列因果推断领域的应用较少，主要原因是：用于推断时延因果关系时，需要对不同的时延分别建立模型，过程十分繁琐；多数方法需要对噪声变量作非高斯假设，但实践中往往难以满足；对于高维序列，模型的函数和参数较多，计算复杂度很大。此类方法未来应进一步向多变量序列拓展，致力解决时延

因果推断、高维序列建模和噪声假设依赖等问题。

基于非线性状态空间模型的方法。这类方法比其他方法更适用于大型复杂系统。特别是对不可分离、非纯随机、弱耦合的确定性非线性系统，效果较好。但有以下局限：对于强耦合、高随机性的系统，其推断结果可靠性不高；无法识别隐变量；难以应用在高维时间序列中；无法衡量因果关系的强弱；难以识别瞬时因果关系。大型复杂非线性系统的因果推断是时间序列因果推断的难点之一，此类方法未来潜力巨大，应着重在解决高维、强耦合和多噪声序列中的计算效率和鲁棒性问题上下功夫。

表2 根据时间序列的不同情形和重难点,给出了5种方法的优缺点对比和主要应用领域。

表 2 5 种基于观测数据的时间序列因果推断方法的优缺点对比

主要方法		多 变 量	非 线 性	非 平 稳	隐 变 量	因 果 强 度	时 延 因 果	瞬 时 因 果	噪 声 依 赖	高 维 序 列	应用领域	参考 文献
Granger 因果分析	Granger 因果分析						✓		✓			[9]
	多元 Granger 因果分析	✓					✓		✓			[12]
	条件 Granger 因果分析											[13]
	非线性 Granger 因果分析		✓				✓		✓		经济、金融、	[14]
	RBF Granger 因果分析										生物、医疗、	[15]
	Kernel Granger 因果分析	✓	✓				✓		✓		航天、气象、	[16]
	Lasso-Granger 因果分析	✓					✓		✓	✓	地理、生态、	[20]
	Copula Granger 因果分析	✓	✓				✓		✓		各种工程领 域	[22]
	时变自回归 (TVAR)	✓	✓	✓			✓		✓			[18]
	ARIMA-Granger 因果分析	✓	✓	✓			✓		✓			[17]
	nsHP-GC	✓	✓	✓	✓		✓		✓			[19]
基于信息论的方法	群 Lasso-Granger 因果分析	✓	✓				✓		✓	✓		[21]
	TE		✓			✓	✓					[24]
	CTE	✓	✓			✓	✓					[24]
	STE		✓	✓		✓	✓				金融、神经	[26]
	PSTE	✓	✓	✓		✓	✓				科学、生物、	[27]
	SDNDTE	✓	✓	✓		✓	✓			✓	气象、地理、	[28]
	NETE	✓	✓	✓		✓	✓				生态、制造、	[29]
	CE	✓	✓			✓	✓				网络、航天、	[30]
	NCE	✓	✓			✓	✓				机械	[31]
	MESS	✓	✓			✓	✓			✓		[32]
	IMTE	✓	✓			✓	✓			✓		[33]
因果网络结构学习算法	PC	✓	✓				✓	✓		✓		[39]
	IC	✓	✓				✓	✓		✓		[40]
	GES	✓	✓				✓	✓		✓		[42]
	MMHC	✓	✓				✓	✓		✓		[43]
	CD-NOD	✓	✓	✓	✓		✓	✓		✓	气象、生态、	[44]
	FCI	✓	✓		✓		✓	✓		✓	工业、机械 等	[46]
	RFCI	✓	✓		✓		✓	✓		✓		[47]
	GRFCI											[48]
	PCMCi	✓	✓				✓	✓		✓		[51]
	PCMCi+	✓	✓				✓	✓		✓		[52]
	LiNGAM							✓	✓		经济金融和 遥感	[53]
基于 SCM 的方法	DirectLiNGAM	✓						✓	✓			[56]
	ParceLiNGAM	✓			✓			✓	✓			[61]

	lvLiNGAM	✓		✓	✓	✓		[62]
	LSTC	✓	✓	✓				[63]
	ANM	✓	✓			✓		[57]
	PNLCM	✓	✓			✓		[60]
	IGCI	✓	✓			✓		[64]
	ETPIA	✓	✓	✓		✓	✓	[65]
基于非线性状态空间的方法	NLIM: S, H, N, M, L		✓	✓		✓	地球系统科学、生态、	[68-70]
	CCM		✓	✓		✓	脑神经科	[66]
	PSDR-TE	✓	✓	✓		✓	学、天文学、	[72]
	CNN		✓	✓		✓	工业、机械	[73]
	CP 和 PI	✓	✓	✓		✓		[74]

4.2 展望

因果推断已经成为时间序列分析的研究重点之一。基于观测数据的时间序列因果推断方法成本低、速度快，相比基于控制实验和模拟实验的因果推断方法更加高效可行。随着研究的不断深入，多元、非线性、非平稳时间序列的因果推断等问题已逐步得到有效解决。然而，大规模、高维、含隐变量时间序列的因果推断精度以及因果强度的衡量仍是亟待解决的难题，特别是在大规模高维数据中如何合理选择条件变量，准确估计概率密度，以及消除偏差影响等。时序数据采样尺度、间隔和样本量的不同也给基于观测数据推断因果关系带来了很大的挑战。此外，如何将现有的因果推断方法与其他模型相结合，进一步提升其可用性和可靠性，也是未来需要研究的重点内容。综上所述，今后的研究工作可以从以下几个方面展开：

（1）对于高维、大规模时间序列，可以Granger因果分析方法和基于信息论的方法为基础，通过字典学习、稀疏表示、信息分解、贪婪算法等现有方法，降低变量选择和概率计算的复杂度，提升估计精度。

（2）对于隐变量识别问题，可以将因果网络结构学习算法与Granger因果分析、基于信息论的方法相结合，利用因果网络结构学习算法能够发现隐变量的特点，解决隐变量和数据分布带来的混淆偏差和选择偏差。

（3）将潜在因果模型和贝叶斯评估等方法与现有的时间序列因果推断方法相结合，解决时间聚合、时间尺度、小样本数据、离散数据、不确定性评估等问题，利用反事实推理验证解释推断结果。

（4）将现有的方法与相关应用领域的先验知识或物理模型、数学模型相结合，比如可靠性评估中的退化模型，以及自动控制系统中的传递模型、响应模型等，提升因果推断模型的可解释性。

（5）发挥机器学习、深度学习等先进技术在数据的模式学习方面的特有优势，利用卷积神经网络、循环神经网络、生成对抗网络等模型实现时间序列的因果推断。

（6）效仿Causeme.net网站^[54]和JIDT工具箱^[32]等平台 and 工具，设计开发更加高效实用、

方便快捷的因果推断软件或网络平台，推动理论成果向实践工具的转化。

参考文献：

- [1] BROCKWELL P J , DAVIS R A , BERGER J O , et al. Time Series: Theory and Methods[M]. Berlin, Springer-Verlag, 2015: 2-35.
- [2] 韩敏,任伟杰,李柏松,冯守渤.混沌时间序列分析与预测研究综述[J].信息与控制,2020,49(01):24-35.
HAN M, REN W ,LI B, et al. A Review of Chaotic Time Series Analysis and Forecasting[J].Information and Control (Chinese),2020,49(01):24-35.
- [3] PEARL J , GLYMOUR M , JEWELL N P . Causal Inference in Statistics: A Primer [M]. New Jersey, John Wiley & Sons, 2016: 22-58.
- [4] 任伟杰,韩敏.多元时间序列因果关系分析研究综述[J].自动化学报,2021,47(01):64-78.
REN W, HAN M. A review of causality analysis of multiple time series[J].Acta Automation Sinica,2021,47(01):64-78.
- [5] PARUNAK H . Elements of causal inference: foundations and learning algorithms[J]. Computing reviews, 2018, 59(11):588-589.
- [6] PEARL, J, MACKENZIE, D. The Book of Why: The New Science of Cause and Effect [M]. London, Allen Lane Press, 2018: 126-155.
- [7] YAO L , CHU Z , LI S , et al. A Survey on Causal Inference[J]. ACM Transactions on Knowledge Discovery from Data, 2021, 15(5):1-46.
- [8] SCHÖLKOPF B, LOCATELLO F, BAUER S, et. al. Toward Causal Representation Learning[C]//in Proceedings of the IEEE, 2021, 109(5), 612-634.
- [9] GRANGER C , WIENER J. Investigating causal relations by econometric models and cross-spectral methods[J], Econometrica, 1969, 37: 424-438.
- [10] N. WIENER. The Theory of Prediction. I n Modern Mathematics for Engineers[M].New York, NY, McGraw-Hill, 1956: 368-398.
- [11] SETH A K, BARRETT A B, BARNETT L. Granger causality analysis in neuroscience and neuroimaging[J]. The Journal of Neuroscience, 2015, 35(8): 3293–3297.
- [12] AUGUSTINE C. ARIZE. Determinants of Income Velocity in the United Kingdom: Multivariate Granger Causality [J].The American Economist,1993, 37(2): 40 – 45.

- [13] GEWEKE, JOHN. Measurement of Linear Dependence and Feedback between Multiple Time Series[J]. Publications of the American Statistical Association, 1982, 77(378):304-313.
- [14] CRAIG, HIEMSTRA, JOHNATHAN, et al. Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation[J]. The Journal of Finance, 1994, 49(5):1639-1664.
- [15] ANCONA N , MARINAZZO D , STRAMAGLIA S . Radial basis function approach to nonlinear Granger causality of time series[J]. Phys Rev E Stat Nonlin Soft Matter Phys, 2004, 70(5 Pt 2):056221.
- [16] WEI L , D MARINAZZO, PA N Z , et al. Kernel Granger Causality Mapping Effective Connectivity on fMRI Data[J]. 2009, 28(11):1825-1835.
- [17] ZHOU Y , KANG Z , LIN Z , et al. Causal analysis for non-stationary time series in sensor-rich smart buildings[C]// Proceedings of IEEE International Conference on Automation Science & Engineering. IEEE, 2013: 593-598.
- [18] SCHAECK T , MUMA M , FENG M , et al. Robust Nonlinear Causality Analysis of Non-Stationary Multivariate Physiological Time Series[J]. IEEE Transactions on Biomedical Engineering, 2017, 65(6):1213-1225.
- [19] WEI CHEN, JIBIN CHEN , RUICHU CAI , et al. Learning granger causality for non-stationary Hawkes processes[J]. Neurocomputing, 2022, 468:22-32.
- [20] SCHINDLER K . Granger Lasso Causal Models in Higher Dimensions - Application to Gene Expression Regulatory Networks[C]// Proceedings of Machine Learning and Knowledge Discovery in Databases European Conference. 2013: 1-14.
- [21] YANG G , WANG L , WANG X . Reconstruction of Complex Directional Networks with Group Lasso Nonlinear Conditional Granger Causality[J]. Scientific Reports, 2017, 7(1):2991.
- [22] BAHADORI, T., Y. LIU, Y. An Examination of Large-Scale Granger Causality Inference[C]// Proceedings of SIAM Conference on Data Mining (SDM'13),2013: 2301-2309.
- [23] SCHREIBER, THOMAS. Measuring Information Transfer[J]. Physical Review Letters, 2000, 85(2):461-464.
- [24] BOSSOMAIER T , BARNETT L , M HARRÉ, et al. Transfer Entropy[M]. Springer International Publishing, 2016.

- [25] KRASKOV A , H STOEGBAUER, GRASSBERGER P . Estimating Mutual Information[J].Physical Review E, 2004, 69:066138.
- [26] STANIEK M , LEHNERTZ K . Symbolic Transfer Entropy[J]. Physical Review Letters, 2008, 100(15):158101.
- [27] KUGIUMTZIS D . Partial transfer entropy on rank vectors[J]. European Physical Journal Special Topics, 2013, 222(2):401-420.
- [28] RASHIDI B , ZHAO Q . Autonomous Root-Cause Fault Diagnosis Using Symbolic Dynamic Based Causality Analysis[J]. Neurocomputing, 2020, 401(3):10-27.
- [29] ZENG Z, JIN G, XU C, CHEN S, ZHANG L. Spacecraft Telemetry Anomaly Detection Based on Parametric Causality and Double-Criteria Drift Streaming Peaks over Threshold[J]. Applied Sciences, 2022; 12(4):1803.
- [30] SUN J , BOLLT E M . Causation Entropy Identifies Indirect Influences, Dominance of Neighbors and Anticipatory Couplings[J]. Physica D Nonlinear Phenomena, 2014, 267:49-57.
- [31] 郝志峰, 谢蔚涛, 蔡瑞初,等. 基于因果强度的时序因果关系发现算法[J]. 计算机工程与设计, 2017, 38(1):6.
- HAO ZF, XIE WT, CAI RC, et al. Time-series causal relationship discovery algorithm based on causal strength. Computer Engineering and Design, 2017, 38(1):6.
- [32] LIZIER J , RUBINOV M . Multivariate construction of effective computational networks from observational data[J]. Avian Diseases, 2012, 30(1):1-2.
- [33] CHEN S , JIN G , MA X . Detection and Analysis of Real-time Anomalies in Large-Scale Complex System[J]. Measurement, 2021(2):109929.
- [34] BARNETT L , BARRETT A B , SETH A K . Granger causality and transfer entropy are equivalent for Gaussian variables[J]. Physical Review Letters, 2009, 103.
- [35] NEAPOLITAN R E . Learning Bayesian Networks[M].New Jersey, Prentice Hall, 2003: 1-33.
- [36] 张连文, 郭海鹏. 贝叶斯网引论[M]. 北京: 科学出版社, 2006.
- Zhang L W, Guo H P. Introduction to Bayesian Networks [M]. Beijing: Science Press, 2006.
- [37] 蔡瑞初, 郝志峰.大数据中的因果关系发现[M].北京: 科学出版社, 2018: 22-49.
- CAI R H, HAO Z F. Causal Relationship Discovery in Big Data . Beijing: Science Press,

2018: 22-49.

- [38] SPIRITES P, GLYMOUR C, SCHEINES R. Causation, Prediction, and Search[M]. Cambridge, MIT Press, 2000.
- [39] MARKUS KALISCH, PETER BÜHLMANN. Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm[J]. Journal of Machine Learning Research, 2007, 8:613-636.
- [40] VERMA T, PEARL J. Equivalence and synthesis of causal models[C]// Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence. Cambridge, UK, 1990: 255-268.
- [41] HUND L, SCHROEDER B. A causal perspective on reliability assessment[J]. Reliability Engineering and System Safety, 2019, 195:106678.
- [42] CHICKERING D M. Learning Equivalence Classes of Bayesian Networks Structures[J]. [arxiv: 1302/1302.3566.pdf](https://arxiv.org/abs/1302.1302), 2013.
- [43] TSAMARDINOS I, BROWN L E, ALIFERIS C F. The max-min hill-climbing Bayesian network structure learning algorithm[J]. Machine Learning, 2006, 65(1): 31-78.
- [44] TSAMARDINOS I, ALIFERIS C F, STATNIKOV A. Time and sample efficient discovery of Markov blankets and direct causal relations[C]//Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, USA, 2003: 673-678.
- [45] HUANG B, ZHANG K, ZHANG J, et al. Causal Discovery from Heterogeneous/Nonstationary Data[J]. [arXiv:1903.01672v5](https://arxiv.org/abs/1903.01672)[cs.LG], 2019.
- [46] SPIRITES P, MEEK C, RICHARDSON T. An Algorithm for causal inference in the presence of latent variables and selection bias[J]. Computation Causation & Discovery, 2016:211-252.
- [47] COLOMBO D, MAATHUIS M H, KALISCH M, et al. Learning high-dimensional directed acyclic graphs with latent and selection variables[J]. The Annals of Statistics, 2012, 40(1): 294-321.
- [48] OGARRIO J M, SPIRITES P, RAMSEY J. A hybrid causal search algorithm for latent variable models[C]//Proceedings of Conference on Probabilistic Graphical Models. 2016: 368-379.

- [49] KUMMERFELD E, RAMSEY J, YANG R, et al. Causal clustering for 2-factor measurement models//Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2014). Nancy, France, 2014: 34-49.
- [50] RUNGE J . Causal network reconstruction from time series: From theoretical assumptions to practical estimation[J]. Chaos, 2018, 28(7):075310.
- [51] RUNGE J , SEJDINOVIC D , FLAXMAN S . Detecting causal associations in large nonlinear time series datasets[J]. [arXiv:1702.07007v2](https://arxiv.org/abs/1702.07007v2) [stat.ME],2017.
- [52] RUNGE J . Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets[J]. [arXiv:2003.03685v2](https://arxiv.org/abs/2003.03685v2) [stat.ME], 2020.
- [53] SHIMIZU S, HOYER P O, HYVÄRINEN A, et al. A linear non-Gaussian acyclic model for causal discovery[J]. Journal of Machine Learning Research, 2006, 7: 2003-2030.
- [54] RUNGE J , BATHIANY S , BOLT E , et al. Inferring causation from time series in Earth system sciences[J]. Nature Communications, 2019, 10(1).
- [55] AAPO HYVÄRINEN, HOYER P O , INKI M . Topographic ICA as a Model of Natural Image Statistics[C]// Proceedings of IEEE International Workshop on Biologically Motivated Computer Vision. Springer-Verlag, 2000: 83-88.
- [56] SHIMIZU S, INAZUMI T, SOGAWA Y, et al. DirectLiNGAM: A direct method for learning a linear non-gaussian structural equation model[J]. Journal of Machine Learning Research, 2011, 12(2): 1225-1248.
- [57] HOYER P O, JANZING D, MOOIJ J M, et al. Nonlinear causal discovery with additive noise models[C]//Proceedings of the 23rd Annual Conference on Neural Information Processing Systems (NIPS 2009). Vancouver, Canada, 2009: 689-696.
- [58] M. MOOIJ, D. JANZING, J. PETERS, AND B. SCHÖLKOPF. Regression by dependence minimization and its application to causal inference[C]// Proceedings of the 26thInternational Conference on Machine Learning (ICML), 2009,: 745–752.
- [59] NOWZOHOUR , P. BÜHLMANN. Score-based causal learning in additive noise models[J]. Statistics, 2016, 50(3):471–485.
- [60] ZHANG K, HYVÄRINEN A. On the identifiability of the post-nonlinear causal model[C]//Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI 2009). Montreal, Canada, 2009: 647-655.

- [61] TASHIRO T, SHIMIZU S, HYVÄRINEN A, et al. ParceLiNGAM: A causal ordering method robust against latent confounders[J]. *Neural Computation*, 2014, 26(1): 57-83.
- [62] HOYER P O, SHIMIZU S, KERMINEN A J, et al. Estimation of causal effects using linear non-Gaussian causal models with hidden variables[J]. *International Journal of Approximate Reasoning*, 2008, 49(2): 362-378.
- [63] CAI R, XIE F, GLYMOUR C, et al. Triad Constraints for Learning Causal Structure of Latent Variables[C]//*Advances in Neural Information Processing Systems*. 2019: 12863-12872.
- [64] JANZING D, MOOIJ J, ZHANG KUN, et al. Information-geometric approach to inferring causal directions[J]. *Artificial Intelligence*, 2012, 182: 1-31.
- [65] XIE F, CAI R, ZENG Y, et al. An Efficient Entropy-Based Causal Discovery Method for Linear Structural Equation Models With IID Noise Variables[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, PP(99):1-14.
- [66] SUGIHARA G, MAY R, YE H, et al. Detecting Causality in Complex Ecosystems[J]. *Science*, 2012, 338(6106):496-500.
- [67] F TAKENS. *Dynamical Systems and Turbulence* [M]. New York, Springer-Verlag, 1981:67-88.
- [68] ARNHOLD J, GRASSBERGER P, LEHNERTZ K, et al. A Robust Method for Detecting Interdependences: Application to Intracranially Recorded EEG[J]. *Physica D-nonlinear Phenomena*, 1999, 134(4):419-430.
- [69] QUIROGA R Q, ARNHOLD J, GRASSBERGER P. Learning driver-response relationships from synchronization patterns[J]. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 2000, 61(5 Pt A):5142-5148.
- [70] ANDRZEJAK R G, KRASKOV A, STOGBAUER, et al. Bivariate surrogate techniques: Necessity, strengths, and caveats[J]. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2003, 68(6):066202.
- [71] CHICHARRO D, ANDRZEJAK R G. Reliable detection of directional couplings using rank statistics[J]. *Physical Review E Statal Nonlinear & Soft Matter Physics*, 2009, 80(2):026217.
- [72] MAO X, SHANG P. Transfer entropy between multivariate time series[J]. *Communications in Nonlinear Science and Numerical Simulation*, 2016, 47:338-347.

- [73] KRUG D , OSTERHAGE H , ELGER C E , et al. Estimating nonlinear interdependences in dynamical systems using cellular nonlinear networks.[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2007, 76(4):041916.
- [74] ANNA KRAKOVSKÁ, JOZEF JAKUBÍK. Implementation of two causal methods based on predictions in reconstructed state spaces[J]. Physical Review E, 2020, 102:022203.
- [75] CHEN S , JIN G , MA X . Satellite On-orbit Anomaly Detection Method Based on a Dynamic Threshold and Causality Pruning[J]. IEEE Access, 2021, PP(99):1-1.
- [76] 张贵生. 数据驱动的金融时间序列预测模型研究[D].太原: 山西大学,2016.
Zhang GS. Research on Data-Driven Financial Time Series Prediction Model [D]. Taiyuan: Shanxi University, 2016.
- [77] CHIMOBİ, OMOKE, PHILIP. The causal Relationship among Financial Development, Trade Openness and Economic Growth in Nigeria.[J]. International Journal of Economics & Finance, 2010, 2(2):137-147.
- [78] FIEDOR P . Causal Non-Linear Financial Networks[J]. Papers, [arXiv:1407.5020v1](https://arxiv.org/abs/1407.5020v1)[q-fin.ST] 2014.
- [79] SEBRI M , BEN-SALHA O . On the causal dynamics between economic growth, renewable energy consumption, CO2 emissions and trade openness: Fresh evidence from BRICS countries[J]. Renewable & Sustainable Energy Reviews, 2014, 39:14-23.
- [80] JUDGE G . Adaptive Intelligent Behavior and Causal Entropy Maximization as a Basis for Microeconomic Information Recovery[J]. SSRN Electronic Journal, 2014. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.2463323>.
- [81] LIU A , CHEN J , YANG S Y , et al. The Flow of Information in Trading: An Entropy Approach to Market Regimes[J]. Entropy, 2020, 22(9):1064.
- [82] Sarkar M , Leong T Y . Characterization of medical time series using fuzzy similarity-based fractal dimensions[J]. Artificial Intelligence in Medicine, 2003, 27(2):201-222.
- [83] WALTER L , FRANCISCO T , REDELICO F O , et al. Analysis of ischaemic crisis using the informational causal entropy-complexity plane[J]. Chaos, 2018, 28(7):075518-.
- [84] STEPHAN K E , PENNY W D , MORAN R J , et al. Ten simple rules for dynamic causal modeling[J]. Neuroimage, 2010, 49(4):3099-3109.

- [85] LEE J , NEMATIS S , SILVA I , et al. Transfer Entropy Estimation and Directional Coupling Change Detection in Biomedical Time Series[J]. BioMedical Engineering OnLine, 2012, 11(1):19.
- [86] FAES L , MARINAZZO D , MONTALTO A , et al. Lag-specific transfer entropy as a tool to assess cardiovascular and cardiorespiratory information transfer[J]. IEEE transactions on bio-medical engineering, 2014, 61(10):2556-68.
- [87] VALENZA G , FAES L , CITI L , et al. Instantaneous transfer entropy for the study of cardio-respiratory dynamics[C]// Proceedings of Engineering in Medicine & Biology Society. IEEE, 2015:7885-7888.
- [88] PEREZ-SUAY A, CAMPS-VALLS G. Causal inference in geoscience and remote sensing from observational. Data[J]. IEEE Transactions on Geoscience and Remote Sensing. 2018, 3: 1502–1513.
- [89] SILVA F N , VEGA-OLIVEROS D A , YAN X , et al. Detecting climate teleconnections with Granger causality[J]. <https://doi.org/10.48550/arXiv.2012.03848>, 2020.
- [90] CONTRERAS-REYES J E , CAROLA HERNÁNDEZ-SANTORO. Assessing Granger-Causality in the Southern Humboldt Current Ecosystem Using Cross-Spectral Methods[J]. Entropy, 2020, 22(10):1071.
- [91] LI J , CONVERTINO M . Inferring Ecosystem Networks as Information Flows[J]. Scientific Reports, 2021, 11:7094.
- [92] OH M , KIM S , LIM K , et al. Time series analysis of the Antarctic Circumpolar Wave via symbolic transfer entropy[J]. Physica A Statistical Mechanics & Its Applications, 2018, 499:233-240.
- [93] 郑树泉, 覃海焕, 王倩. 工业大数据技术与架构[J]. 大数据, 2017, 3(4):14.
Zheng S Q, Qin H H, Wang Q. Industrial Big Data Technology and Architecture [J]. Big Data Research, 2017, 3(4):14.
- [94] YU W , YANG F . Detection of Causality between Process Variables Based on Industrial Alarm Data Using Transfer Entropy[J]. Entropy, 2015, 17(8):5868-5887.
- [95] SHI D , GUO Z , JOHANSSON K H , et al. Causality Countermeasures for Anomaly Detection in Cyber-Physical Systems[J]. IEEE Transactions on Automatic Control, 2017, 63(2):386-401.

- [96] ZENG Z, JIN G, XU C, et al. Satellite Telemetry Data Anomaly Detection Using Causal Network and Feature-Attention-based LSTM [J], IEEE Transactions on Instrumentation and Measurement, 2022, Early Access, doi:[10.1109/TIM.2022.3151930](https://doi.org/10.1109/TIM.2022.3151930).
- [97] YAO C Z, LIN J N, LIN Q W, et al. A study of causality structure and dynamics in industrial electricity consumption based on Granger network[J]. Physica A: Statistical Mechanics and its Applications, 2016, 462:297-320.
- [98] YOSHIDA T, NAKADAI K. Active audio-visual integration for Voice Activity Detection based on a Causal Bayesian Network[C]// Proceedings of IEEE-RAS International Conference on Humanoid Robots. IEEE, 2013:370-375.
- [99] TIAN C, ZHAO C, FAN H, et al. Causal network construction based on convergent cross mapping (CCM) for alarm system root cause tracing of nonlinear industrial process[J]. IFAC-Papers Online, 2020, 53(2):13619-13624.

[作者简介]



曾泽凡（1993- ），男，国防科技大学系统工程学院硕士研究生，主要研究方向为数据分析与数据建模。



陈思雅（1998- ），女，国防科技大学系统工程学院博士研究生，主要研究方向为时间序列异常检测、故障诊断。



龙洗（1999- ），男，国防科技大学空天科学与工程学院博士研究生，主要研究方向为航天任务规划，因果推断，强化学习。



金光（1973- ）男，博士，国防科技大学系统工程学院，研究员，主要研究方向为寿命预测与健康管理、系统试验与评估。