

多元时间序列因果关系分析研究综述

任伟杰^{1,2} 韩敏^{1,2}

摘要 多元时间序列的因果关系分析是数据挖掘领域的研究热点. 时间序列数据包含着与时间动态有关的、未知的、有价值的信息, 因此若能挖掘出这些知识进而对时间序列未来趋势进行预测或干预, 具有重要的现实意义. 为此, 本文综述了多元时间序列因果关系分析的研究进展、应用与展望. 首先, 本文归纳了主要的因果分析方法, 包括 Granger 因果关系分析、基于信息理论的因果分析和基于状态空间的因果分析; 然后, 总结了不同方法的优缺点、适用范围和发展方向, 并概述了其在不同领域的典型应用; 最后, 讨论了多元时间序列因果分析方法待解决的问题和未来研究趋势.

关键词 多元时间序列, Granger 因果分析, 转移熵, 状态空间

引用格式 任伟杰, 韩敏. 多元时间序列因果关系分析研究综述. 自动化学报, 2021, 47(1): 64–78

DOI 10.16383/j.aas.c180189

Survey on Causality Analysis of Multivariate Time Series

REN Wei-Jie^{1,2} HAN Min^{1,2}

Abstract The causality analysis of multivariate time series is a research hotspot in data mining. Time series data contains unknown, valuable information related to temporal dynamics. Therefore, it is of great practical significance to be able to mine these knowledge and then predict or intervene the future trend of time series. For this reason, this paper reviews the research progress, application and prospects of causality analysis of multivariate time series. Firstly, this paper summarizes the main causality analysis methods, including Granger causality analysis, causality analysis based on information theory and causality analysis based on state space. Then, we summarize the advantages and disadvantages, scope of application and development directions of different methods, and outline their typical applications in different fields. Finally, the problems to be solved and future research trends of the causality analysis methods of multivariate time series are discussed.

Key words Multivariate time series, Granger causality analysis, transfer entropy, state space

Citation Ren Wei-Jie, Han Min. Survey on causality analysis of multivariate time series. *Acta Automatica Sinica*, 2021, 47(1): 64–78

时间序列是指现实世界中的某个观测变量, 按照其发生的时间先后顺序排列的一组数字序列. 时间序列可以分为一元时间序列和多元时间序列, 多元时间序列是指多个一元时间序列的组合, 可以认为是一次采样中可以获得不同来源的多个观测变量. 多元时间序列广泛存在于自然^[1]、医学^[2]、社会^[3]、

工业^[4]等各个领域的复杂系统中, 多个变量之间具有复杂的关联关系, 相互影响作用不明确. 随着数据采集和存储技术的发展, 时间序列数据的维度和规模不断增加, 为建立准确的预测模型增加了难度. 同时, 随着数据维度的增加, 出现了大量冗余和无关变量, 容易掩盖重要变量的作用, 对模型的建立产生负面的影响^[5]. 时间序列数据挖掘^[6-7]是当前研究的热门问题, 研究如何有效地从多元时间序列中挖掘潜在的有用信息、构建预测模型, 能够为自然、医学、社会、工业等领域的控制、决策与调控提供理论指导, 具有十分重要的现实意义^[8]. 因此, 本文主要研究多元时间序列的分析手段, 解释未知系统的动力学特性与运行规律, 从而为建立更加精确的系统模型奠定基础.

在多变量系统中, 通过分析可观测变量之间的相关关系, 可以找出对建模贡献度大的相关变量, 从而推断出系统的运行机理. 目前, 多元时间序列相关性分析主要集中于统计学手段, 例如 Pearson

收稿日期 2018-04-02 录用日期 2018-11-22

Manuscript received April 2, 2018; accepted November 22, 2018

国家自然科学基金 (61773087), 中央高校基本科研业务费 (DUT18RC(6)005) 资助

Supported by National Natural Science Foundation of China (61773087) and Fundamental Research Funds for the Central Universities (DUT18RC(6)005)

本文责任编辑 何海波

Recommended by Associate Editor HE Hai-Bo

1. 大连理工大学电子信息与电气工程学部 大连 116023 2. 大连理工大学工业装备智能控制与优化教育部重点实验室 大连 116023

1. Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116023 2. Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, Dalian University of Technology, Dalian 116023

相关系数、秩相关系数、典型相关分析^[9]、互信息^[10]、最大信息系数^[11]、灰色关联分析^[12]、Copula 分析^[13]等. 这些方法能够有效处理线性或非线性相关关系, 其分析结果具有对称性. 然而, 多个变量之间不仅存在直接相互作用, 还存在以中间变量为桥梁的间接相互作用, 并且影响关系通常具有非对称性. 传统的相关性分析方法难以处理间接关系、非对称影响关系, 在实际应用中受到很大限制.

随着系统复杂度的增加, 相关性分析难以满足建模需求, 因果关系分析方法得到广泛关注^[14]. 因果关系是一个系统(因)与另一个系统(果)之间的作用关系, 其中第 1 个系统是第 2 个系统的原因, 第 2 个系统依赖于第 1 个系统. 1969 年, Granger^[15]首次提出了一种评价二变量时间序列之间是否存在相互作用的因果关系分析方法, 即 Granger 因果关系分析方法. 该方法基于系统的可预测性, 基本思想是: 对于两个时间序列, 如果一个时间序列未来时刻的预测误差, 能够通过引入另一个时间序列的历史信息而减小, 则称第 2 个时间序列对第 1 个时间序列具有因果影响. 由于传统的 Granger 因果分析建立在线性模型的基础上, 仅对二元时间序列进行分析, 在提出之后出现了大量改进模型^[16]. Granger 因果分析方法具有很强的可解释性, 但是此类方法只能给出定性分析结果, 并且对于高维时间序列容易产生虚假因果现象. 基于信息测度的因果分析是一类非参数方法, 包括转移熵、条件熵、条件互信息等, 这类方法通过建立评价函数, 能够定量分析因果关系的强弱^[17]. 此外, 基于状态空间的因果模型^[18]、贝叶斯网络等模型^[19-20], 同样用于分析各种类型的因果关系. 因此, 针对多变量系统的建模要求, 合理利用因果分析方法的优点, 研究系统各个变量之间的驱动响应关系, 进而推断系统内部结构和运行机理, 是当前研究的热点问题^[21].

综上所述, 相比于常规的相关性分析方法, 因果分析方法能够分析出具有方向性的直接因果关系, 更加适用于多变量系统的分析与建模. 本文针对多元时间序列因果关系分析的几类典型方法进行综述, 包括 Granger 因果关系分析、基于信息理论的因果分析和基于状态空间的因果分析, 并结合当前流行的机器学习方法、不同领域时间序列建模的需求等, 讨论因果分析方法的实际应用和未来发展趋势.

1 Granger 因果关系分析

Granger 因果关系是由诺贝尔经济学奖得主

Granger 提出的一种因果关系分析模型, 在金融时间序列分析中发挥了重要作用, 目前已经成为自然、医学等领域普遍使用的因果模型. 本节将对 Granger 因果分析及其改进模型的基本原理、适用范围进行分析与总结.

1.1 Granger 因果分析基本方法

Granger 因果关系分析作为一种判别二元时间序列之间是否存在因果关系的方法, 从提出以来受到研究人员的广泛关注. 这种思想最初由 Wiener 提出, 之后由 Granger 通过随机过程的线性回归模型实现^[15], 其基本思想是: 若采用时间序列 X 和 Y 的历史信息对 Y 进行预测, 优于仅采用 Y 的历史信息对 Y 进行预测的结果, 即时间序列 X 有助于解释时间序列 Y 的未来变化趋势, 那么时间序列 X 是时间序列 Y 的 Granger 原因. 建立如下两个向量自回归 (Vector autoregressive, VAR) 模型:

$$\begin{aligned} Y_{t+1} &= \sum_{j=0}^{m-1} \alpha_j Y_{t-j} + \varepsilon_{Y,t+1} \\ Y_{t+1} &= \sum_{j=0}^{m-1} a_j X_{t-j} + \sum_{j=0}^{m-1} b_j Y_{t-j} + \varepsilon_{Y|X,t+1} \end{aligned} \quad (1)$$

其中, α_j , a_j 和 b_j 为模型的系数, m 为模型的阶数, ε_Y 和 $\varepsilon_{Y|X}$ 为模型的残差. 根据回归预测结果, 通过比较 VAR 模型残差的方差大小, 判断 $X \rightarrow Y$ 是否存在 Granger 因果关系, Granger 因果指数 (Granger causality index, GCI) 定义为

$$GCI_{X \rightarrow Y} = \ln \frac{\text{var}(\varepsilon_Y)}{\text{var}(\varepsilon_{Y|X})} \quad (2)$$

如果满足 $\text{var}(\varepsilon_{Y|X}) < \text{var}(\varepsilon_Y)$, 即 $GCI_{X \rightarrow Y} > 0$, 表明 $X \rightarrow Y$ 存在统计意义下的 Granger 因果关系. 采用类似步骤, 我们可以对 $Y \rightarrow X$ 进行 Granger 因果检验. 需要特别注意, 进行 Granger 因果关系分析的前提是时间序列为平稳序列, 否则可能出现虚假因果.

在提出之后的几十年中, Granger 因果模型在许多领域都得到了广泛的应用, 但是由于其建立在线性模型的基础上, 如果直接应用于非线性系统, 在某些情况下将失去作用. 此外, 多变量系统存在复杂的因果关系, GCI 仅进行两个变量的因果分析, 处理复杂系统的能力严重不足. 因此, 学者们提出了大量改进模型^[22], 用于分析多变量、非线性系统的因果关系, 包括条件 Granger 因果模型、Lasso-Granger 因果模型、非线性 Granger 因果模型和频域 Granger 因果模型等, 具体如表 1 所示.

表 1 Granger 因果关系分析及其改进方法
Table 1 Granger causality analysis and its improvement methods

类别	研究者	发表年份	方法名称	文献
Granger因果模型	Granger	1969	Granger 因果指数 (GCI)	[15]
	Geweke	1982	条件 Granger 因果指数 (CGCI)	[23]
条件Granger因果模型	Chen 等	2004	条件扩展 Granger 因果指数 (CEGCI)	[24]
	Siggiridou 等	2016	限制条件 Granger 因果指数 (RCGCI)	[25]
Lasso-Granger因果模型	Arnold 等	2007	Lasso-Granger 因果模型	[26]
	Shojaie 等	2010	截断 Lasso-Granger 因果模型	[27]
	Bolstad 等	2011	Grouped-Lasso-Granger 因果模型	[28]
	Yang 等	2017	Grouped-Lasso 非线性条件 Granger 因果模型	[29]
	Ancona 等	2004	RBF-Granger 因果模型	[30]
非线性Granger因果模型	Marinazzo 等	2008	Kernel-Granger 因果模型	[31-32]
	Wu 等	2011	KCCA-Granger 因果模型	[33]
	Hu 等	2014	Copula-Granger 因果模型	[34]
	Montalto 等	2015	NN-Granger 因果模型	[35]
频域Granger因果模型	Geweke	1982	Spectral-Granger 因果模型	[23]
	Baccalá 等	2001	偏定向相干性 (PDC)	[36]
	Kamiński 等	2001	直接传递函数 (DTF)	[37]

1.2 条件 Granger 因果模型

传统的 Granger 因果模型仅用于分析两个变量之间的因果关系, 对于多变量系统的因果分析, 通常不考虑多个变量之间的关联关系, 直接转化为多个二变量问题进行分析. 然而, 多变量系统的变量之间存在直接或间接的联系, 在分析任意两个变量因果关系时, 可能存在中间变量的影响, 此时传统的 Granger 因果模型容易产生虚假因果. 为了解决上述问题, Geweke^[23] 提出了条件 Granger 因果分析方法, 引入条件变量, 建立两个 VAR 模型:

$$\begin{aligned}
 Y_{t+1} &= \sum_{j=0}^{m-1} \alpha_j Y_{t-j} + \sum_{j=0}^{m-1} \beta_j Z_{t-j} + \varepsilon_{Y|Z,t+1} \\
 Y_{t+1} &= \sum_{j=0}^{m-1} a_j X_{t-j} + \sum_{j=0}^{m-1} b_j Y_{t-j} + \\
 &\quad \sum_{j=0}^{m-1} c_j Z_{t-j} + \varepsilon_{Y|XZ,t+1}
 \end{aligned} \quad (3)$$

其中, Z 表示条件变量. 条件 Granger 因果指数 (Conditional Granger causality index, CGCI) 定义为

$$\text{CGCI}_{X \rightarrow Y|Z} = \ln \frac{\text{var}(\varepsilon_{Y|Z})}{\text{var}(\varepsilon_{Y|XZ})} \quad (4)$$

条件 Granger 因果模型建立在多变量回归模型基础上, 通过将条件变量加入到回归模型中, 有效区分变量的直接和间接联系, 得到直接因果关系. Chen等^[24] 引入非线性预测思想, 提出了条件扩展

Granger因果指数 (Conditional extended Granger causality index, CEGCI), 应用于多元混沌时间序列因果分析. 由于回归模型 (3) 中包含很多待求参数, 对于包含大量观测变量的系统, 其计算过程十分复杂且容易失效. Siggiridou 等^[25] 引入了延迟变量选择策略, 限制 VAR 模型的阶数, 提出了限制条件 Granger 因果指数 (Restricted conditional Granger causality index, RCGCI), 成功应用于高维时间序列的因果分析.

1.3 Lasso-Granger 因果模型

针对多变量系统因果分析, Granger 因果模型和条件 Granger 因果模型需要对任意两个变量进行 Granger 因果检测, 具有很高的计算复杂度. 尽管建立 VAR 模型可以考虑多个变量之间的相互影响, 仍然难以获得理想的分析结果. 针对高维变量 Granger因果分析问题, Arnold 等^[26] 提出了 Lasso-Granger因果模型, 根据输入变量选择的结果识别 Granger因果关系, 其基本思想是: 应用全部输入变量进行 Lasso 回归, 根据模型回归系数识别 Granger 因果关系的强弱. 目标函数如下所示:

$$\min \left\{ \|Y - \mathbf{X}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \right\} \quad (5)$$

其中, Y 为预测变量, \mathbf{X} 为全部输入变量, $\boldsymbol{\alpha}$ 为回归系数, λ 为正则化参数, 用于控制惩罚项大小. 如果时间序列 X_j 对应的系数 α_j 为零或接近于零, 则表明时间序列 $X_j \rightarrow Y$ 不存在 Granger 因果关系,

反之则存在 Granger 因果关系. Lasso-Granger 因果模型通过建立一个回归模型, 分析出全部输入变量对预测变量的因果关系, 大大缩减了计算量.

Shojaie 等^[27]提出了截断 Lasso-Granger 因果模型, 能够准确估计时间序列回归模型的阶数, 从而提高模型的计算准确度. 为避免群组效应, Bolstad 等^[28]提出了 Grouped-Lasso-Granger 因果模型, 能够减少错误因果关系的产生. Yang 等^[29]提出了 Grouped-Lasso 非线性条件 Granger 因果模型, 该方法利用不同集合的径向基函数近似非线性关系, 并结合群组变量选择算法, 将 Lasso-Granger 因果模型扩展到非线性复杂网络重构.

1.4 非线性 Granger 因果模型

传统的 Granger 因果模型仅用于分析线性因果关系, 随着应用范围的扩大以及研究的逐渐深入, 人们发现大量系统存在非线性因果关系, 从而涌现出很多非线性 Granger 因果模型. 根据前面介绍的 Granger 因果分析方法可以看出, 已经有学者提出应用非线性预测理论改进线性模型, 实现由线性到非线性的扩展. 下面详细介绍几类代表性的非线性 Granger 因果模型.

Ancona 等^[30]提出了基于径向基函数 (Radial basis functions, RBF) 的非线性预测模型, 用于衡量二变量之间的非线性 Granger 因果关系. 建立如下两个回归模型:

$$\begin{aligned} Y_{t+1} &= \mathbf{v} \times (\mathbf{Y}_t) + \varepsilon_{Y,t+1} \\ Y_{t+1} &= \mathbf{w}_1 \times (\mathbf{X}_t) + \mathbf{w}_2 \times (\mathbf{Y}_t) + \varepsilon_{Y|X,t+1} \end{aligned} \quad (6)$$

其中, \mathbf{v} , \mathbf{w}_1 , \mathbf{w}_2 为模型系数, $\mathbf{X}_t = [X_t, X_{t-1}, \dots, X_{t-m+1}]$ 和 $\mathbf{Y}_t = [Y_t, Y_{t-1}, \dots, Y_{t-m+1}]$ 表示时间序列 X 和 Y 的历史信息, 和为径向基函数. 通过判断模型残差的方差大小, 可以分析是否存在非线性 Granger 因果关系.

Marinazzo 等^[31]提出了基于核方法的非线性 Granger 因果模型, 在再生核 Hilbert 空间中进行线性 Granger 因果检测, 根据核函数映射实现线性到非线性的转换. 该方法的关键在于核函数的选择, 经过核函数的内积运算, 很容易实现高维变量的因果关系分析^[32]. Wu 等^[33]提出了基于核典型相关分析 (Kernel canonical correlation analysis, KCCA) 的非线性 Granger 因果模型, 同样在典型相关分析的基础上引入了核映射, 使得该方法具有处理多变量、非线性系统因果关系的能力.

Hu 等^[34]提出了一种基于 Copula 的 Granger 因果模型, 成功应用于非线性、多变量系统因果分析. 该方法是一种非参数模型方法, 基于 Granger

因果分析的基本思想, 应用 Copula 函数描述系统的条件概率分布, 实现因果关系分析. Montalto 等^[35]提出了基于神经网络 (Neural networks, NN) 的 Granger 因果模型, 该方法不需要任何先验假设条件, 直接根据神经网络模型的预测结果判断因果关系.

1.5 频域 Granger 因果模型

前面介绍的 Granger 因果分析方法均为时域模型, 然而在频域中能够更好地描述神经动力学系统^[36], 从而产生了频域 Granger 因果模型. Geweke^[23]提出了第一个频域 Granger 因果模型, 首先建立多变量 VAR 模型, 经过傅里叶变换将时域模型转换为频域模型, 进而分析因果关系. Barrett 等^[39]在前面模型的基础上, 引入了线性变化, 得到了简化的频域 Granger 因果模型.

Baccalá 等^[36]提出了另一种频域因果模型—偏定向相干性 (Partial directed coherence, PDC), 该方法将包含 K 个变量的 VAR 模型系数进行傅里叶变换, 定义时间序列 $X_j \rightarrow X_i$ 的因果关系为

$$PDC_{X_j \rightarrow X_i | Z}(f) = \frac{\bar{A}_{i,j}(f)}{\sqrt{\sum_{k=1}^K |\bar{A}_{k,j}(f)|^2}} \quad (7)$$

其中, Z 表示条件变量, $\mathbf{A}(f)$ 为 VAR 模型系数的傅里叶变换, $\bar{A}_{i,j}(f)$ 为矩阵 $\bar{\mathbf{A}}(f) = \mathbf{I} - \mathbf{A}(f)$ 的对应元素. $PDC_{X_j \rightarrow X_i | Z}(f)$ 表示在频率 f 下 $X_j \rightarrow X_i$ 的因果关系, 结果归一化到 $[0, 1]$ 之间, 其值接近于 0 表示无因果关系, 大于一定的阈值表明有因果关系.

Kamiński 等^[37]提出了直接传递函数 (Directed transfer function, DTF) 方法, 同样在频域分析因果关系. 与 PDC 模型类似, DTF 对建立的 VAR 模型系数进行傅里叶变换, 定义 $\mathbf{H}(f) = \mathbf{A}^{-1}(f)$ 为传递系数矩阵, 则时间序列 $X_j \rightarrow X_i$ 的因果关系为

$$DTF_{X_j \rightarrow X_i | Z}(f) = \frac{|H_{i,j}(f)|^2}{\sum_{k=1}^K |H_{k,j}(f)|^2} \quad (8)$$

其中, $H_{i,j}(f)$ 为矩阵 $\mathbf{H}(f)$ 的对应元素. DTF 描述在频率 f 下时间序列 $X_j \rightarrow X_i$ 的直接因果关系.

2 基于信息理论的因果分析

信息理论能够度量任意类型的相关关系, 是分析两个系统或多个系统之间信息流的重要手段. 在转移熵概念提出之后, 出现了各种类型的基于信息

理论的因果模型. 下面首先介绍信息理论的基本概念, 然后总结基于转移熵、条件熵和条件互信息的因果模型.

2.1 信息理论基础

熵表示一个系统混乱的程度, 系统混乱程度越高, 其熵值越大. 在信息理论中, 熵通常也称作信息熵或香农熵, 它以数值形式表达随机变量取值的不确定性程度, 目的是刻画信息量的多少. 假设 X 为一个离散的随机变量, $p(x)$ 表示 X 的概率密度函数, 则变量 X 的信息熵定义为

$$H(X) = - \sum_x p(x) \log p(x) \quad (9)$$

联合熵主要用来考察多个变量间共同拥有信息的含量, 其定义与信息熵类似. 假设 X 和 Y 为两个离散的随机变量, 其联合概率密度函数为 $p(x, y)$, 则 X 和 Y 的联合熵定义为

$$H(X, Y) = - \sum_{x, y} p(x, y) \log p(x, y) \quad (10)$$

条件熵 (Conditional entropy, CE) 表示已知一个变量的情况下, 衡量变量包含的信息量大小. 引入条件概率函数 $p(x|y)$, 条件熵具体定义为

$$H(X|Y) = - \sum_{x, y} p(x, y) \log p(x|y) \quad (11)$$

互信息反映两个变量之间的统计依赖程度, 表示两个变量共同拥有的信息量大小. 变量 X 和 Y 之间的互信息定义为

$$I(X; Y) = \sum_{x, y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (12)$$

由定义可知, 当变量 X 和 Y 完全无关或相互独立时, 互信息值为 0, 表明变量之间不存在相同的信息; 反之, 当它们相互依赖程度越高时, 互信息值越大, 所包含的相同信息也越多. 根据熵和互信息的表达式, 可以得到互信息与熵的关系为

$$\begin{aligned} I(X; Y) &= H(X) + H(Y) - H(X, Y) = \\ &= H(X) - H(X|Y) = \\ &= H(Y) - H(Y|X) \end{aligned} \quad (13)$$

基于信息理论的基本概念, 学者提出了一系列因果分析模型, 主要包括转移熵、条件熵、条件互信息等, 具体如表 2 所示. 转移熵^[40] 最初用于评价二变量时间序列之间的非线性因果关系, 是一种非参数模型方法, 受到广泛关注. 在此之后, 学者们提出了大量改进模型, 如偏转移熵、条件熵、条件互信息等. 此外, 部分时间序列呈现非平稳特性, 因此有学者致力于研究非平稳时间序列的因果关系分析, 提出了符号转移熵、偏符号转移熵等因果分析模型.

2.2 转移熵及其改进方法

2000 年, Schreiber^[40] 提出了转移熵 (Transfer entropy, TE) 的概念, 首次根据信息转移来判断变量之间的因果关系. 转移熵建立在信息理论的基本框架下, 是一种非参数模型方法, 能够很好地分析两个系统的耦合强度和不对称驱动响应关系. 考虑时间序列 X 与时间序列 Y , 转移熵定义为

$$\text{TE}_{X \rightarrow Y} = \sum_{y_{t+1}, \mathbf{x}_t, \mathbf{y}_t} p(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) \log \frac{p(y_{t+1} | \mathbf{x}_t, \mathbf{y}_t)}{p(y_{t+1} | \mathbf{y}_t)} \quad (14)$$

其中, \mathbf{X}_t 和 \mathbf{Y}_t 分别表示时间序列 X 与 Y 的历史观测值, $p(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t)$ 、 $p(y_{t+1} | \mathbf{x}_t, \mathbf{y}_t)$ 和 $p(y_{t+1} | \mathbf{y}_t)$ 分别表示联合概率密度函数和条件概率密度函数. 根据转移熵的大小判断因果关系的强弱, 当 $\text{TE}_{X \rightarrow Y} > 0$ 时, 存在由时间序列 X 到 Y 的因果关系, 数值越大表明因果关系越强. Barnett 等^[41] 证明了 Granger 因果分析与转移熵之间的联系, 在变量服从高斯分布的假设条件下, Granger 因果分析与转移熵是等价的, 从而为 Granger 因果分析与基于信息理论的因果分析方法建立了联系.

为了检测多个变量的因果关系, 考虑到中间变量的影响, 提出了多变量转移熵, 也称为偏转移熵 (Partial transfer entropy, PTE). 考虑时间序列 X 、

表 2 基于信息理论的因果关系分析方法
Table 2 Causality analysis methods based on information theory

类别	研究者	发表年份	方法名称	文献
转移熵	Schreiber	2000	转移熵 (TE)	[40]
	Staniek 等	2008	符号转移熵 (STE)	[42]
	Kugiumtzis	2013	偏符号转移熵 (PSTE)	[43]
条件熵	Faes 等	2011	条件熵 (CE)	[44]
条件互信息	Frenzel 等	2007	偏互信息 (PMI)	[45]
	Kugiumtzis	2013	基于混合嵌入的偏互信息 (PMIME)	[46]

Y 以及其他变量 \mathbf{Z} , 在给定时间序列 \mathbf{Z} 的条件下, $X \rightarrow Y$ 的转移熵定义为

$$\text{PTE}_{X \rightarrow Y | \mathbf{Z}} = \sum_{y_{t+1}, \mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t} p(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) \times \log \frac{p(y_{t+1} | \mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t)}{p(y_{t+1} | \mathbf{y}_t, \mathbf{z}_t)} \quad (15)$$

偏转移熵是二变量转移熵的扩展, 能够评价多个相互耦合系统中任意两个变量之间的信息转移. 根据信息熵和互信息的定义, 可以得出转移熵与信息熵、互信息之间的等价关系为

$$\text{TE}_{X \rightarrow Y} = H(Y_{t+1} | \mathbf{Y}_t) - H(Y_{t+1} | \mathbf{X}_t, \mathbf{Y}_t) = I(Y_{t+1}; \mathbf{X}_t | \mathbf{Y}_t) \quad (16)$$

$$\text{PTE}_{X \rightarrow Y | \mathbf{Z}} = H(Y_{t+1} | \mathbf{Y}_t, \mathbf{Z}_t) - H(Y_{t+1} | \mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t) = I(Y_{t+1}; \mathbf{X}_t | \mathbf{Y}_t, \mathbf{Z}_t) \quad (17)$$

可以看出, 转移熵可以表示为条件熵和条件互信息的形式. 上述关系式通常用于转移熵的计算.

Staniek 等^[42] 提出了符号转移熵 (Symbolic transfer entropy, STE), 将输入变量转化为秩向量, 可以应用于非平稳时间序列的因果关系分析, 表达式为

$$\text{STE}_{X \rightarrow Y} = H(\hat{\mathbf{Y}}_{t+1} | \hat{\mathbf{Y}}_t) - H(\hat{\mathbf{Y}}_{t+1} | \hat{\mathbf{X}}_t, \hat{\mathbf{Y}}_t) \quad (18)$$

其中, $\hat{\mathbf{X}}$ 和 $\hat{\mathbf{Y}}$ 表示排序后的秩向量. 为了分析多变量系统, Kugiumtzis^[43] 提出了偏符号转移熵 (Partial symbolic transfer entropy, PSTE), 其形式为

$$\text{PSTE}_{X \rightarrow Y} = H(\hat{\mathbf{Y}}_{t+1} | \hat{\mathbf{Y}}_t, \hat{\mathbf{Z}}_t) - H(\hat{\mathbf{Y}}_{t+1} | \hat{\mathbf{X}}_t, \hat{\mathbf{Y}}_t, \hat{\mathbf{Z}}_t) \quad (19)$$

2.3 用于因果分析的其他信息测度

在转移熵提出之后, 出现了一系列用于识别非线性时间序列因果关系的信息测度, 例如条件熵、条件互信息等, 其基本思想与转移熵理论一致.

Faes 等^[44] 提出了基于条件熵的因果关系分析方法, 用于检测多变量因果关系

$$C_{X \rightarrow Y | \mathbf{Z}} = 1 - \frac{H(Y_{t+1} | \mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t)}{H(Y_{t+1} | \mathbf{Y}_t, \mathbf{Z}_t)} \quad (20)$$

该方法可以看作偏转移熵的归一化形式.

Frenzel 等^[45] 提出了偏互信息 (Partial mutual information, PMI) 的概念, 即采用条件互信息检测因果关系. 在互信息的基础上, 引入了条件变量, 具体表达式为

$$I(X; Y | Z) = H(X, Z) + H(Y, Z) - H(Z) - H(X, Y, Z) \quad (21)$$

该方法能够建立三变量系统的因果关系图. 可以证明, 偏互信息与偏转移熵是等价的.

Kugiumtzis^[46] 应用条件互信息检测多个变量的因果关系, 提出了基于混合嵌入的偏互信息准则 (Partial mutual information from mixed embedding, PMIME), 表达式为

$$R_{X \rightarrow Y | \mathbf{Z}} = \frac{I(Y_{t+1}; \mathbf{X}_t | \mathbf{Y}_t, \mathbf{Z}_t)}{I(Y_{t+1}; \mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t)} \quad (22)$$

其中, $R_{X \rightarrow Y | \mathbf{Z}}$ 表示变量 $X \rightarrow Y$ 的因果关系, \mathbf{Z} 表示条件变量. 该方法能够检测变量之间的直接因果关系, 效果优于偏转移熵.

在实际应用中, 由于转移熵、条件熵和条件互信息需要计算概率密度函数, 随着输入变量维度的增加, 高维概率密度函数计算困难, 限制了其应用范围. 因此, 选择条件变量十分关键, 根据状态空间重构理论, 学者提出应用非均匀嵌入方法^[47], 可以有效处理高维变量. Runge 等^[48] 提出了基于图模型的方法, 将偏转移熵分解为多个有限维转移熵的组合, 从而避免维数灾难问题.

3 基于状态空间的时间序列分析

建立系统的数学模型是分析系统的基础. 现代控制理论引入了状态空间的概念, 通过建立状态空间模型, 能够很好地描述系统内部变量、输入变量和输出变量之间的关系, 并解释系统的运动规律. 根据观测到的时间序列建立系统的状态空间模型, 是分析结构和参数未知系统的重要手段, 同时揭示系统内部的驱动-响应关系. 然而, 对于未知结构的非线性系统, 状态空间模型的建立十分困难. 为了有效分析非线性系统的动力学特性, Takens 提出了基于时间序列的状态空间重构理论^[49], 为分析非线性系统的运行机制提供了理论支撑. 本节首先介绍时间序列的状态空间模型, 然后总结基于状态空间重构理论的因果模型.

3.1 状态空间模型

状态空间模型是描述系统动态过程的有力工具, 为时间序列分析提供了理论基础. 状态空间模型最早由 Kalman^[50] 提出, 是一种通过观测值研究确定性和随机动态系统的重要手段. 状态空间模型将物理系统表示为由输入变量、输出变量和状态变量构成的一阶微分 (或差分) 方程组, 一般由状态方程和输出方程组成

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \\ \mathbf{y}(t) &= \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t), t) \end{aligned} \quad (23)$$

其中, $\mathbf{u}(t)$ 为输入变量, $\mathbf{y}(t)$ 为输出变量, $\mathbf{x}(t)$ 为状

态变量, $f(\cdot)$ 和 $h(\cdot)$ 为线性或非线性函数. 状态空间模型是一类线性或非线性的时域模型, 用状态方程描述动态系统, 用输出方程描述量测信息. 系统的状态方程描述系统内部结构和信号的作用方向, 即反映了系统状态变量的因果关系. 建立状态空间模型主要有分析和辨识两种方式. 分析方式适用于结构和参数已知的系统, 基于物理或化学机理直接建立状态空间模型. 针对结构和参数未知的系统, 一般采用辨识方式, 即通过实际观测的输入输出数据建立状态空间模型. 状态空间模型参数估计方法主要有 Kalman 滤波、贝叶斯推理、EM 算法等.

状态空间模型利用状态变量表示一个时间序列, 状态变量包含与预测值相关的所有历史信息, 从而建立了多元时间序列模型^[51]. 状态空间模型是一类应用十分广泛的模型, 任何时间序列模型都可以写成状态空间的形式, 如自回归模型、滑动平均模型等, 在时间序列建模和因果分析方向得到了广泛应用. Jinno 等^[52]建立了非线性状态空间模型, 采用二阶泰勒展开式近似非线性系统, 根据扩展 Kalman 滤波算法更新模型参数, 实现非线性系统辨识与时间序列预测. Hong 等^[53]针对中长期径流量时间序列预测, 选择状态空间模型的结构为基于二阶泰勒展开式的非线性微分方程组, 并利用遗传算法更新状态空间模型参数. 可以看出, 状态空间模型能够识别线性或低阶非线性系统的内部结构, 从而推断系统的因果关系, 实现对未来信息的预测.

3.2 基于状态空间的因果模型

建立时间序列状态空间模型, 需要预先假设模型结构, 如一阶线性微分方程、泰勒展开式近似的非线性微分方程等, 然后根据输入输出数据辨识模型参数. 然而, 很多动力学系统具有很强的非线性, 甚至表现出混沌特性, 难以建立准确的状态空间模型^[54]. 20 世纪 80 年代, Takens 提出的延迟坐标状态空间重构方法^[49]解决了此类问题, 重构系统可以在高维状态空间中恢复原系统的动力学特性, 并与原系统保持微分同胚, 为深入研究时间序列和非线性系统奠定了理论基础. 在状态空间重构理论的基础上, 学者提出了一系列因果分析模型, 下面分别介绍状态空间重构理论和两类因果分析模型.

3.2.1 Takens 状态空间重构理论

状态空间重构的目的是在高维状态空间中恢复混沌吸引子, 是分析混沌动力学系统的第一步. 对于混沌系统, 系统中任一分量的演化过程均由与其相互作用的其他分量所决定, 产生的时间序列包含了系统全部变量的运动信息. 因此, 通过研究观测

到的时间序列, 将某些固定时间的延迟点观测值构成新的坐标, 重构出与原系统等价的状态空间, 就能够恢复出原动力学系统的运行规律.

Takens 定理^[49]指出找到状态空间嵌入维数的下界, 即延迟坐标维数 $m \geq 2D + 1$ (D 为系统的维数), 就能够在状态空间中恢复吸引子的动力学特性, 重构出的状态空间与原系统保持微分同胚. Takens 定理为混沌系统的分析与预测提供了坚实的理论依据, 问题的关键在于如何选取合适的延迟时间 τ 和嵌入维数 m . 根据嵌入方式的不同, 状态空间重构可分为均匀嵌入和非均匀嵌入两种方法.

均匀嵌入指延迟时间 τ 设定为固定数值, 考虑时间序列 $X(t)$, 根据 Takens 定理建立重构的状态空间为

$$\mathbf{X}(t) = [X(t), X(t - \tau), \dots, X(t - (m - 1)\tau)] \quad (24)$$

对于延迟时间 τ 和嵌入维数 m 的选取, 主要有两种观点. 一种思想是分别选取延迟时间 τ 和嵌入维数 m . 对于延迟时间的选择, 主要有自相关、互信息等方法. 对于嵌入维数的选择, 主要有伪最近邻^[55]、饱和关联维数、Cao 方法^[56]、最小描述长度^[57]等方法. 另一种思想认为延迟时间 τ 和嵌入维数 m 是相关的, 同时选择两个参数. Kugiumtzis^[58]提出了嵌入窗口的概念, 即 τ 和 m 由时间窗口 $\tau_w = (m - 1)\tau$ 决定. Kim 等^[59]提出的 C-C 方法常用于求解时间窗口和延迟时间.

非均匀嵌入指延迟时间 τ 选择不同的数值, 时间序列 $X(t)$ 的状态空间为

$$\mathbf{X}(t) = [X(t - \tau_1), X(t - \tau_2), \dots, X(t - \tau_m)] \quad (25)$$

其中, 嵌入维数为 m , 延迟时间为 $[\tau_1, \tau_2, \dots, \tau_m]$. 非均匀嵌入的状态空间能够完全描述原系统的动力学特性, 同时保持嵌入变量之间相互独立, 即具有低的冗余性. 相比于均匀嵌入, 非均匀嵌入能够用更精简的状态变量描述原系统, 适合解决多变量系统的重构问题. Vlachos 等^[47]提出了基于联合互信息的非均匀嵌入方法, Faes 等^[44]提出了基于条件熵的非均匀嵌入方法. 此外, 为了寻找最优的状态空间, 合理的搜索策略是十分必要的, 常用方法有顺序前向选择、遗传算法、蚁群优化算法^[60]等.

3.2.2 非线性相互依赖指标

非线性相互依赖指标 (Nonlinear interdependence measures) 是基于状态空间重构和近邻距离的方法, 用于判定因果关系的方向和大小. 对于两个系统 X 和 Y , 根据状态空间重构理论建立两个系统

的状态空间.

对于状态空间 \mathbf{X} 中的样本点 $\mathbf{x}_n, \mathbf{x}_{r_{n,1}}, \mathbf{x}_{r_{n,2}}, \dots, \mathbf{x}_{r_{n,k}}$ 表示 \mathbf{x}_n 在状态空间 \mathbf{X} 中的 k 个近邻点, 计算 \mathbf{x}_n 与 k 个近邻点的欧氏距离平均值

$$R_n^{(k)}(\mathbf{X}) = \frac{1}{k} \sum_{j=1}^k \|\mathbf{x}_n - \mathbf{x}_{r_{n,j}}\|^2 \quad (26)$$

对于状态空间 \mathbf{Y} 中的样本点 $\mathbf{y}_n, \mathbf{y}_{s_{n,1}}, \mathbf{y}_{s_{n,2}}, \dots, \mathbf{y}_{s_{n,k}}$ 表示 \mathbf{y}_n 在状态空间 \mathbf{Y} 中的 k 个近邻点, 将其映射到状态空间 \mathbf{X} 中, 计算 \mathbf{x}_n 与 k 个近邻点 $\mathbf{x}_{s_{n,1}}, \mathbf{x}_{s_{n,2}}, \dots, \mathbf{x}_{s_{n,k}}$ 的欧氏距离平均值

$$R_n^{(k)}(\mathbf{X}|\mathbf{Y}) = \frac{1}{k} \sum_{j=1}^k \|\mathbf{x}_n - \mathbf{x}_{s_{n,j}}\|^2 \quad (27)$$

为了简化计算, 可以采用 \mathbf{x}_n 与全部 N 个样本点的平均距离

$$R_n(\mathbf{X}) = \frac{1}{N-1} \sum_{j=1}^N \|\mathbf{x}_n - \mathbf{x}_j\|^2 \quad (28)$$

非线性相互依赖指标为状态空间方法, 根据状态空间的映射关系判断系统的因果关系. Arnhold 等^[6] 首先提出了指标 S , 定义为

$$S_{\mathbf{X} \rightarrow \mathbf{Y}} = \frac{1}{N} \sum_{n=1}^N \frac{R_n^{(k)}(\mathbf{X})}{R_n^{(k)}(\mathbf{X}|\mathbf{Y})} \quad (29)$$

根据定义可以得出, $0 < S_{\mathbf{X} \rightarrow \mathbf{Y}} \leq 1$. 当 $S_{\mathbf{X} \rightarrow \mathbf{Y}}$ 趋近于 0 时, 系统 \mathbf{X} 和 \mathbf{Y} 相互独立; 当 $S_{\mathbf{X} \rightarrow \mathbf{Y}}$ 明显大于 0 时, 存在由系统 \mathbf{X} 到 \mathbf{Y} 的因果关系. 同时, 指标 S 是非对称的, 可以通过判定 $S_{\mathbf{X} \rightarrow \mathbf{Y}}$ 和 $S_{\mathbf{Y} \rightarrow \mathbf{X}}$ 的大小, 分析出系统之间因果关系的强弱. 该方法具有较强的鲁棒性, 并且能够检测出弱的因果关系.

Arnhold 等^[6] 提出了另一个指标 H , 定义为

$$H_{\mathbf{X} \rightarrow \mathbf{Y}} = \frac{1}{N} \sum_{n=1}^N \ln \frac{R_n(\mathbf{X})}{R_n^{(k)}(\mathbf{X}|\mathbf{Y})} \quad (30)$$

可以看出, $H_{\mathbf{X} \rightarrow \mathbf{Y}}$ 没有上界. 当 $H_{\mathbf{X} \rightarrow \mathbf{Y}} = 0$ 时, 系统 \mathbf{X} 和 \mathbf{Y} 完全独立; 当 $H_{\mathbf{X} \rightarrow \mathbf{Y}} > 0$ 时, 存在由系统 \mathbf{X} 到 \mathbf{Y} 的因果关系. 该方法同样具有较强的鲁棒性, 对弱因果关系的灵敏度更高.

Quiroga 等^[62] 提出了第三个指标 N , 定义为

$$N_{\mathbf{X} \rightarrow \mathbf{Y}} = \frac{1}{N} \sum_{n=1}^N \frac{R_n(\mathbf{X}) - R_n^{(k)}(\mathbf{X}|\mathbf{Y})}{R_n(\mathbf{X})} \quad (31)$$

与指标 H 相比, 指标 N 采用了算术平均和标准化操作. $N_{\mathbf{X} \rightarrow \mathbf{Y}} = 1$ 的充要条件是 $R_n^{(k)}(\mathbf{X}|\mathbf{Y}) = 0$. 由于 $R_n^{(k)}(\mathbf{X}|\mathbf{Y}) \geq R_n^{(k)}(\mathbf{Y})$, 并且只有周期性系统满足 $R_n^{(k)}(\mathbf{Y}) = 0$, 所以对于一般系统 $N_{\mathbf{X} \rightarrow \mathbf{Y}} < 1$. 指

标 N 的大小主要由 $R_n^{(k)}(\mathbf{X}|\mathbf{Y})$ 决定, 受自相关性和系统有限维数影响较大. Andrzejak 等^[63] 提出了指标 M , 改进了指标 N 的不足, 具体定义为

$$M_{\mathbf{X} \rightarrow \mathbf{Y}} = \frac{1}{N} \sum_{n=1}^N \frac{R_n(\mathbf{X}) - R_n^{(k)}(\mathbf{X}|\mathbf{Y})}{R_n(\mathbf{X}) - R_n^{(k)}(\mathbf{X})} \quad (32)$$

为了避免负值的产生, 将负数替换为 0, 即 $M_{\mathbf{X} \rightarrow \mathbf{Y}}^* = \max\{M_{\mathbf{X} \rightarrow \mathbf{Y}}, 0\}$. 因此, 指标 M 限制在 $[0, 1]$, 结果为 0 表示系统 \mathbf{X} 和 \mathbf{Y} 相互独立, 数值越大因果关系越强.

Chicharro 等^[64] 提出了指标 L , 该方法采用秩统计量代替距离统计量. 对于样本 \mathbf{x}_n , 令 $g_{n,j}$ 表示距离 $\|\mathbf{x}_n - \mathbf{x}_j\|$ 在所有距离 ($j = 1, 2, \dots, N, j \neq n$) 中的排序. 因此, 状态空间 \mathbf{Y} 中样本点 \mathbf{y}_n 的 k 个近邻点映射到状态空间 \mathbf{X} 时, 计算平均秩统计量为

$$G_n^{(k)}(\mathbf{X}|\mathbf{Y}) = \frac{1}{k} \sum_{j=1}^k g_{n,s_{n,j}}$$

$$L_{\mathbf{X} \rightarrow \mathbf{Y}} = \frac{1}{N} \sum_{n=1}^N \frac{G_n(\mathbf{X}) - G_n^{(k)}(\mathbf{X}|\mathbf{Y})}{G_n(\mathbf{X}) - G_n^{(k)}(\mathbf{X})} \quad (33)$$

其中, $G_n(\mathbf{X}) = n/2$ 和 $G_n^{(k)}(\mathbf{X}) = (k+1)/2$ 分别表示全部样本和 k 个近邻点的平均秩统计量. 与指标 M 类似, 指标 L 同样限制在 $[0, 1]$. 相比于基于距离统计量的指标, 基于秩统计量的指标对定向耦合关系具有更强的敏感性与特异性.

3.2.3 收敛交叉映射

2012 年, Sugihara 等^[65] 提出了收敛交叉映射 (Convergent cross mapping, CCM) 方法, 论文发表于 *Science* 上, 引起国内外学者的广泛关注. 该方法建立在非线性状态空间重构的基础上, 分析两个系统之间的非线性因果关系, 其基本思想是: 如果系统 \mathbf{Y} 对系统 \mathbf{X} 有因果关系, 则认为系统 \mathbf{X} 中包含系统 \mathbf{Y} 的演化信息, 通过分析系统 \mathbf{X} 和 \mathbf{Y} 重构流形之间的相关性, 进而检测出系统之间的因果关系.

假设 $X(t)$ 和 $Y(t)$ 分别为系统 \mathbf{M} 投影于一维空间产生的两个时间序列. 对于时间序列 $X(t)$ 和 $Y(t)$, 设重构流形的嵌入维数为 m , 延迟时间为 τ , 重构出的状态空间为

$$\mathbf{X}(t) = [X(t), X(t-\tau), \dots, X(t-(m-1)\tau)] \quad (34)$$

$$\mathbf{Y}(t) = [Y(t), Y(t-\tau), \dots, Y(t-(m-1)\tau)] \quad (35)$$

根据状态空间重构理论^[49], 重构流形 \mathbf{X} 、 \mathbf{Y} 与系统 \mathbf{M} 是微分同胚的. 在系统 \mathbf{X} 中寻找样本 $\mathbf{X}(i)$ 的 m 个近邻点 $\{\mathbf{X}(i, k)\} = \{\mathbf{X}(i, 1), \mathbf{X}(i, 2), \dots, \mathbf{X}(i, m)\}$, 将其映射到流形 \mathbf{Y} 中, 对应的样本点为 $\{\mathbf{Y}(i, k)\}$, 计算 $Y(i)$ 的估计值

$$\hat{Y}(i) = \sum_{k=1}^m w_k Y(i, k) \quad (36)$$

其中

$$u_k = \exp\left(-\frac{\|\mathbf{X}(i) - \mathbf{X}(i, k)\|}{\|\mathbf{X}(i) - \mathbf{X}(i, 1)\|}\right)$$

$$w_k = \frac{u_k}{\sum_{k=1}^m u_k}$$

$\|\cdot\|$ 表示样本之间的欧氏距离. 定义 $\hat{Y}(t)$ 为 $Y(t)$ 从流形 \mathbf{X} 到 \mathbf{Y} 的交叉映射, 计算 $\hat{Y}(t)$ 与 $Y(t)$ 的相关系数, 相关系数计算公式为

$$r = \frac{\sum_{i=1}^L (Y(i) - \bar{Y}(i)) (\hat{Y}(i) - \bar{\hat{Y}}(i))}{\sqrt{\sum_{i=1}^L (Y(i) - \bar{Y}(i))^2 \sum_{i=1}^L (\hat{Y}(i) - \bar{\hat{Y}}(i))^2}} \quad (37)$$

随着样本长度 L 增加, $\hat{Y}(t)$ 逐渐收敛于 $Y(t)$, 最终相关系数收敛到 $[0, 1]$, 表明存在由系统 \mathbf{Y} 到系统 \mathbf{X} 的因果关系.

图 1 给出了收敛交叉映射的基本原理示意图, 建立了两个系统 \mathbf{X} 和 \mathbf{Y} 之间的交叉映射. 如图 1(a) 所示, 流形 \mathbf{X} 中的样本点 $\mathbf{X}(i)$ 及其邻近点映射到流形 \mathbf{Y} 中, 对应邻近点收敛于样本点 $\mathbf{Y}(i)$, 则表明存在由系统 \mathbf{Y} 到系统 \mathbf{X} 的因果关系; 如图 1(b) 所示, 经过交叉映射, 邻近点呈现发散现象, 则不存在

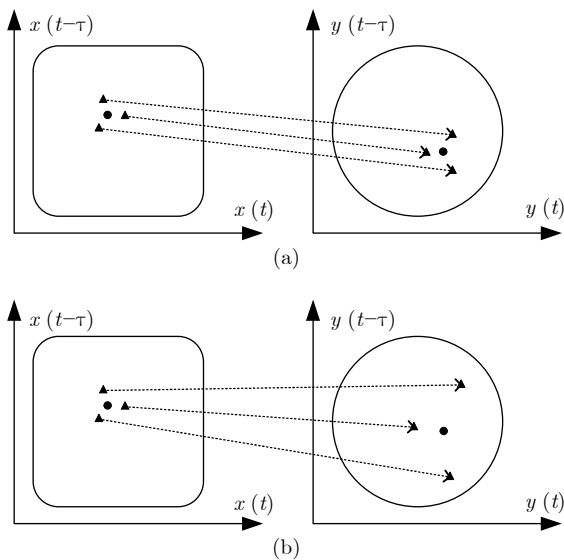


图 1 收敛交叉映射基本原理示意图

Fig. 1 Schematic diagram of the basic principle of convergence cross mapping

由系统 \mathbf{Y} 到系统 \mathbf{X} 的因果关系.

4 时间序列因果模型的对比及应用

本文的研究对象为多元时间序列, 前面详细介绍了三类时间序列因果分析模型, 分别为 Granger 因果关系分析、基于信息理论的因果分析和基于状态空间的因果分析. 三类方法分别从不同的角度提出, 本节将总结每类方法的优势、不足以及适用范围, 并讨论存在的问题和未来发展方向. 然后, 针对不同的应用领域, 分别介绍时间序列因果分析模型的典型应用, 并进一步明确模型的适用范围.

4.1 因果分析模型对比与发展方向

针对不同维数、特性的时间序列, Granger 因果分析模型、基于信息理论的因果模型和基于状态空间的因果模型具有特定的适用范围, 例如传统的 Granger 因果模型只能分析二变量线性因果关系, 转移熵可以分析二变量非线性因果关系, 收敛交叉映射能够实现二变量非线性时间序列的因果分析. 下面详细分析三类因果模型的适用范围、存在的问题以及发展方向, 表 3 给出了三类方法对非线性、多变量和非平稳时间序列因果分析的应用情况.

从表 3 可以看出, 随着研究的深入, 时间序列因果模型的适用范围由二变量、线性因果向多变量、非线性因果发展. 此外, 少数因果模型突破了对时间序列平稳性的要求, 能够实现对非平稳时间序列的因果分析. 下面, 分别对三类方法进行具体对比分析.

1) Granger 因果关系分析方法应用十分广泛, 其建立在时间序列模型的基础上, 形式简单且具有很强的可解释性. 然而, Granger 因果分析方法是一种定性的因果分析模型, 不能根据检验结果直接判断因果关系的强弱. Granger 因果分析是基于模型的方法, 因此对于时间序列模型阶数的确定是一个至关重要的问题, 常用的确定模型阶数的方法有 AIC (Akaike information criterion)、BIC (Bayesian information criterion) 等信息准则. 由于时间序列模型的参数较多, 因此在进行 Granger 因果检验时, 计算复杂度较高. 例如, 对于具有 n 个样本的时间序列 X 和 Y , 建立一个模型阶数为 m 的 VAR 模型的计算复杂度介于 $O(m^2n^2)$ 到 $O(mn)$ 之间. 对 l 个时间序列进行两两因果分析, 计算复杂度将达到 $O(l^2m^2n^2)$, Lasso-Granger 因果模型可以将计算复杂度降为 $O(lm^2n^2)$ ^[26]. 此外, Granger 因果模型的应用对象是平稳时间序列, 因此在进行 Granger 因果检验之前, 需要对时间序列进行平稳性检

表 3 因果分析方法应用范围比较
Table 3 Comparison of application range of causality analysis methods

研究者	方法名称	非线性	多变量	非平稳	文献
Granger	Granger 因果指数				[15]
Geweke	条件 Granger 因果指数		√		[23]
Chen 等	条件扩展 Granger 因果指数	√	√		[24]
Sigiridou 等	限制条件 Granger 因果指数	√	√		[25]
Arnold 等	Lasso-Granger 因果模型		√		[26]
Shojaie 等	截断 Lasso-Granger 因果模型		√		[27]
Bolstad 等	Grouped-Lasso-Granger 因果模型		√		[28]
Yang 等	Grouped-Lasso 非线性条件 Granger 因果模型	√	√		[29]
Ancona 等	RBF-Granger 因果模型	√			[30]
Marinazzo 等	Kernel-Granger 因果模型	√	√		[31–32]
Wu 等	KCCA-Granger 因果模型	√	√		[33]
Hu 等	Copula-Granger 因果模型	√	√		[34]
Montalto 等	NN-Granger 因果模型	√	√		[35]
Geweke	Spectral-Granger 因果模型		√		[23]
Baccalá 等	偏定向相干性		√		[36]
Kamiński 等	直接传递函数		√		[37]
Schreiber	转移熵	√			[40]
Staniek 等	符号转移熵	√		√	[42]
Kugiumtzis	偏符号转移熵	√	√	√	[43]
Faes 等	条件熵	√	√		[44]
Frenzel 等	偏互信息	√	√		[45]
Kugiumtzis	基于混合嵌入的偏互信息	√	√		[46]
Arnhold 等	非线性相互依赖指标 S 和 H	√			[61]
Quiroga 等	非线性相互依赖指标 N	√			[62]
Andrzejak 等	非线性相互依赖指标 M	√			[63]
Chicharro 等	非线性相互依赖指标 L	√		√	[64]
Sugihara 等	收敛交叉映射	√			[65]

验和平稳化. 学者提出了非平稳时间序列的因果分析模型^[66], 在 VAR 模型中引入时变参数, 实现了非平稳时间序列的因果关系分析. 随着时间序列数据维度和规模的不断增加, 挖掘复杂环境下时间序列存在的因果关系, 是当前面临的重要挑战. 因此, Granger 因果模型未来将着重解决非线性、多变量、非平稳等复杂环境的因果分析, 同时提高模型的计算效率.

2) 基于信息理论的因果模型是一类定量分析方法, 通过建立评价函数实现对时间序列因果关系的定量描述. 该方法只需要计算信息指标就可以得出因果分析结果, 对于低维复杂系统, 其分析结果明显优于只能定性分析的 Granger 因果模型. 此类方法也是建立在平稳时间序列的基础之上, 对于非平稳时间序列, 有学者应用符号转移熵实现非平稳时间序列的因果分析, 它将输入变量转化为秩向量,

为非平稳时间序列的因果分析提供了指导性思想. 虽然基于信息理论的因果分析方法形式简单, 但嵌入变量的选择需要进行深入研究, 文献^[67]对基于信息理论的因果模型的参数求解做了详细分析. 在实际的应用中, 信息测度需要计算概率密度函数, 当变量维数增加时, 其计算复杂度增加, 计算精度下降. 因此, 未来我们需要重点关注条件变量的选取, 通过合理选取条件变量, 达到简化计算的目的.

3) 状态空间模型是一类基于模型的方法, 根据模型内部信号的作用方向判断因果关系, 该方法需要预先假设模型结构, 对于非线性时间序列的因果分析能力较弱. 随着状态空间重构理论的提出, 使得非线性系统的分析与建模更为便捷, 出现了基于状态空间的因果分析方法. 此类方法在非线性的因果分析问题上具有很好的效果, 特别是小规模、短期时间序列的因果关系分析^[68]. Clark 等^[69]提出了多

空间收敛交叉映射,能够分析出小数据量时间序列(长度小于10)之间存在的因果关系.基于状态空间的因果分析方法具有较强的因果识别能力,如收敛交叉映射对于弱耦合系统的因果关系的灵敏度较高,但在强耦合系统可能会产生错误结果^[18].然而,该方法建立在状态空间基础上,根据样本点的邻域信息分析因果关系,受到噪声点的影响较大^[70],严重影响分析结果的准确性,需要提升方法的抗噪能力.此外,该方法常用于分析两个系统之间的直接因果关系,将其扩展至分析多个系统的因果关系,是未来的研究方向.

4.2 应用

多元时间序列的因果关系分析技术广泛应用于自然、医学、社会科学等领域,在日常生活中发挥着重要作用,已经成为大规模数据挖掘的重要手段.不同领域的时间序列具有不同的特性,下面具体介绍时间序列因果分析在不同领域的典型应用,并进一步讨论模型的适用范围.

在自然界中,气象、水文、环境等系统的内部和系统之间具有复杂的驱动响应关系,借助于时间序列因果分析手段,有助于揭示系统的运行规律,模拟和预测自然现象的未来发展趋势.例如,随着经济快速发展,以雾霾为代表的大气污染已经成为主要环境问题之一,我国积极参与一系列国际大气污染防治的公约和协议,并制定了具体措施和监管机制.然而,雾霾的成因十分复杂,以PM_{2.5}空气质量指数为例,其浓度不仅受到NO₂、CO、O₃、SO₂等大气污染物的影响,而且受到气温、气压、湿度、风速、风向等环境变量的影响,如果分析出PM_{2.5}的主要污染物和生成机理,能够为国家治理大气污染以及各地制定针对性的治理手段发挥重要作用.因果分析方法能够利用一定范围内的观测序列,快速识别PM_{2.5}与影响变量之间的因果关系,判断主要影响因素,从而为决策与调控提供理论依据^[71].此外,在其他实际问题中,因果分析方法同样发挥了重要作用.Liang^[72]通过建立回归模型,研究了厄尔尼诺和印度洋偶极子两个气象子系统的因果关系,确定了系统之间的非对称因果关系.Faybishenko^[73]应用条件Granger因果分析方法,分析具有非线性混沌特性的水文过程,并建立了时空分布的因果循环图.Zhu等^[74]将Granger因果分析扩展到时空空间,分析气象、交通等影响变量对空气质量的影响.Sugihara等^[65]在状态空间重构理论的基础上,提出了收敛交叉映射,分析复杂生态系统的因果关系.Chen等^[75]应用收敛交叉映射方法,分析京

津冀区域气象因子与PM_{2.5}浓度的因果关系,得出定量的分析结果.基于以上研究成果可以看出,Granger因果分析方法广泛用于自然界复杂系统的因果分析,通过建立时间序列模型,从可预测性角度评价因果关系.此外,收敛交叉映射是针对生态系统提出的因果分析模型,目前在气象、环境等自然领域取得了很好的应用效果.

在医学领域,随着信息采集和存储技术的不断发展,医学信号的分析与判别是一项十分复杂的工作,在医疗辅助决策中发挥着重要作用,因此时间序列分析方法在医学领域具有广阔的应用前景.例如,常见的脑部神经疾病癫痫,由大脑皮层神经元异常放电引起,其发病机制十分复杂.脑电信号是鉴别癫痫病是否发作的重要依据,从多通道脑电信号的功能性连接的角度,建立癫痫发作期大脑不同区域神经活动的因果关系网络,能够有效定位癫痫病灶,对癫痫发作机理的研究具有重要价值.近年来,应用时域和频域Granger因果分析方法,为癫痫病的诊断和手术治疗提供了有力保障^[76].此外,在其他医学研究中,因果分析方法同样取得了突出成果.Dhamala等^[77]应用二变量和条件Granger因果分析方法,分析猴子大脑局部场电位信号的因果关系,研究感觉运动任务.Wu等^[78]提出了一种基于扩展典型相关分析的多变量Granger因果分析方法,分析癫痫病患者的大脑皮层和深层脑电信号的网络连接.Li等^[79]提出了一种基于 L_p 范数的Granger因果分析方法,应用于包含眼电伪迹的脑电信号分析,能够有效消除干扰并还原网络结构.Hu等^[80]提出了一种基于Copula的Granger因果分析方法,应用于神经脉冲序列数据,揭示脉冲序列的非线性、高阶因果关系.Faes等^[81]应用具有特定延迟变量的转移熵指标,度量心血管和心肺系统的信息转移,并得出确定的方向、大小和时间.Wang等^[82]根据定向信息建立因果分析模型,应用于功能性磁共振成像数据分析,很好地反映出非线性因果关系.Heskamp等^[83]采用具有非线性分析能力的收敛交叉映射,定性分析大脑自动调节能力.综上所述,Granger因果分析方法广泛应用于脑电等生理学时间序列,其中频域Granger因果模型能够更好地刻画神经动力学系统,在医学领域的因果分析中占据重要位置.由于医学信号通常包含大量噪声且具有非平稳特性,基于信息理论的因果分析方法适用于解决此类问题,取得了很好的应用效果.

在社会科学领域,金融、能源等序列的分析与预测具有十分重要的战略意义,并且受到政治、经济、气候变化等多个因素的共同影响,借助于时间

序列因果关系分析技术, 进行准确的定性分析与定量预测是重要的研究内容. 例如, 我国人口众多、经济快速发展, 导致能源需求巨大, 节能减排已经成为国家战略, 是维持可持续发展的长久之计. 然而, 节能减排必须注重社会效益和环境效益的均衡, 关于能源消耗与经济增长相互影响关系的研究具有重要意义. 因果分析方法在经济学领域发挥着重要作用, 合理解释能源消耗与经济增长之间的因果关系, 能够为节能减排政策的制定提供理论指导, 在可持续发展和国民经济建设方面具有重要意义^[84]. 目前, 因果分析方法在社会科学领域取得了广泛应用. Zhou 等^[85]应用 Granger 因果分析方法, 调查中国的经济结构、能源消费结构、收入、城市化、外商直接投资和贸易总额等变量对中国二氧化碳排放量的影响. Rafindadi 等^[86]提出采用向量误差修正和 Granger 因果分析模型, 评价可再生能源消费对德国经济增长的影响. Tiwari^[87]应用频域 Granger 因果分析方法, 定性分析生产商的批发价格指数与消费者物价指数之间的相互作用, 得到二者之间的双向因果关系. Bekiros 等^[88]应用转移熵和复杂网络理论, 研究美国股票和商品期货市场之间的动态因果关系. Papan 等^[89]提出了一种基于偏转移熵的因果分析方法, 并成功应用于非平稳金融时间序列分析. 根据上述研究成果可以发现, Granger 因果分析模型最早在金融领域取得突破, 目前仍然在社会科学领域发挥着重要作用. 另外, 金融时间序列同样具有非平稳特性, 基于信息理论的因果模型在金融领域具有广阔的应用前景.

5 总结与展望

随着海量时间序列的出现, 时间序列因果关系分析已经成为当前的研究热点. 首先, 本文对多元时间序列因果分析的研究现状进行了详细的综述, 包括三类典型方法. 第 1 类方法是 Granger 因果关系分析, 从可预测性角度出发, 根据时间序列模型预测结果定性分析因果关系, 该方法易于操作且可解释性强. 第 2 类方法是基于信息理论的因果分析, 根据信息测度建立因果关系评价指标, 定量分析时间序列的因果关系, 对时间序列的类型没有特殊要求, 具有广泛的适用范围. 第 3 类方法是基于状态空间的因果分析, 根据重构状态空间的映射关系, 定量评价两个系统之间的因果关系强弱, 在非线性系统因果分析问题上具有很好的效果. 然后, 对三类方法的代表性模型进行了对比分析, 并指出每类方法适用范围、存在的主要问题和发展方向. 最后, 本文总结了因果分析方法在自然、医学、社会科学

领域的实际应用, 并进一步讨论了模型的适用范围.

根据因果分析模型的适用范围, 可以将其划分为线性和非线性、二变量和多变量等不同类型方法, 随着研究的深入, 当前因果分析方法主要面向非线性、多变量、非平稳系统. 对于今后的研究工作可以从以下几个方向展开:

1) 针对非线性因果关系分析, 可以从三个角度展开研究: a) 根据 Granger 提出的可预测性理论, 建立非线性预测模型, 如核方法、神经网络等; b) 基于非线性相关性指标建立因果关系模型, 如互信息、Copula 分析等; c) 根据非线性状态空间重构理论, 应用状态空间模型建立因果关系. 在实际应用中, 采用几种不同方法的组合, 能够实现对复杂系统的非线性因果关系分析.

2) 针对多变量因果关系分析, 可以从两个角度展开研究: a) 建立多变量回归模型, 根据 Granger 因果理论分析多变量系统因果关系; b) 引入条件变量, 从条件概率的角度建立多变量因果关系指标. 目前, 大部分研究成果集中于二维或多维变量的因果分析, 对于高维或超高维时间序列的因果分析缺少有效的处理手段. 借助于稀疏化建模等技术手段, 展开对海量数据的因果分析, 是未来的重点研究内容之一.

3) 针对非平稳时间序列的因果关系分析, 可以从以下两个方面展开研究: a) 对时间序列本身进行处理, 实现平稳化, 如差分方法、符号化等, 然后对平稳化后的时间序列进行因果关系分析; b) 建立时变的回归模型, 实现非平稳时间序列的因果关系分析, 如时变广义部分有向相干方法. 建立时变参数模型对非平稳时间序列进行因果关系分析是未来的一个研究方向.

4) 对系统的历史信息或状态空间的选择, 严重影响因果分析的结果. 对于 Granger 因果模型, 确定模型的阶数, 应用输入变量选择算法构建合适的输入变量, 能够有效提高因果分析的准确性. 对于基于状态空间的因果分析指标及模型, 应用非均匀嵌入方法建立状态空间, 有助于降低模型的复杂度, 提高因果分析的计算精度.

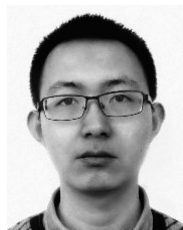
References

- Gao Yue, Su Chong, Li Hong-Guang. A kind of deep belief networks based on nonlinear features extraction with application to PM2.5 concentration prediction and diagnosis. *Acta Automatica Sinica*, 2018, 44(2): 318-329
(高月, 宿翀, 李宏光. 一类基于非线性 PCA 和深度置信网络的混合分类器及其在 PM2.5 浓度预测和影响因素诊断中的应用. *自动化学报*, 2018, 44(2): 318-329)
- Han M, Liu X X. Feature selection techniques with class separ-

- ability for multivariate time series. *Neurocomputing*, 2013, **110**: 29–34
- 3 He J Y, Shang P J. Comparison of transfer entropy methods for financial time series. *Physica A: Statistical Mechanics and Its Applications*, 2017, **482**: 772–785
 - 4 Baek S, Kim D Y. Empirical sensitivity analysis of discretization parameters for fault pattern extraction from multivariate time series data. *IEEE Transactions on Cybernetics*, 2017, **47**(5): 1198–1209
 - 5 Zhou Ping, Liu Ji-Ping. Data-driven multi-output ARMAX modeling for online estimation of central temperatures for cross temperature measuring in blast furnace ironmaking. *Acta Automatica Sinica*, 2018, **44**(3): 552–561
(周平, 刘记平. 基于数据驱动多输出 ARMAX 建模的高炉十字测温中心温度在线估计. *自动化学报*, 2018, **44**(3): 552–561)
 - 6 Fu T C. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 2011, **24**(1): 164–181
 - 7 Esling P, Agon C. Time-series data mining. *ACM Computing Surveys*, 2012, **45**(1): 12
 - 8 Liu Qiang, Qin S Joe. Perspectives on big data modeling of process industries. *Acta Automatica Sinica*, 2016, **42**(2): 161–171
(刘强, 秦泗钊. 过程工业大数据建模研究展望. *自动化学报*, 2016, **42**(2): 161–171)
 - 9 Hardoon D R, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 2004, **16**(12): 2639–2664
 - 10 Han M, Ren W J. Global mutual information-based feature selection approach using single-objective and multi-objective optimization. *Neurocomputing*, 2015, **168**: 47–54
 - 11 Reshef D N, Reshef Y A, Finucane H K, Grossman S R, McVean G, Turnbaugh P J, et al. Detecting novel associations in large data sets. *Science*, 2011, **334**(6062): 1518–1524
 - 12 Shi J, Ding Z H, Lee W J, Yang Y P, Liu Y Q, Zhang M M. Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features. *IEEE Transactions on Smart Grid*, 2014, **5**(1): 521–526
 - 13 Liebscher E. Copula-based dependence measures. *Dependence Modeling*, 2014, **2**(1): 49–64
 - 14 Sun Y Q, Li J Y, Liu J X, Chow C W, Sun B Y, Wang R J. Using causal discovery for feature selection in multivariate numerical time series. *Machine Learning*, 2015, **101**(1-3): 377–395
 - 15 Granger C W J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 1969, **37**(3): 424–438
 - 16 Barnett L, Seth A K. The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, 2014, **223**: 50–68
 - 17 Hlaváčková-Schindler K, Paluš M, Vejmelka M, Bhattacharya J. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 2007, **441**(1): 1–46
 - 18 Cummins B, Gedeon T, Spendlove K. On the efficacy of state space reconstruction methods in determining causality. *SIAM Journal on Applied Dynamical Systems*, 2015, **14**(1): 335–381
 - 19 Zou C L, Feng J F. Granger causality vs. dynamic Bayesian network inference: A comparative study. *BMC Bioinformatics*, 2009, **10**(1): 122–122
 - 20 Kleinberg S, Hripcsak G. A review of causal inference for biomedical informatics. *Journal of Biomedical Informatics*, 2011, **44**(6): 1102–1112
 - 21 Porta A, Faes L. Wiener-Granger causality in network physiology with applications to cardiovascular control and neuroscience. *Proceedings of the IEEE*, 2016, **104**(2): 282–309
 - 22 Seth A K, Barrett A B, Barnett L. Granger causality analysis in neuroscience and neuroimaging. *The Journal of Neuroscience*, 2015, **35**(8): 3293–3297
 - 23 Geweke J. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 1982, **77**(378): 304–313
 - 24 Chen Y H, Rangarajan G, Feng J F, Ding M Z. Analyzing multiple nonlinear time series with extended Granger causality. *Physics Letters A*, 2004, **324**(1): 26–35
 - 25 Siggiridou E, Kugiumtzis D. Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model. *IEEE Transactions on Signal Processing*, 2016, **64**(17): 1759–1773
 - 26 Arnold A, Liu Y, Abe N. Temporal causal modeling with graphical granger methods. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Jose, California, USA: ACM, 2007. 66–75
 - 27 Shojaie A, Michailidis G. Discovering graphical Granger causality using the truncating lasso penalty. *Bioinformatics*, 2010, **26**(18): i517–i523
 - 28 Bolstad A, Van Veen B D, Nowak R. Causal network inference via group sparse regularization. *IEEE Transactions on Signal Processing*, 2011, **59**(6): 2628–2641
 - 29 Yang G X, Wang L, Wang X F. Reconstruction of complex directional networks with group lasso nonlinear conditional Granger causality. *Scientific Reports*, 2017, **7**(1): 2991
 - 30 Ancona N, Marinazzo D, Stramaglia S. Radial basis function approach to nonlinear Granger causality of time series. *Physical Review E*, 2004, **70**(5): 056221
 - 31 Marinazzo D, Pellicoro M, Stramaglia S. Kernel method for nonlinear Granger causality. *Physical Review Letters*, 2008, **100**(14): 144103
 - 32 Marinazzo D, Pellicoro M, Stramaglia S. Kernel-Granger causality and the analysis of dynamical networks. *Physical Review E*, 2008, **77**(5): 056215
 - 33 Wu G R, Duan X J, Liao W, Gao Q, Chen H F. Kernel canonical-correlation Granger causality for multiple time series. *Physical Review E*, 2011, **83**(4): 041921
 - 34 Hu M, Liang H L. A copula approach to assessing Granger causality. *NeuroImage*, 2014, **100**: 125–134
 - 35 Montalto A, Stramaglia S, Faes L, Tessitore G, Prevede R, Marinazzo D. Neural networks with non-uniform embedding and explicit validation phase to assess Granger causality. *Neural Networks*, 2015, **71**: 159–171
 - 36 Baccalá L A, Sameshima K. Partial directed coherence: A new concept in neural structure determination. *Biological Cybernetics*, 2001, **84**(6): 463–474
 - 37 Kamiński M, Ding M Z, Truccolo W A, Bressler S L. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics*, 2001, **85**(2): 145–157

- 38 Stokes P A, Purdon P L. A study of problems encountered in Granger causality analysis from a neuroscience perspective. *Proceedings of the National Academy of Sciences*, 2017, **114**(34): E7063–E7072
- 39 Barrett A B, Barnett L, Seth A K. Multivariate Granger causality and generalized variance. *Physical Review E*, 2010, **81**(4): 041907
- 40 Schreiber T. Measuring information transfer. *Physical Review Letters*, 2000, **85**(2): 461–464
- 41 Barnett L, Barrett A B, Seth A K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Physical Review Letters*, 2009, **103**(23): 238701
- 42 Staniek M, Lehnertz K. Symbolic transfer entropy. *Physical Review Letters*, 2008, **100**(15): 158101
- 43 Kugiumtzis D. Partial transfer entropy on rank vectors. *The European Physical Journal Special Topics*, 2013, **222**(2): 401–420
- 44 Faes L, Nollo G, Porta A. Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. *Physical Review E*, 2011, **83**(5): 051112
- 45 Frenzel S, Pompe B. Partial mutual information for coupling analysis of multivariate time series. *Physical Review Letters*, 2007, **99**(20): 204101
- 46 Kugiumtzis D. Direct-coupling information measure from nonuniform embedding. *Physical Review E*, 2013, **87**(6): 062918
- 47 Vlachos I, Kugiumtzis D. Nonuniform state-space reconstruction and coupling detection. *Physical Review E*, 2010, **82**(1): 016207
- 48 Runge J, Heitzig J, Petoukhov V, Kurths J. Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Physical Review Letters*, 2012, **108**(25): 258701
- 49 Takens F. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*. Heidelberg, Germany: Springer-Verlag, 1981. 366–381
- 50 Kalman R E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 1960, **82**(1): 35–45
- 51 Solo V. State-space analysis of Granger-Geweke causality measures with application to fMRI. *Neural Computation*, 2016, **28**(5): 914–949
- 52 Jinno K, Xu S G, Berndtsson R, Kawamura A, Matsumoto M. Prediction of unspots using reconstructed chaotic system equations. *Journal of Geophysical Research: Space Physics*, 1995, **100**(A8): 14773–14781
- 53 Hong M, Wang D, Wang Y K, Zeng X K, Ge S S, Yan H Q, Singh V P. Mid-and long-term runoff predictions by an improved phase-space reconstruction model. *Environmental Research*, 2016, **148**: 560–573
- 54 Yin Li-Sheng, He Yi-Gang, Dong Xue-Ping, Lu Zhao-Quan. Research on the multi-step prediction of Volterra neural network for traffic flow. *Acta Automatica Sinica*, 2014, **40**(9): 2066–2072 (殷礼胜, 何怡刚, 董学平, 鲁照权. 交通流量 VNNTF 神经网络模型多步预测研究. 自动化学报, 2014, **40**(9): 2066–2072)
- 55 Luo S H, Gao C H, Zeng J S, Huang J. Blast furnace system modeling by multivariate phase space reconstruction and neural networks. *Asian Journal of Control*, 2013, **15**(2): 553–561
- 56 Cao L. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena*, 1997, **110**(1): 43–50
- 57 Molkov Y I, Mukhin D N, Loskutov E M, Feigin A M, Fidelin G A. Using the minimum description length principle for global reconstruction of dynamic systems from noisy time series. *Physical Review E*, 2009, **80**(4): 046207
- 58 Kugiumtzis D. State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length. *Physica D: Nonlinear Phenomena*, 1996, **95**(1): 13–28
- 59 Kim H, Eykholt R, Salas J D. Nonlinear dynamics, delay times, and embedding windows. *Physica D: Nonlinear Phenomena*, 1999, **127**(1–2): 48–60
- 60 Shen M, Chen W N, Zhang J, Chung H S H, Kaynak O. Optimal selection of parameters for nonuniform embedding of chaotic time series using ant colony optimization. *IEEE Transactions on Cybernetics*, 2013, **43**(2): 790–802
- 61 Arnhold J, Grassberger P, Lehnertz K, Elger C E. A robust method for detecting interdependences: application to intracranially recorded EEG. *Physica D: Nonlinear Phenomena*, 1999, **134**(4): 419–430
- 62 Quiroga R Q, Arnhold J, Grassberger P. Learning driver-response relationships from synchronization patterns. *Physical Review E*, 2000, **61**(5): 5142
- 63 Andrzejak R G, Kraskov A, Stögbauer H, Mormann F, Kreuz T. Bivariate surrogate techniques: Necessity, strengths, and caveats. *Physical Review E*, 2003, **68**(6): 066202
- 64 Chicharro D, Andrzejak R G. Reliable detection of directional couplings using rank statistics. *Physical Review E*, 2009, **80**(2): 026217
- 65 Sugihara G, May R, Ye H, Hsieh C H, Deyle E, Fogarty M, Munch S. Detecting causality in complex ecosystems. *Science*, 2012, **338**(6106): 496–500
- 66 Schäck T, Muma M, Feng M L, Guan C T, Zoubir A M. Robust nonlinear causality analysis of nonstationary multivariate physiological time series. *IEEE Transactions on Biomedical Engineering*, 2017, **65**(6): 1213–1225
- 67 Montalto A, Faes L, Marinazzo D. MuTE: A MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy. *PloS One*, 2014, **9**(10): e109462
- 68 Ma H F, Aihara K, Chen L N. Detecting causality from nonlinear dynamics with short-term time series. *Scientific Reports*, 2014, **4**: 7464
- 69 Clark A T, Ye H, Isbell F, Deyle E R, Cowles J, Tilman G D, Sugihara G. Spatial convergent cross mapping to detect causal relationships from short time. *Ecology*, 2015, **96**(5): 1174–1181
- 70 Mønster D, Fusaroli R, Tylén K, Roepstorff A, Sherson J F. Causal inference from noisy time-series data—testing the convergent cross-mapping algorithm in the presence of noise and external influence. *Future Generation Computer Systems*, 2017, **73**: 52–62
- 71 Zhu J Y, Zhang C, Zhang H C, Zhi S, Li V O K, Han J W, Zheng Y. pg-Causality: Identifying spatiotemporal causal pathways for air pollutants with urban big data. *IEEE Transactions on Big Data*, 2018, **4**(4): 571–585

- 72 Liang X S. Unraveling the cause-effect relation between time series. *Physical Review E*, 2014, **90**(5): 052150
- 73 Faybishenko B. Detecting dynamic causal inference in nonlinear two-phase fracture flow. *Advances in Water Resources*, 2017, **106**: 111–120
- 74 Zhu J Y, Sun C, Li V O K. An extended spatio-temporal Granger causality model for air quality estimation with heterogeneous urban big data. *IEEE Transactions on Big Data*, 2017, **3**(3): 307–319
- 75 Chen Z Y, Cai J, Gao B B, Xu B, Dai S, He B, Xie X M. Detecting the causality influence of individual meteorological factors on local PM 2.5 concentration in the Jing-Jin-Ji region. *Scientific Reports*, 2017, **7**: 40735
- 76 Hu S Q, Dai G J, Worrell G A, Dai Q H, Liang H L. Causality analysis of neural connectivity: Critical examination of existing methods and advances of new methods. *IEEE Transactions on Neural Networks*, 2011, **22**(6): 829–844
- 77 Dhamala M, Rangarajan G, Ding M Z. Analyzing information flow in brain networks with nonparametric Granger causality. *NeuroImage*, 2008, **41**(2): 354–362
- 78 Wu G R, Chen F Y, Kang D Z, Zhang X Y, Marinazzo D, Chen H F. Multiscale causal connectivity analysis by canonical correlation: Theory and application to epileptic brain. *IEEE Transactions on Biomedical Engineering*, 2011, **58**(11): 3088–3096
- 79 Li P Y, Huang X Y, Li F L, Wang X R, Zhou W W, Liu H, et al. Robust Granger analysis in Lp norm space for directed EEG network analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2017, **25**(11): 1959–1969
- 80 Hu M, Li W, Liang H L. A copula-based Granger causality measure for the analysis of neural spike train data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2018, **15**(2): 562–569
- 81 Faes L, Marinazzo D, Montalto A, Nollo G. Lag-specific transfer entropy as a tool to assess cardiovascular and cardiorespiratory information transfer. *IEEE Transactions on Biomedical Engineering*, 2014, **61**(10): 2556–2568
- 82 Wang Z, Alahmadi A, Zhu D C, Li T T. Causality analysis of fMRI data based on the directed information theory framework. *IEEE Transactions on Biomedical Engineering*, 2016, **63**(5): 1002–1015
- 83 Heskamp L, Meel-van den Abeelen A S, Lagro J, Claassen J A. Convergent cross mapping: A promising technique for cerebral autoregulation estimation. *International Journal of Clinical Neurosciences and Mental Health*, 2014, **1**(1): S20
- 84 Wang S, Li Q, Fang C, Zhou C. The relationship between economic growth, energy consumption, and CO₂ emissions: Empirical evidence from China. *Science of the Total Environment*, 2016, **542**: 360–371
- 85 Zhou C S, Wang S J, Feng K S. Examining the socioeconomic determinants of CO₂ emissions in China: A historical and prospective analysis. *Resources, Conservation and Recycling*, 2018, **130**: 1–11
- 86 Rafindadi A A, Ozturk I. Impacts of renewable energy consumption on the German economic growth: Evidence from combined cointegration test. *Renewable and Sustainable Energy Reviews*, 2017, **75**: 1130–1141
- 87 Tiwari A K. Causality between wholesale price and consumer price indices in India: An empirical investigation in the frequency domain. *Indian Growth and Development Review*, 2012, **5**(2): 151–172
- 88 Bekiros S, Nguyen D K, Junior L S, Uddin G S. Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets. *European Journal of Operational Research*, 2017, **256**(3): 945–961
- 89 Papan A, Kyrtsov C, Kugiuntzis D, Diks C. Detecting causality in non-stationary time series using partial symbolic transfer entropy: evidence in financial data. *Computational Economics*, 2016, **47**(3): 341–365



任伟杰 大连理工大学电子信息与电气工程学部博士研究生。主要研究方向为时间序列分析和特征选择。

E-mail: renweijie@mail.dlut.edu.cn
(**REN Wei-Jie** Ph.D. candidate at the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. His research interest covers time series analysis and feature selection.)



韩敏 大连理工大学电子信息与电气工程学部教授。主要研究方向为模式识别, 复杂系统建模及时间序列预测。本文通信作者。

E-mail: minhan@dlut.edu.cn
(**HAN Min** Professor at the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. Her research interest covers pattern recognition, modeling of complex system, and time series prediction. Corresponding author of this paper.)