

基于贝叶斯网络的时间序列因果关系学习*

王双成^{1,2}, 郑飞¹, 张立¹



¹(上海立信会计金融学院 信息管理学院, 上海 201620)

²(上海立信会计金融学院 数据科学交叉研究院, 上海 201209)

通讯作者: 王双成, E-mail: wangsc@lixin.edu.cn

摘要: 贝叶斯网络是研究变量之间因果关系的有力工具, 基于贝叶斯网络的因果关系学习包括结构学习与参数学习两部分, 其中结构学习是核心。目前, 贝叶斯网络主要用于发现非时间序列数据中所蕴含的因果关系(非时间序列因果关系), 从数据中学习得到的也均是一般变量之间的因果关系。针对这些情况, 本文结合时间序列预处理、时间序列变量排序、转换数据集构建和局部贪婪打分-搜索等进行时间序列的因果关系学习; 再将包括分段在内的时间序列预处理、时间序列段的因果关系结构学习、因果关系结构数据集构建、因果关系变量排序和局部贪婪打分-搜索等相结合, 来进行元因果关系(因果关系变量之间的因果关系)学习, 从而实现两个层次的时间序列因果关系学习, 为进一步的量化因果分析奠定了基础。分别使用模拟、UCI 和金融时间序列数据进行实验与分析, 实验结果显示, 基于贝叶斯网络能够有效地进行时间序列的因果关系和元因果关系学习。

关键词: 时间序列; 因果关系; 贝叶斯网络; 转换数据集; 结构数据集

中图法分类号: TP181

中文引用格式: 王双成, 郑飞, 张立. 基于贝叶斯网络的时间序列因果关系学习. 软件学报. <http://www.jos.org.cn/1000-9825/6012.htm>

英文引用格式: Wang SC, Zheng F, Zhang L. Learning causal relationship from time series based on Bayesian network. Ruan Jian Xue Bao/Journal of Software, (in Chinese). <http://www.jos.org.cn/1000-9825/6012.htm>

Learning Causal Relationship from Time Series Based on Bayesian Network

WANG Shuang-Cheng^{1,2}, ZHENG Fei¹, ZHANG Li¹

¹(School of Information Management, Shanghai Lixin University of Accounting and Finance, Shanghai 201620, China)

²(Institute of Data Science and Interdisciplinary Studies, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China)

Abstract: Bayesian network is a powerful tool for studying the causal relationship between variables. Causal learning, based on Bayesian network, consists of two parts: structure learning and parameter learning. Structural learning is the core of causal learning. At present, Bayesian network is mainly used to discover the causality in non time series data (non time series causality) and what is learned from the data is the causal relationship between general variables. In this paper, the causality of time series is learned by time series preconditioning, time series variable sorting, the construction of transformation data set, local greedy search-scoring, maximum likelihood parameter estimation and so on. Combining the time series preconditioning of including segmentation, the structure learning of causal relationship for time series segments, the construction of causality structure data set, the variable sorting of causal relationship, local greedy search-scoring and maximum likelihood parameter estimation etc, meta causal relationship (used to study the randomness of causal relationship) is established. Two levels of causality learning can be realized to lay the foundation for further quantitative causal analysis. Experiments and analyses are done by using simulation, UCI and finance time series to verify the validity, reliability and practicability of learning causal relationship and meta causality based on Bayesian network.

* 基金项目: 国家社会科学基金(18BTJ020)

Foundation item: Foundation item: National Social Science Foundation of China (18BTJ020)

收稿时间: 2018-10-23; 修改时间: 2019-11-06; 采用时间: 2020-01-20; jos 在线出版时间: 2021-01-15

Key words: time series; causal relationship; Bayesian network; transformation data set; structure data set

人类对现实世界中现象的一种强烈渴望就是因果联系,从古至今人们不间断地从不同层次和角度探索因果理论和发现因果关系的方法,以达到更好地认识和改造世界的目的.早期的因果关系属于哲学的范畴,现代更强调从数据中的因果关系发现.时间序列是现实世界数据的重要表现形式之一,在宏观经济与金融等领域,数据主要以时间序列的形式存在,大量宏观经济与金融时间序列真实地记录了系统在不同时间点(或时间片)的各种重要信息,其中蕴含着丰富而有价值的因果关系和映射规则等方面的知识,这些知识往往是诊断宏观经济与金融体系运行情况,揭示经济运行规律,以及制定相应的调控政策的重要依据.

目前主要采用格兰杰(Granger)方法探索和发现时间序列中的因果关系,如Ferreira^[1]采用面板数据格兰杰因果关系方法所进行的欧盟债务与经济增长研究, Fredrik^[2]对中国金融增长的Granger 因果分析,Chang^[3]基于格兰杰方法的金砖五国煤炭消费与经济增长的因果关系检验等.但这种方法存在如下一些局限性:① 进行格兰杰因果关系检验的一个前提条件是时间序列必须具有平稳性,否则可能会出现虚假回归现象;② 格兰杰因果关系检验的结论只是一种预测,是统计意义上的“格兰杰因果性”,而不是真正意义上的因果关系,不能作为肯定或否定因果关系的依据;③ 格兰杰因果关系是建立在线性回归基础之上,因此,从某种意义上可以说格兰杰因果关系是线性因果关系.近些年也有一些将其它线性和非线性回归模型用于检验因果关系的研究,如Kristofer^[4]基于岭回归的因果关系检验,David^[5]采用广义回归发现因果关系,Ryutah^[6]使用线性分位数回归研究因果关系,Luo^[7]依据逐步回归的因果效应分析等.这些因果关系发现(或检验)方法所依据的是回归计算,往往是针对特定的问题和具体的方面.

概率和因果关系具有密切的联系,基于概率理论探索因果关系已是目前因果关系研究的一个重要方向.在概率与因果关系的基础理论方面,已有许多研究,如Rubin^[8]的因果模型,Pearl^[9]的因果图、Heckman^[10]的科学因果关系模型、John^[11]的因果推断原理与方法研究等,其中最有影响的是Rubin 和Pearl 的因果关系模型,在2000年,Pearl 介绍了这两个模型的等价性.Rubin 因果模型强调局部细节,一般用于较少变量之间的精细因果分析;Pearl 的因果模型(或贝叶斯网络)更加直观,突出整体因果联系,适合于复杂多因素之间的因果知识表示与推理,本文基于Pearl 的贝叶斯网络进行时间序列的因果关系学习研究.

贝叶斯网络是描述随机变量(简称为变量)之间相互影响与制约关系的有向图模型,由结构(有向无环图)和参数(条件概率分布表)两部分构成,其结构中弧的方向具有因果语义(贝叶斯网络的创始人pearl也因将概率与因果关系的结合而获得2011年图灵奖),因此是因果建模与分析的有力工具,将贝叶斯网络用于时间序列因果分析(离散变量)是一种发展的趋势,也将会与格兰杰方法等(连续变量)形成互补.基于贝叶斯网络进行因果关系研究的核心是贝叶斯网络(或因果关系网络)学习.上世纪九十年代之前,以根据专家知识的贝叶斯网络构建为主,如以Kullback^[12]为代表的整体和局部学习方法,这种方法使得贝叶斯网络的学习与更新都比较困难,而且往往带有主观倾向性,影响了贝叶斯网络的广泛应用.九十年代以后,随着数据获取和存储技术的发展,以及人们数据意识的增强,开始逐渐转向将专家主观知识与客观数据相结合来建立贝叶斯网络,以及完全基于数据的贝叶斯网络学习,这期间比较有影响的工作有Cheng^[13]、Liu^[14]、Parviainen^[15]和Xiao^[16]等的依赖分析方法,Heckerman^[17]、Suzuki^[18]、Gheisar^[19]和Liu^[20]等的打分-搜索方法.这些贝叶斯网络学习研究所针对的是非时间序列(数据集中的记录之间需满足独立同分布的假设)数据和一般的随机变量,不适合于时间序列(数据集中的记录之间具有时序依赖)的情况.Friedman在1998年将贝叶斯网络同马尔科夫链相结合,在平稳性和马尔科夫性两个假设下,给出了动态贝叶斯网络(dynamic Bayesian network)^[21],用于变量的时序变化研究,随后的动态贝叶斯网络研究基本遵循Friedman的框架,它们需要使用完整的面板数据(获得这样的数据非常困难)和复杂的打分函数运算进行学习,对时间序列的因果关系学习不具有实用性.

本文的主要贡献如下:

(1) 提出了建立时间序列转换数据集的思想与方法,通过转换数据集来实现时序与非时序信息的整合与统一,使在转换数据集的基础上采用贝叶斯网络方法能够进行时滞、非时滞和混合因果关系学习,以及因果推理计

算,这将突破已有贝叶斯网络学习的局限性和动态贝叶斯网络框架的约束,使贝叶斯网络也成为时间序列因果分析的有力工具.

(2) 结合时间序列预处理、时间序列变量排序、转换数据集构建、MDL(minimal description length)标准、局部贪婪搜索等,建立了多变量时间序列的因果关系学习方法,其中能够反映需求的基于模式的离散化使因果关系更具针对性和可解释性,而通过打分-搜索发现时序父结点来确定时滞变量,相对于传统的相关性计算能够更有效地界定转换数据集中的时滞信息,来避免信息丢失和冗余,这种因果关系学习也将与格兰杰方法等形成互补,推动因果分析理论与应用的发展.

(3) 在时间序列段因果关系结构学习的基础上,将因果关系结构数据集构建,因果关系变量选择和排序、局部贪婪打分-搜索等相结合,提出了建立元因果关系(因果关系变量之间的因果关系)的思想与方法,这将扩展贝叶斯网络的应用范围,也为进一步的因果关系演化分析和未来发展趋势判断奠定了坚实的基础,并分别采用模拟、UCI 和金融时间序列对因果关系与元因果关系学习的有效性、可靠性和实用性进行了实验验证与分析.

文章包括五个部分:第一部分对因果关系和贝叶斯网络的产生与发展等进行了回顾与分析;第二部分是时间序列变量的因果关系学习;第三部分是因果关系变量的元因果关系学习;第四部分是实验与分析;第五部分是结论和进一步的工作.另外,在文章中,概率模式中的变量,以及表示概率模式的图形模式中的结点有时不加区分.

1 时间序列的因果关系学习

时间序列的因果关系(或时间序列变量之间的因果关系)学习由时间序列预处理、因果关系结构学习和因果关系参数学习三部分构成.

1.1 时间序列预处理

我们首先采用滑动平均的方法初始化丢失的数据,然后将时间序列离散化,最后对初始化的离散丢失数据进行迭代修复. 用 $\mathbf{Z}[t] = (Z_1[t], Z_2[t], \dots, Z_n[t])$ 表示时间序列向量(简称为时序向量),

$\mathbf{z}[t] = (z_1[t], z_2[t], \dots, z_n[t])$ 为其配置, $D_c[n, T] = \{z_1[t], z_2[t], \dots, z_n[t]\} (1 \leq t \leq T)$ 是具有 T 个记录的时间序列数据集(简称为时序数据集).

(1) 丢失数据的初始化

丢失数据的修复一直是一个重要且困难的研究课题,时间序列的丢失数据普遍存在,如果删除这些丢失数据所在的记录,会导致信息损失,而使用不当的丢失数据修复方法,则可能引入大量的噪声.对给定的窗体大小

$\omega \geq 1$,用 $\hat{z}_i[t]$ 表示丢失数据 $z_i[t]$ 的初始值, $\hat{z}_i[t]$ 的计算方法如下:

$$\hat{z}_i[t] = \begin{cases} \frac{1}{\omega} (z_i[t_1] + z_i[t_2] + \dots + z_i[t_\omega]), & t \leq \omega \\ \frac{1}{\omega} (z_i[t - \omega] + z_i[t - \omega + 1] + \dots + z_i[t - 1]), & t > \omega \end{cases}$$

(1)

其中 $t \leq \omega$ 时的 $z_i[t_1], z_i[t_2], \dots, z_i[t_\omega]$ 是从 $z_i[1]$ 开始依次出现的 ω 个非丢失数据(包括初始化后的丢失数据),当 $t > \omega$ 时便不需要从 $z_i[1]$ 开始选取.

(2) 时间序列的离散化

离散化是一种抽象(离散值是连续值的抽象,一个离散值对应一个连续值的集合,去除连续值的个性内容,保留共性部分),也可以认为是一种概念的形成(每一个离散值都可以看作是对应连续值所形成的概念,离散化就是只保留我们所关注的特征,去除不需要的特征).在连续变量离散化的过程中,我们对一个离散值对应的连续值集合中的所有值不加以区分,因此会丢弃一些细节信息(相对于需要解决的问题,也可能是冗余或干扰信息),但离散化后的数据更有意义且容易理解(人类的知识和推理均不具有连续性),这对解决一些特定的问题十分必要.Pearl 因果关系中的变量要求取离散值,这样所有因果关系中的变量便形成一个概率分布,因果关系蕴含在离散变量的概率分布之中(也可以将Pearl 的因果关系称为概率因果关系).我们需要将连续的数据(或变量)离散化,也就是去除与因果关系无关的细节信息,再去发现数据中所蕴含的因果关系(因果关系学习).在理论上,一般的离散化方法是NP 困难问题(每一种子集的划分都对应一个离散化策略),在因果关系学习时,我们通常都采用特定的离散化方法,以降低运算复杂性和使因果关系更具有针对性.不同的离散化方法将产生不同的分布,这样会使学习得到的因果关系也具有差异和不同的含义,因此,需要根据实际问题的需要来选择(或设计)离散化方法,以便于解释学习得到因果关系.

① 离散化方法的构成

现已发展了许多离散化的方法,但这些方法主要基于数据的内部结构,无法满足时间序列数据变化方面的需求(或数据随时间变化的规律与模式),我们给出基于时间序列段模式的离散化方法来解决这一问题,关于时间序列因果关系学习中数据(或变量)离散化的具体情况如图1 所示.

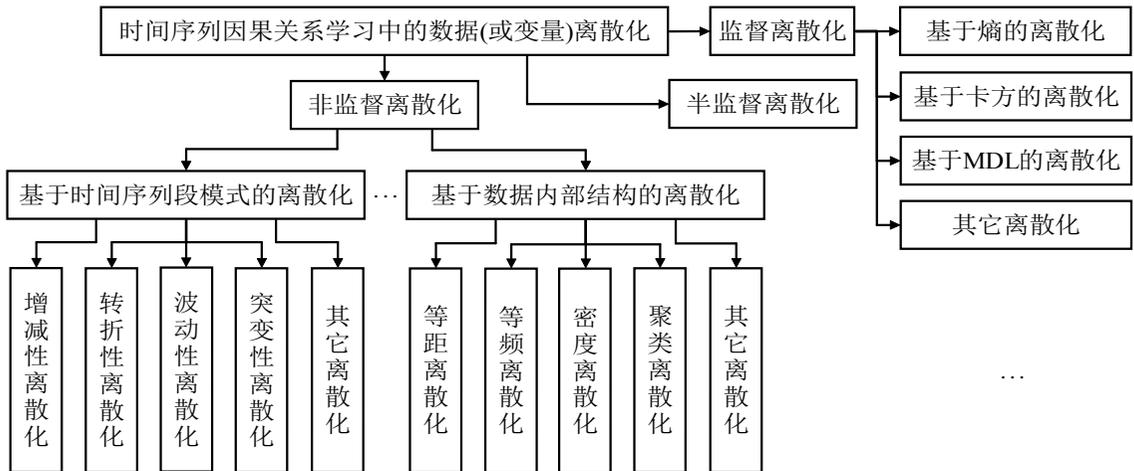


Fig.1 Constitution of data discretization methods in time series causality learning

图1 时间序列因果关系学习中的数据离散化方法的构成

图1 显示,用于时间序列因果关系学习中的离散化方法可分成非监督离散化和监督离散化两种,本文采用非监督离散化方法.在非监督离散化方法中,即可以采用反映时序变化的模式离散化(如反映增减性、转折性和波动性等方面的离散化,这些离散化方法往往与实际需求密切相关,使学习得到的因果关系具有更丰富的内涵),也可以使用与时序无关的结构离散化(如等频、等距和聚类离散化等,所描述的是数据内部的结构和特征,一般不会增加因果关系的内涵).离散化方法的变化对因果关系会产生影响,在本文的第四部分我们将进行相应的实验与分析.

用 $\mathbf{X}[t] = (X_1[t], X_2[t], \dots, X_n[t])$ 表示离散化后的时序向量, $\mathbf{x}[t] = (x_1[t], x_2[t], \dots, x_n[t])$ 为其配置, $D_d[n, T] = \{x_1[t], x_2[t], \dots, x_n[t]\} (1 \leq t \leq T)$ 表示离散化后时序数据集, $G[t]$ 表示因果

关系结构(有向无环图), $\{x_i^1, x_i^2, \dots, x_i^{r_i}\}$ 是 $x_i[t]$ 的值域, $x_i^j (1 \leq j \leq r_i)$ 可以是数字, 也可以是没有数字含义的符号等, ϕ (或 ψ) 是一个有限值映射.

② 基于时间序列段模式的离散化

用 $M^k[i, t] (1 \leq k \leq r_i)$ 表示在时间序列段 $z_i[t - \eta], z_i[t - \eta + 1], \dots, z_i[t], \dots, z_i[t + \tau - 1], z_i[t + \tau] (\Theta_i(t, \eta, \tau))$ 表示时间序列段, $\eta \geq 0, \tau \geq 0, \eta$ 和 τ 不能同时为 0) 上的一个模式, $M[i, t]$ 表示所有模式的集合, 时间序列段模式离散化(简称为模式离散化)方法可以表示为:

$$x_i[t] = \phi(z_i[t]) = \begin{cases} x_i^1, \Theta_i(t, \eta, \tau) \sim M^1[i, t] \\ x_i^2, \Theta_i(t, \eta, \tau) \sim M^2[i, t] \\ \dots \\ x_i^{r_i}, \Theta_i(t, \eta, \tau) \sim M^{r_i}[i, t] \end{cases}$$

(2)

其中 $\Theta_i(t, \eta, \tau) \sim M^k[i, t]$ 表示 $\Theta_i(t, \eta, \tau)$ 满足模式 $M^k[i, t]$, $M^1[i, t] \cup M^2[i, t] \cup \dots \cup M^{r_i}[i, t] = M[i, t]$, $M^h[i, t] \cap M^k[i, t] = \varphi$, 其中 $1 \leq h, k \leq r_i, h \neq k$.

③ 基于数据内部结构的离散化

用 $D_c[n, T, i]$ 表示单变量时序数据集 $\{z_i[t] | 1 \leq t \leq T\}$, $D_c^k[n, T, i] (1 \leq k \leq r_i)$ 是 $D_c[n, T, i]$ 的一个子集, 描述数据内部结构的离散化(简称为结构离散化)方法可以表示为:

$$x_i[t] = \psi(z_i[t]) = \begin{cases} x_i^1, z_i[t] \in D_c^1[n, T, i] \\ x_i^2, z_i[t] \in D_c^2[n, T, i] \\ \dots \\ x_i^{r_i}, z_i[t] \in D_c^{r_i}[n, T, i] \end{cases}$$

(3)

其中 $D_c^1[n, T, i] \cup D_c^2[n, T, i] \cup \dots \cup D_c^{r_i}[n, T, i] = D_c[n, T, i]$, $D_c^h[n, T, i] \cap D_c^k[n, T, i] = \varphi$, 其中 $1 \leq h, k \leq r_i, h \neq k$.

一种模式离散化方法可能包含许多具体的离散化方案, 下面给出转折性离散化中的三个方案.

四值离散化:

$$x_i[t] = \phi(z_i[t]) = \begin{cases} x_i^1, & z_{i-1}[t] \geq z_{i-2}[t] \text{ 且 } z_i[t] \leq z_{i-1}[t] \\ x_i^2, & z_{i-1}[t] \leq z_{i-2}[t] \text{ 且 } z_i[t] \geq z_{i-1}[t]; \\ x_i^3, & z_{i-1}[t] \leq z_{i-2}[t] \text{ 且 } z_i[t] \leq z_{i-1}[t] \\ x_i^4, & z_{i-1}[t] \geq z_{i-2}[t] \text{ 且 } z_i[t] \geq z_{i-1}[t] \end{cases}$$

三值离散化:

$$x_i[t] = \phi(z_i[t]) = \begin{cases} x_i^1, & z_{i-1}[t] \geq z_{i-2}[t] \text{ 且 } z_i[t] \leq z_{i-1}[t] \\ x_i^2, & z_{i-1}[t] \leq z_{i-2}[t] \text{ 且 } z_i[t] \geq z_{i-1}[t]; \\ x_i^3, & \text{其它情况} \end{cases}$$

二值离散化:

$$x_i[t] = \phi(z_i[t]) = \begin{cases} x_i^1, & z_{i-1}[t] \geq z_{i-2}[t] \text{ 且 } z_i[t] \leq z_{i-1}[t] \text{ 或 } z_{i-1}[t] \leq z_{i-2}[t] \text{ 且 } z_i[t] \geq z_{i-1}[t]. \\ x_i^2, & \text{其它情况} \end{cases}$$

(3) 初始化数据的迭代修复

最大似然树具有确定的建树算法,而且在效率(其参数最多只是一阶条件概率)、可靠性(最大似然树是与贝叶斯网络具有最好拟合的树形结构^[22],往往构成贝叶斯网络结构的主体骨架,所确定的联合概率分布也是产生数据的联合概率分布的良好近似)和鲁棒性(建立最大似然树只使用互信息由大到小排序后的顺序信息,而顺序信息是比较稳定的信息)等方面均有优势,我们将最大似然树和 Gibbs 抽样相结合,对初始化后的丢失数据进行迭代修复,直到满足终止条件结束迭代。

最大似然树的建树过程如下^[22]:

① 把变量之间的互信息作为边的权重,计算所有边的权重,并按权重由大到小排序边。

② 按照不产生环路的原则,根据边权重的大小,由大到小依次选择边来建树,直到选择 $n - 1$ 条边为止,得到最大似然树 *Tree*。

选择一个结点作为根结点,由根结点向外的方向为边定向,定向的目的是便于分解联合概率,这种方向不具有因果含义。

根据概率公式可得:

$$\begin{aligned} & p(x_i[t] \mid x_1[t], \dots, x_{i-1}[t], x_{i+1}[t], \dots, x_n[t]) \\ &= \frac{p(x_1[t], x_2[t], \dots, x_n[t])}{p(x_1[t], \dots, x_{i-1}[t], x_{i+1}[t], \dots, x_n[t])} \\ &= \alpha \prod_{i=1}^n p(x_i[t] \mid \pi(x_i[t]), \text{Tree}) \end{aligned}$$

(4)

其中 α 是与 $x_i[t]$ 无关的量, $\pi(x_i[t])$ 是最大似然树中 $X_i[t]$ 的父结点 $\Pi(X_i[t])$ 的取值.对(4)式进行归一化处

$$\text{理,记 } \xi(k) = \frac{\prod_{i=1}^n p(x_i^k | \pi(x_i[t]), Tree)}{\sum_{h=1}^{r_i} \prod_{i=1}^n p(x_i^h | \pi(x_i[t]), Tree)},$$

$$\text{对生成的随机数 } \lambda, \text{ 变量 } X_i[t] \text{ 丢失数据的修正值为: } \hat{x}_i[t] = \begin{cases} x_i^1, & 0 < \lambda \leq \xi(1) \\ \dots\dots\dots \\ x_i^k, & \sum_{j=1}^{k-1} \xi(j) < \lambda \leq \sum_{j=1}^k \xi(j) \\ \dots\dots\dots \\ x_i^{r_i}, & \lambda > \sum_{j=1}^{r_i-1} \xi(j) \end{cases}$$

(5)

我们能够发现,丢失数据的迭代修复实际上采用的是贝叶斯网络局部随机模拟技术,由于 $X_1[t], X_2[t], \dots, X_n[t]$ 的真实贝叶斯网络未知,因此使用最大似然树来代替贝叶斯网络,在迭代中,使用最大似然树产生的局部模拟数据依次更新上一次迭代修正过的丢失数据(或者初始化的数据),这一迭代过程在理论上将收敛到全局平稳分布.这样的迭代修复能够有效减少信息的丢失和降低噪声的引入量,提高经过丢失数据修复的数据集质量.

1.2 时间序列的因果关系结构学习

时间序列的因果关系结构学习是因果关系学习的核心,在时间序列预处理的基础上,我们结合变量排序,转换数据集构建和局部贪婪打分-搜索进行时间序列的因果关系结构学习.

(1) 时间序列变量的排序

目前,基于数据的变量排序主要采用打分-搜索方法(打分函数一般使用卡方或 MDL 标准等,搜索策略则选择贪婪或启发式搜索等),不但效率低,而且往往不能完全排序(存在一些变量无法排序),我们基于信息理论给出变量删除法来排序变量,不但效率高,还可完全排序变量.时间序列变量(简称为时序变量)包括非时滞变量和时滞变量,我们首先采用变量删除法来排序非时滞变量(时序变量排序的主要部分),然后依据专家知识(数据产生的先后顺序)对非时滞变量的顺序进行调整.调整的具体过程是:当依据变量删除法对非时滞变量进行排序之后,如果存在两个非时滞变量之间的顺序与专家知识不一致,交换这两个非时滞变量在序列中的位置,这时也可能产生新的不一致,继续调整,直到非时滞变量的顺序与专家知识没有冲突为止,最后再按照时间顺序加入时滞变量,得到最终时序变量的顺序.

对于非时滞变量矩阵 $(\delta_{ij})_{n \times n}$, 首先选取 $\delta_{ij} \geq 0 (i, j = 1, \dots, n, \text{ 当且仅当 } i = j \text{ 时 } \delta_{ij} = 0)$ 的行所对应的变量作为第一个变量(排在最前面的变量), $\delta_{ij} \geq 0$ 的行表示 X_i 对 $X_j (j = 1, \dots, i-1, i+1, \dots, n)$ 的影响要大于 X_j 对 X_i 的影响,在矩阵中删除满足条件的行和对应的列(对应的第 i 列则表示 X_j 对 X_i 的影响要小于 X_i 对 X_j 的影响),剩下的是一个 $n-1$ 阶矩阵,在这个 $n-1$ 阶矩阵中重复第一个变量的选择方法可得到第二个变量,如此下去便可对所有变量排序.为表述方便,仍用

$X_1[t], X_2[t], \dots, X_n[t]$ 和 $D_d[n, T]$ 表示排序后的非时滞变量和对应的时序数据集, δ_{ij} 的计算方法如下:

$$\delta_{ij} = -\delta_{ji} = \begin{cases} +1, & \frac{H(X_i[t] | X_j[t])}{H(X_i[t])} < \frac{H(X_j[t] | X_i[t])}{H(X_j[t])} \\ 0, & \frac{H(X_i[t] | X_j[t])}{H(X_i[t])} = \frac{H(X_j[t] | X_i[t])}{H(X_j[t])} \\ -1, & \frac{H(X_i[t] | X_j[t])}{H(X_i[t])} > \frac{H(X_j[t] | X_i[t])}{H(X_j[t])} \end{cases}$$

(6)

其中 $H(X_i[t] | X_j[t])$ 和 $H(X_j[t] | X_i[t])$ 是条件熵, $\frac{H(X_i[t] | X_j[t])}{H(X_i[t])}$ 是条件相对熵, 表示 $X_j[t]$ 为

$X_i[t]$ 提供的相对信息. 如果 $X_i[t]$ 是 $X_j[t]$ 的原因, 那么相对于 $X_j[t]$ 为 $X_i[t]$ 提供的相对信息, $X_i[t]$ 将为 $X_j[t]$ 提供更多的相对信息, 我们将这种非时滞变量的排序方法称为变量删除法. 我们可以证明使用变量删除法排序非时滞变量不会产生环路.

(2) 转换数据集构建

转换数据集能够实现时滞与非时滞信息的统一, 从而可在转换数据集的基础上采用贝叶斯网络方法发现多变量时间序列中所蕴含的时滞、非时滞和混合因果关系.

定义 1 对给定的时序数据集 $D_d[n, T] = \{x_1[t], x_2[t], \dots, x_n[t]\}$ ($n \geq 1, 1 \leq t \leq T$), 称数据集

$$D_d[n, T, \mathbf{q}] = \{x_1[t - v_{q_1}^1], \dots, x_1[t - v_1^1], x_1[t], \dots, x_n[t - v_{q_n}^n], \dots, x_n[t - v_1^n], x_n[t]\} (\max\{q_1, q_2, \dots, q_n\} + 1 \leq t \leq T)$$

(7)

为 $D_d[n, T]$ 的转换数据集, 其中 $\mathbf{q} = (q_1, q_2, \dots, q_n)$ ($q_i \geq 0, 1 \leq i \leq n$) 是时滞阶数向

量, $\{X_i[t - v_{q_i}^i], \dots, X_i[t - v_i^i]\}$ 是 $X_i[t]$ 的时滞变量集, 时序变量的数量是 $n + \sum_{i=1}^n q_i$, 记录的数量是

$$T - \max\{q_1, q_2, \dots, q_n\}.$$

定义 1 给出的转换数据集可用于单变量和多变量时间序列的因果关系学习, 对于单变量时间序列 ($n = 1$), 一般设置 $v_{q_i}^i = q_i$ ($1 \leq i \leq n$). 假设 $n > 1$, 在 $D_d[n, T, \mathbf{q}]$ 中, 我们将

$X_1[t], \dots, X_{i-1}[t], X_{i+1}[t], \dots, X_n[t]$ 与 $X_i[t]$ 之间的因果关系称为非时滞因果关系, $X_i[t - v_u^i]$ ($1 \leq u \leq q_i$) 与 $X_i[t]$ 之间为时滞因果关系, 其它时序变量与 $X_i[t]$ 之间为混合因果关系, 确

定 $X_i[t]$ 的时滞变量的方法将在下面给出.

(3) 基于打分-搜索的因果关系结构学习

我们采用打分-搜索方法进行时间序列转换数据集的因果关系结构学习, 打分函数选择 MDL 标准, 基于 MDL 标准的可分解性和变量顺序, 只需进行父子局部结构的打分与比较, 为避免搜索的指数复杂性, 搜索策略采用贪婪搜索. 首先建立单变量时间序列 $\{x_i[t]\} (1 \leq i \leq n, 1 \leq t \leq T)$ 的转换数据集 $D_d[1, T, i] = \{x_i[t - s_i], x_i[t - s_i + 1], \dots, x_i[t - 1], x_i[t]\}$, 在此基础上, 进行单变量时间序列的因果关系结构学习, 假设学习得到 $X_i[t]$ 的父结点集是 $\{X_i[t - v_{q_i}^i], \dots, X_i[t - v_1^i]\} (q_i \leq s_i)$, 就将这个父结点集作为多变量时间序列的转换数据集构建中 $X_i[t]$ 的时滞变量集.

仍用 $X_1[t], X_2[t], \dots, X_n[t]$ 表示排序后的非时滞变量, 转换数据集中时序变量的顺序是 $X_1[t - v_{q_1}^1], \dots, X_1[t - v_1^1], X_1[t], X_2[t - v_{q_2}^2], \dots, X_2[t - v_2^2], X_2[t], \dots, X_n[t - v_{q_n}^n], \dots, X_n[t - v_1^n], X_n[t]$, 在此基础上, 通过局部贪婪打分-搜索建立多变量时间序列的因果关系结构. $X_i[t]$ 的父结点集的构成如图 2 所示, 其中 $\Pi_{X_i[t]}$ 是 $X_i[t]$ 在 $\{X_1[t], X_2[t], \dots, X_{i-1}[t]\}$ 中的非时滞父结点集, $\Pi_{X_i[t]}^{X_i[t - q_i]}$ 是 $X_i[t]$ 在 $\{X_i[t - v_{q_i}^i], \dots, X_i[t - v_2^i], X_i[t - v_1^i]\}$ 中的时滞父结点集, $\Pi_{X_i[t]}^{X_1[t - q_1]}$ 和 $\Pi_{X_i[t]}^{X_{i-1}[t - q_{i-1}]}$ 则是 $X_i[t]$ 在 $\{X_1[t - v_{q_1}^1], \dots, X_1[t - v_2^1], X_1[t - v_1^1]\}$ 和 $\{X_{i-1}[t - v_{q_{i-1}}^{i-1}], \dots, X_{i-1}[t - v_2^{i-1}], X_{i-1}[t - v_1^{i-1}]\}$ 中的混合父结点集.

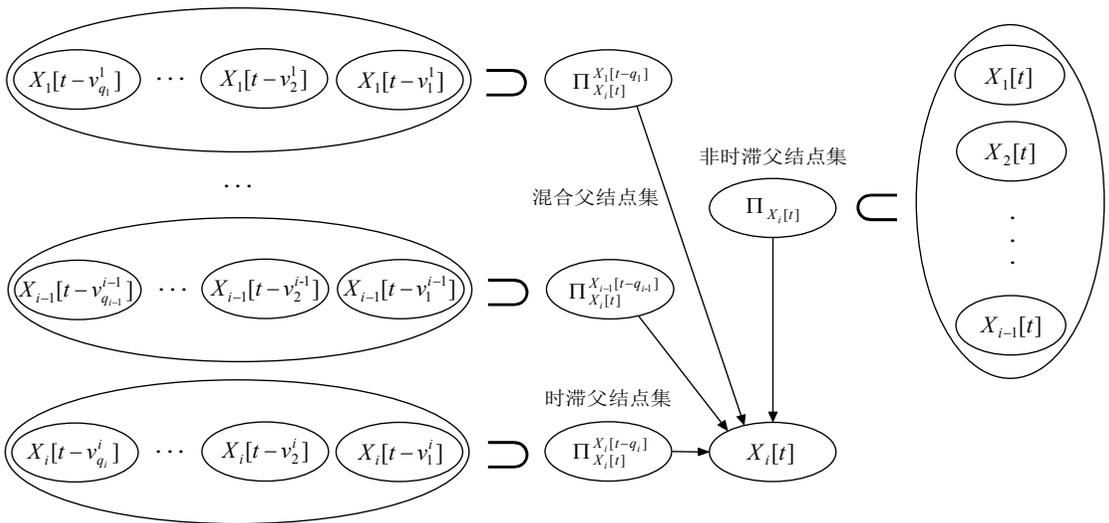


Fig.2 Composition of $X_i[t]$ parent node set

图2 $X_i[t]$ 父结点集的构成

1.3 时间序列的因果关系参数学习

因果关系的参数学习存在多种方法,比较常用的有两种^[17,23],它们是最大似然估计(因果关系参数学习)和最大后验估计(因果关系和元因果关系的参数学习),因果关系的参数可以表示为:

$$\theta_{ijk}[t] = p(x_i^k[t] | \pi_i^j[t], \theta_i[t], G[t]) \geq 0$$

(8)

其中 $\pi_i^1[t], \pi_i^2[t], \dots, \pi_i^{w_i}[t]$ ($w_i = \prod_{X_j[t] \in \Pi_i[t]} r_j$) 是 $X_i[t]$ 的父结点集 $\Pi_i[t]$ 所有可能的配

置, $\theta_i[t] = ((\theta_{ijk}[t])_{k=2}^{r_i})_{j=1}^{w_i}$ ^[24] 是参数向量(由 $\sum_{k=1}^{r_i} \theta_{ijk}[t] = 1$ 可得 $\theta_{ij1}[t] = 1 - \sum_{k=2}^{r_i} \theta_{ijk}[t]$).

分别用 $\hat{\theta}_{ijk}[t]$ 和 $\tilde{\theta}_{ijk}[t]$ 表示 $\theta_{ijk}[t]$ 的最大似然和最大后验估计,那么

$$\hat{\theta}_{ijk}[t] = \frac{N_{ijk}[t]}{N_{ij}[t]}$$

(9)

$$\tilde{\theta}_{ijk}[t] = \frac{\alpha_{ijk}[t] + N_{ijk}[t] - 1}{\alpha_{ij}[t] + N_{ij}[t] - r_i}$$

(10)

其中 $N_{ijk}[t]$ 是充分统计因子,是在时间序列的转换数据集中 $X_i[t] = x_i^k[t]$ 且 $\Pi_i[t] = \pi_i^j[t]$ 的情况数

量, $N_{ij}[t] = \sum_{k=1}^{r_i} N_{ijk}[t]$, $\alpha_{ijk}[t]$ 和 $\alpha_{ij}[t]$ 是先验充分统计因子,可以依据经验和历史数据获得.我们采用

最大似然估计方法进行因果关系的参数学习,多变量时间序列的因果关系学习算法如算法1所示.

算法1. 多变量时间序列的因果关系学习算法.

输入: 多变量时间序列数据集 $D_c[n, T]$

输出: 多变量时间序列的因果关系(包括结构和参数)

时间序列丢失数据的初始化、离散化和迭代修复

时序变量的排序

For $i=1$ to n

 基于 $\{x_i[t]\}$ ($1 \leq t \leq T$) 的转换数据集,通过局部贪婪打分-搜索发现 $X_i[t]$ 的时滞父结点集

 确定 $X_i[t]$ 的时滞变量

End for

For $i=1$ to n

 基于 $\{x_1[t], x_1[t], \dots, x_n[t]\}$ ($1 \leq t \leq T$) 的转换数据集,通过局部贪婪打分-搜索发现 $X_i[t]$ 的

三种父结点集

依据最大似然估计方法进行 $X_i[t]$ 的参数学习

End for

在多变量时间序列的因果关系学习算法中,运算的主要部分是确定时滞变量(单变量时间序列的因果关系结构学习)和多变量时间序列的因果关系结构学习.在单变量时间序列的因果关系结构学习中,需要进行不超过 $n(4s^* - 6)$ 次的MDL计算(我们限定父结点数量不超过4),其中 $s^* = \max\{s_1, s_2, \dots, s_n\}$;而在多变量时间序列的因果关系结构学习中,最多需要进行 $\frac{1}{2}q^*n(q^*n - 1) + 1$ 次的MDL计算,其中 $q^* = \max\{q_1, q_2, \dots, q_n\}$, $q_i \leq s_i (1 \leq i \leq n)$. s^* 和 q^* 是与 n 无关的量,因此,相对于MDL打分计算,多变量时间序列因果关系学习算法的时间复杂度是 $O(n^2)$.

2 时间序列的元因果关系学习

时间序列的元因果关系学习是因果关系学习的深化和拓展.时间序列的因果关系学习适合于低频常规时间序列(也可以用于高频时间序列的局部或子列),而时间序列的元因果关系学习则主要针对高频大时间序列.时间序列的元因果关系学习包括时间序列预处理、在时间序列段因果关系结构学习基础上的因果关系结构数据集构建和变量选择,以及元因果关系的结构与参数学习三个部分.

2.1 时间序列预处理

在时间序列预处理中,除丢失数据修复(如果需要突出效率,可以只采用滑动平均的方法来修复丢失的数据)和离散化(适合于使用模式离散化方法)之外,还需要时间序列分段.如果时间序列中的数据是按照固定时间跨度采集(时间均匀地采集),那么可选择确定的时间跨度来为时间序列分段;当时间序列中的数据不是按时间均匀分布,则可按数据量为时间序列分段;当然,也可以根据实际情况和需求为时间序列分段,时间序列段的大小可以一致,也可以不一致,无论采用哪种方式分段,时间序列段的大小和数量应该能够有效地进行因果关系和元因果关系学习.

假设将时间序列分为 Γ 个段,用 $X_1^{(m)}[t], X_2^{(m)}[t], \dots, X_n^{(m)}[t]$ 表示第 m 个时间序列段的排序后(同样采用变量删除法来排序)非时滞变量,其中 $T_1, T_2, \dots, T_{\Gamma-1}$ 是时间序列段的分点, $T_{m-1} + 1 \leq t \leq T_m, 1 \leq m \leq \Gamma, T_0 = 0, T_\Gamma = T, D_d^{(m)}[n, T]$ 是对应的时序数据集,那么

$$D_d^{(m)}[n, T, \mathbf{q}] = \left\{ x_1^{(m)}[t - v_{q_1}^1], \dots, x_1^{(m)}[t - v_1^1], x_1^{(m)}[t], \dots, x_n^{(m)}[t - v_{q_n}^n], \dots, x_n^{(m)}[t - v_1^n], x_n^{(m)}[t] \right\}$$

$(T_{m-1} + \max\{q_1, \dots, q_n\} + 1 \leq t \leq T_m)$ 是 $\{x_1^{(m)}[t], x_2^{(m)}[t], \dots, x_n^{(m)}[t]\} (T_{m-1} + 1 \leq t \leq T_m)$ 的转换数据集.在时间序列预处理、时间序列段的时序变量排序与转换数据集构建的基础上,同样采用打分-搜索方法进行时间序列段的因果关系结构学习,用 $G^{(m)}[t]$ 表示学习得到的因果关系结构.接下来是因果关系结构数据集构建和因果关系变量选择与排序,然后进行元因果关系的结构与参数学习,最终得到时间序列的元因果关系.

2.2 因果关系结构数据集构建和变量选择

在时间序列段的时序变量因果关系结构学习的基础上,需要将这些因果关系结构转化为用于元因果关系学习的数据集,即因果关系结构数据集.

定义 2 在第 m 段时间序列的因果关系结构 $G^{(m)}[t]$ 中,不考虑混合因果关系,时序变量的顺序依次为

$$X_1^{(m)}[t - v_{q_1}^{(m)}], \dots, X_1^{(m)}[t - v_1^{(m)}], X_1^{(m)}[t], \dots,$$

$$X_n^{(m)}[t - v_{q_n}^{(m)}], \dots, X_n^{(m)}[t - v_1^{(m)}], X_n^{(m)}[t], \text{用 } Y_{ij}^{(m)}[t] \text{ 表示 } X_i^{(m)}[t - h] \text{ 和 } X_j^{(m)}[t] \text{ 之间的因果关系变量(当}$$

$h = 0$ 和 $i \neq j$ 时,是非时滞因果关系;当 $h \neq 0$ 和 $i = j$ 时,是时滞因果关系),

$Y_{ij}^{(m)}[t] = f(X_i^{(m)}[t - h], X_j^{(m)}[t])$,由 $Y_{ij}^{(m)}[t]$ 构成的记录称为因果关系结构记录,所有因果关系

结构记录构成的数据集则称为因果关系结构数据集. $f(X_i^{(m)}[t - h], X_j^{(m)}[t])$ 的取值情况如下:

$$f(X_i^{(m)}[t - h], X_j^{(m)}[t]) = \begin{cases} 1, & \text{存在弧 } X_i^{(m)}[t] \rightarrow X_j^{(m)}[t] \\ 0, & X_i^{(m)}[t] \text{ 和 } X_j^{(m)}[t] \text{ 之间不存在弧, } h = 0 \\ -1, & \text{存在弧 } X_i^{(m)}[t] \leftarrow X_j^{(m)}[t] \\ \begin{cases} 1, & \text{存在弧 } X_i^{(m)}[t - h] \rightarrow X_j^{(m)}[t] \\ 0, & \text{不存在弧} \end{cases}, & h \neq 0 \end{cases}$$

(11)

我们将因果关系结构记录分成两个部分,第一部分是 $(X_i^{(m)}[t], X_j^{(m)}[t])$ ($i \neq j$) 产生的非时滞因果关系变量

值(这部分的维数固定),第二部分是 $(X_i^{(m)}[t - h], X_i^{(m)}[t])$ ($h \neq 0$) 产生的时滞因果关系变量数对(这部分的

维数不固定,数对为 (h, i)),这样不同因果关系结构记录的维度可能不同,因此,在将因果关系结构记录组建成

因果关系结构数据集时,需要补位,补位的值是 0,通过补位将统一因果关系结构记录的维度,还要将因果关系结构记录中的数对转化为值 1.我们规定在因果关系结构记录中两部分值出现的顺序依次是非时滞因果关系变量值和时滞因果关系变量值(也可以采用其它的方式).在非时滞因果关系变量值部分,值的排列顺序是 i 依次取

$1, 2, \dots, n - 1$ 且 $j = i + 1, i + 2, \dots, n$; 而时滞因果关系变量值部分,值的排列顺序是 i 依次取

$1, 2, \dots, n$ 且 h 值由大到小.虽然因果关系结构数据集中可能有许多变量,但其中包含大量的冗余变量,在建

立元因果关系结构时,可以不考虑这些冗余变量(通过变量选择去除冗余变量).在因果关系结构数据集中,如果一个变量的取值不发生变化(或者没有统计意义的变化),便可将其作为冗余变量,当然,随着时间的变化,冗余变量也可能转化为非冗余变量.

2.3 元因果关系的结构与参数学习

在因果关系结构数据集构建和变量选择的基础上,同样采用变量删除法来排序因果关系变量,并通过局部贪婪打分-搜索来学习元因果关系结构,分别使用最大似然和最大后验估计方法进行元因果关系的初始参数学

习和增量参数学习,具体的元因果关系学习算法如算法2 所示.

算法 2. 元因果关系学习算法.

输入: 时间序列数据集 $D_c[n, T]$

输出: 元因果关系(包括结构和参数)

时间序列的丢失数据处理

时间序列的离散化

时间序列分段

for $m=1$ to Γ

 时间序列段的时序变量排序

 时间序列段的因果关系结构学习

 确定时滞变量

 多变量时间序列段的部分转换数据集构建

 多变量时间序列段的因果关系结构学习

 建立时间序列段的因果关系结构记录

end for

建立时间序列的因果关系结构数据集

去除因果关系结构数据集中的冗余变量

因果关系结构数据集中的因果关系变量排序

基于局部贪婪打分-搜索的元因果关系结构学习

if 参数学习是初始学习 then

 结合元因果关系结构和最大似然估计方法进行参数初始学习

else

 结合元因果关系结构和最大后验方法进行参数增量学习

end if

元因果关系学习算法的主要部分是元因果关系结构学习,在元因果关系结构学习中,因果关系结构数据集构建的时间复杂度是 $O(n^2\Gamma)$, 因果关系结构数据集中最多具有 $\frac{n(n-1)}{2}$ 个三值变量和

$nq_1 + (n-1)q_2 + \dots + 2q_{n-1} + q_n$ 个二值变量,其中 $q_i (1 \leq i \leq n)$ 是与 n 和 Γ 无关的量,因此,相对于

MDL打分计算,元因果关系学习算法的时间复杂度是 $O(n^2(\Gamma + n^2))$.

如果把时间序列变量之间的因果关系看作概念,那么,随着时间的演进而导致的因果关系变化就是概念漂移(或因果关系漂移)。本文假设因果关系变化(或漂移)具有随机性和规律性,因此,可用随机变量(因果关系变量)描述因果关系的变化(或漂移),并可建立描述因果关系变化(或漂移)的贝叶斯网络,即元贝叶斯网络。因果关系变化(或漂移)可以发生在非时滞变量之间,也可以发生在时滞变量与非时滞变量之间,通过元贝叶斯网络研究可以揭示因果关系变化(或漂移)的规律,并进行未来演化态势判断和预测。

3 实验与分析

分别使用模拟数据、UCI 数据和wind 数据库中的金融数据,从模拟时间序列的因果关系与元因果关系的结构学习、真实时间序列的因果关系与元因果关系的结构学习,以及离散化方法对时序因果关系的影响三个方面进行实验与分析。我们采用滑动平均的方法初始化丢失的数据(取 $\omega = 10$),并结合最大似然树和Gibbs 抽样

来迭代修复常规时间序列的丢失数据.

3.1 模拟时间序列的因果关系与元因果关系的结构学习

结合给定的因果关系、元因果关系和Monte Carlo 方法产生时间序列模拟数据集,在此基础上,进行因果关系和元因果关系的结构学习.

(1) 因果关系结构学习

分别从侧重时滞父结点集(用 T_BN_1 和 T_BN_2 表示,总的弧数量是 55 和 79)、非时滞父结点集(用 NT_BN_1 和 NT_BN_2 表示,总的弧数量是 42 和 54)和混合父结点集(用 H_BN_1 和 H_BN_2 表示,总的弧数量是 38 和 65),以及均衡三种父结点集(用 E_BN_1 和 E_BN_2 表示,总的弧数量是 30 和 58)四种情况设计因果关系结构(有向无环图),取 $\max \{q_1, q_2, \dots, q_n\} = 3$,根据变量的顺序和弧的数量随机产生弧的位置,也随机产生因果关系参数(条件概率分布表),每种情况给出两个因果关系网络,每一个因果关系网络产生 5000 个模拟数据.首先进行单变量时间序列的时滞因果关系结构学习,学习得到的因果关系结构与产生模拟数据的因果关系结构完全一致,根据学习的结果确定时滞变量,然后建立模拟数据集的转换数据集,最后结合变量顺序和局部贪婪打分-搜索方法进行多变量时间序列的因果关系结构学习,具体情况如表 1 所示.表 1 中第一列括号中的数字是因果关系结构中弧的数量,最后一列括号中的数字是变化弧的数量占所有弧数量的百分比,“+”表示增加的弧,“-”表示丢失的弧.只给出均衡三种父结点集情况的一个因果关系结构,如图 3 所示.

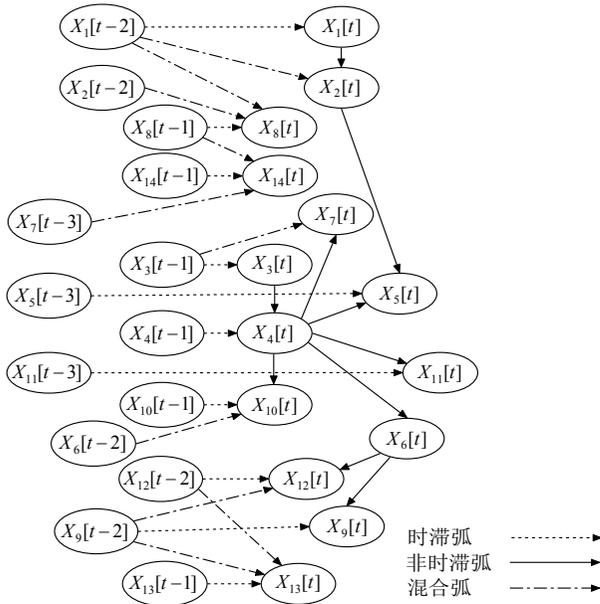


Fig.3 Causal structure of balancing three parent nodes

图 3 均衡三种父结点的因果关系结构(E_BN_1)

Table 1 Causal structure learning for simulating time series

表 1 模拟时间序列的因果关系结构学习

时间序列数据集	时滞弧变化数量	非时滞弧变化数量	混合弧变化数量	总变化数量(百分比)
T_BN_1(45)	0	0	-1	1(2.22%)
T_BN_2(54)	0	-1	-1	2(3.70%)
NT_BN_1(48)	-1	0	-1	2(4.16%)
NT_BN_2(51)	0	-1	-1	2(3.92%)
H_BN_1(57)	0	-1	+1	2(3.51%)

H_BN_2(60)	-1	-1	-1	3(5.00%)
E_BN_1(30)	0	-1	0	1(3.33%)
E_BN_2(42)	0	-1	-1	2(4.76%)

表 1 显示了模拟时序数据因果关系结构学习的情况,没有正确识别弧(主要是丢失的弧)的平均百分比是 3.82%。出现这些丢失弧的主要原因是:随机产生的概率分布会出现小概率的情况,从而导致变量之间的依赖和条件独立性的判断出现差错。总的来看,对于模拟时间序列数据,能够有效识别所蕴含的三种父结点(时滞父结点、非时滞父结点和混合父结点),因此,可以为真实数据的因果关系学习提供保障。

(2) 元因果关系结构学习

对给定的元因果关系,首先基于元因果关系产生模拟数据,然后进行元因果关系结构学习,并比较产生模拟数据的元因果关系结构和学习得到的元因果关系结构之间的差异。

① 元因果关系的随机模拟

元因果关系的随机模拟包括两个阶段:第一阶段是结合因果关系变量的顺序、元因果关系结构、因果关系变量的概率分布(随机生成概率分布表)和 Monte Carlo 方法产生元因果关系的直接模拟数据集(因果关系结构数据集);第二阶段是将时序变量的顺序、元因果关系模拟数据集中的记录(或因果关系结构记录,每个记录都对应着一个因果关系结构)、时间序列变量的概率分布(随机生成概率分布表)和 Monte Carlo 方法相结合生成因果关系的模拟数据集(时间序列数据集)。设置元因果关系模拟数据集的大小(或时间序列段的数量)为 $\Gamma = 5000$,所有时间序列段的大小也均为 $T_m - T_{m-1} = 5000, n = 6$,最终得到元因果关系的随机模拟时间序列大小为 25000000。

(a) 时间序列的因果关系结构

首先需要给出产生第一段时间序列的因果关系结构(也可称为初始因果关系结构),在为变量拓扑排序时,先排序非时滞变量(设为 $X_1[t], X_2[t], X_3[t], X_4[t], X_5[t], X_6[t]$),后排序时滞变量(如 $X_i[t-4], X_i[t-3], X_i[t-2], X_i[t-1], X_i[t]$)。

初始因果关系结构:

$$X_1[t] \leftarrow X_1[t-4], X_1[t-2], X_2[t] \leftarrow X_1[t], X_2[t-2], X_3[t] \leftarrow X_1[t], X_3[t-3], X_3[t-1],$$

$$X_4[t] \leftarrow X_2[t], X_3[t], X_4[t-1], X_5[t] \leftarrow X_2[t], X_4[t], X_5[t-3], X_5[t-2],$$

$$X_6[t] \leftarrow X_3[t], X_5[t], X_6[t-4], X_6[t-3].$$

由初始因果关系结构得到的因果关系结构记录为:

$$(1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, (4, 1), (2, 1), (2, 2), (3, 3), (1, 3), (1, 4), (3, 5), (2, 5), (4, 6), (3, 6)).$$

(b) 时间序列的元因果关系结构

用 $(X_i[t-u], X_j[t-v])$ 表示元因果关系网络中的变量, $(X_i[t-u], X_j[t-v]) \leftarrow (X_h[t-w], X_k[t-z])$ 表示元因果关系网络中的因果关系。

时间序列的元因果关系结构:

$$(X_1[t], X_1[t-4]) \leftarrow (X_1[t], X_1[t-2]), (X_3[t], X_3[t-3]) \leftarrow (X_3[t], X_3[t-1]),$$

$$\begin{aligned} & (X_1[t], X_4[t]) \leftarrow (X_1[t], X_2[t]), (X_1[t], X_3[t]), (X_3[t], X_6[t]) \leftarrow (X_1[t], X_4[t]), (X_1[t], X_3[t]), \\ & (X_2[t], X_5[t]) \leftarrow (X_2[t], X_4[t]), (X_4[t], X_4[t-1]), \\ & (X_5[t], X_6[t]) \leftarrow (X_4[t], X_5[t]), (X_5[t], X_5[t-2]). \end{aligned}$$

结合元因果关系和 Monte Carlo 方法产生模拟时间序列数据用于元因果关系结构学习。

(2) 元因果关系结构学习

元因果关系学习是元因果关系随机模拟的逆过程。在产生元因果关系的模拟时间序列数据集的基础上,首先进行时间序列段的因果关系结构学习,然后建立因果关系结构数据集,并对因果关系结构数据集进行变量选择,基于因果关系结构数据集建立元因果关系结构,通过检验发现学习得到的时间序列元因果关系结构与产生模拟数据的元因果关系结构完全一致。

3.2 真实时间序列的因果关系与元因果关系的结构学习

在wind 数据库中选择股票、期货和汇率时间序列用于因果关系结构学习,使用UCI 时间序列进行元因果关系结构学习。

(1) 股票、期货和汇率时间序列的因果关系结构学习

在wind 数据库中,选择上海银行板块10 支股票、上海10 个贵金属期货和香港10 种汇率,对股票、期货和汇率时间序列进行丢失数据修复、三值增减性离散化和结合专家知识与变量删除法的排序,排序后的股票顺序是:农业银行,宁波银行,兴业银行,平安银行,民生银行,南京银行,中信银行,招商银行,华夏银行和光大银行;期货顺序是:钢,铀,锌,铁,铂,钯,铝,银,铜和金;汇率顺序是:港币兑美元,港币兑英镑,港币兑瑞士法郎,港币兑瑞典克朗,港币兑加元,港币兑欧元,港币兑澳大利亚元,港币兑日元,港币兑挪威克朗和港币兑新西兰元。使用 $X_1[t], X_2[t], \dots, X_{10}[t]$ 来统一表示排序后的变量,股票、期货和汇率时间序列均不存在时滞因果关系,学习得到的非时滞因果关系如表2所示。

Table 2 Non delay causality learning for stock, futures and exchange rate time series

表 2 股票、期货和汇率时间序列的非时滞因果关系学习

变量	股票因果关系	期货因果关系	汇率因果关系
$X_2[t]$	$X_2[t] \leftarrow X_1[t]$	$X_2[t] \leftarrow X_1[t]$	$X_2[t] \leftarrow X_1[t]$
$X_3[t]$	$X_3[t] \leftarrow X_2[t], X_1[t]$	$X_3[t] \leftarrow X_2[t], X_1[t]$	$X_3[t] \leftarrow X_2[t]$
$X_4[t]$	$X_4[t] \leftarrow X_3[t], X_2[t], X_1[t]$	$X_4[t] \leftarrow X_3[t], X_2[t]$	$X_4[t] \leftarrow X_3[t], X_2[t]$
$X_5[t]$	$X_5[t] \leftarrow X_4[t], X_3[t], X_1[t]$	$X_5[t] \leftarrow X_4[t], X_3[t]$	$X_5[t] \leftarrow X_4[t], X_2[t]$
$X_6[t]$	$X_6[t] \leftarrow X_4[t], X_3[t], X_2[t]$	$X_6[t] \leftarrow X_5[t], X_3[t]$	$X_6[t] \leftarrow X_5[t], X_4[t]$
$X_7[t]$	$X_7[t] \leftarrow X_6[t], X_5[t], X_2[t], X_1[t]$	$X_7[t] \leftarrow X_6[t], X_5[t], X_3[t]$	$X_7[t] \leftarrow X_6[t], X_5[t]$
$X_8[t]$	$X_8[t] \leftarrow X_6[t], X_5[t], X_3[t]$	$X_8[t] \leftarrow X_7[t], X_6[t], X_5[t]$	$X_8[t] \leftarrow X_3[t]$
$X_9[t]$	$X_9[t] \leftarrow X_8[t], X_6[t], X_4[t], X_3[t]$	$X_9[t] \leftarrow X_8[t], X_7[t], X_3[t]$	$X_9[t] \leftarrow X_6[t], X_5[t], X_4[t]$
$X_{10}[t]$	$X_{10}[t] \leftarrow X_9[t], X_7[t], X_1[t]$	$X_{10}[t] \leftarrow X_8[t], X_5[t], X_3[t]$	$X_{10}[t] \leftarrow X_9[t], X_5[t], X_4[t]$

虽然股票、期货和汇率时间序列不存在时滞因果关系,但在表2 中我们能够发现,存在比较复杂的非时滞因果关系,也就是不同的时间序列之间存在较强的影响与制约关系,尤其是同一板块的股票之间可能存在非常复杂的因果关系。通过与金融领域专家的交流,表2 显示的股票、期货和汇率的因果关系较好地反映了经验中的因果联系,同时还展示了许多专家经验中所没有,而却在数据中蕴含的因果关系,切实起到了客观因果知识对主观因果经验的修正和补充的作用,因此,有助于深化对股票、期货和汇率客观规律的认识,并提高管理决策的科学化水平。

(2) UCI 时间序列的元因果关系结构学习

选择四个 UCI 时间序列数据集,它们是 GSA(Gas sensor array under dynamic gas mixtures,具有 4178504 个

记录和 19 个时间序列)、URLR(URL reputation,具有 2396130 个记录和 3231961 个时间序列)、PAM(Physical activity monitoring,具有 3850505 个记录和 52 个时间序列)和 IHEPC(Individual household electric power consumption,具有 2075259 个记录和 9 个时间序列),从它们中各选择 6 个时间序列,结合滑动平均与二值增减性离散化进行时间序列预处理,基于变量删除法排序时序变量,用 $X_1[t], X_2[t], X_3[t], X_4[t], X_5[t]$ 和 $X_6[t]$ 表示排序后的时序变量,取时间序列段(等大小的时间序列段)的数量 $\Gamma = T_m - T_{m-1} = 2044, 1547, 1962, 1440$,学习得到的多变量时间序列元因果关系结构如下:

GSA: $(X_1[t], X_2[t]) \leftarrow (X_1[t], X_3[t])$;URLR: $(X_2[t], X_2[t-3]) \leftarrow (X_2[t], X_2[t-2])$;PAM:
 ① $(X_4[t], X_5[t]) \leftarrow (X_1[t], X_4[t])$,② $(X_5[t], X_5[t-3]) \leftarrow (X_5[t], X_5[t-1])$;IHEPC:
 ① $(X_3[t], X_6[t]) \leftarrow (X_2[t], X_6[t]), (X_1[t], X_6[t]),$ ② $(X_6[t], X_6[t-4]) \leftarrow (X_6[t], X_6[t-3])$.

在四个多变量时序数据集中,存在三种因果关系:非时滞、时滞和混合因果关系.我们发现有两种因果关系之间存在因果影响,它们是非时滞因果关系之间的因果影响(因果关系 GSA, PAM①和 IHEPC①)和时滞因果关系之间的因果影响(因果关系 URLR, PAM②和 IHEPC②).

3.3 离散化方法对时序因果关系的影响

离散化方法对时序数据的因果关系会产生影响,下面从离散化(包括模式离散化和结构离散化)方法对非时滞因果关系和时滞因果关系两个方面的影响进行实验与分析.

(1) 离散化方法对非时滞因果关系的影响

在 wind 数据库中,选择 16 个基金时间序列,为便于比较,统一使用三值增减性离散化后的变量顺序,排序后的时间序列依次是融通创业 A ($X_1[t]$)、华商稳 B ($X_2[t]$)、新蓝筹($X_3[t]$)、融通蓝筹($X_4[t]$)、融通行业($X_5[t]$)、融通动力($X_6[t]$)、融通内需($X_7[t]$)、长城久恒($X_8[t]$)、长城回报($X_9[t]$)、长城品牌($X_{10}[t]$)、长城债券($X_{11}[t]$)、长城动力($X_{12}[t]$)、长城积 A ($X_{13}[t]$)、长城优化($X_{14}[t]$)、泰信先行($X_{15}[t]$)和泰信优质($X_{16}[t]$).分别使用模式离散化和结构离散化两种方法来离散化基金时间序列,其中模式离散化方法均取三个离散值,包括增减性离散化(简称为增减)、转折性离散化(简称为转折)、波动性离散化(简称为波动)和突变性离散化(简称为突变),而结构离散化方法也都取三个离散值,它们是等距离散化(简称为等距)、等频离散化(简称为等频)、基于密度的离散化(简称为密度,采用高斯密度)和基于聚类的离散化(简称为聚类,使用贝叶斯聚类),采用不同离散化方法学习得到的非时滞因果关系情况如表 3 所示,其中表示因果关系的数字是对应变量的下标.

Table 3 Influence of different discretization methods on non delay causality

表 3 不同离散化方法对非时滞因果关系的影响

变量	增减	转折	波动	突变	等距	等频	密度	聚类
$X_2[t]$	2←1	2←1	2←1	2←1	2←1	2←1	2←1	2←1
$X_3[t]$	3←2,1	3←2,1	3←2,1	3←2,1	3←2,1	3←2,1	3←2,1	3←2,1
$X_4[t]$	4←3,2,1	4←3,2,1	4←3,2,1	4←3,1	4←3,2,1	4←3,1	4←3,2,1	4←3,1
$X_5[t]$	5←4,3,1	5←4,3,1	5←4,3,1	5←4,3	5←4,3,1	5←4,3,1	5←4,3,1	5←4,3,1
$X_6[t]$	6←4,3,1	6←4,3,1	6←4,3,1	6←4,3,1	6←4,3,1	6←5,4,1	6←5,4,1	6←4,3,1
$X_7[t]$	7←5,4,1	7←5,4,1	7←4,2,1	7←6,4,1	7←6,3,1	7←4,3,1	7←4,3,1	7←4,3,1
$X_8[t]$	8←4,1	8←4,1	8←6,4,1	8←5,4,1	8←7,5,1	8←5,4,1	8←5,1	8←5,1
$X_9[t]$	9←8,4,3	9←7,4,3	9←6,4,3	9←6,4,3	9←8,4,3	9←6,4,3	9←6,4,3	9←6,4,3
$X_{10}[t]$	10←9,3,2	10←9,3	10←9,3	10←9,3	10←9,3	10←9,3,2	10←9,3,2	10←9,3
$X_{11}[t]$	11←8	11←8	11←10,8	11←8	11←8,3	11←8,2	11←8	11←8
$X_{12}[t]$	12←9,8,1	12←9,8,1	12←9,8,1	12←9,6,1	12←9,8,1	12←9,4,1	12←9,4,1	12←9,4,1

$X_{13}[t]$	13←11,10	13←11	13←11,5	13←11	13←11,2	13←11,2	13←11,2	13←11,2
$X_{14}[t]$	14←12,10,4,1	14←12,7,4	14←12,7,4	14←12,4,3,1	14←12,6,5	14←12,5,4	14←12,5,4	14←12,7,5
$X_{15}[t]$	15←12,3,1	15←12,3,1	15←12,3,1	15←12,3,1	15←12,6,1	15←12,4,1	15←12,4,1	15←12,6,1
$X_{16}[t]$	16←15,14,6	16←15,6,1	16←15,14,6	16←15,7,6	16←15,7,1	16←15,4,1	16←15,4,1	16←15,7

从表 3 我们能够发现,采用不同的离散化方法使学习得到的非时滞因果关系具有一定的差距,其主要原因是:不同的离散化方法会产生相应的联合分布,从而导致由联合分布所确定的因果关系具有差异.不同的离散化方法也使学习得到的因果关系具有不同的含义,如增减性因果关系、转折性因果关系和波动性因果关系等.对于模式离散化,所有因果关系数量是 150,共同因果关系数量是 116,因果关系不变率是 77.33%;对于结构离散化,所有因果关系数量是 151,共同因果关系数量是 112,因果关系不变率是 74.17%;对于模式与结构离散化两种情况,所有因果关系数量是 301,共同因果关系数量是 184,因果关系不变率是 61.12%.由表 6 得到的模式和结构离散化的最小(因果关系的交集)与最大因果关系(因果关系的并集)情况如表 4 所示.

Table 4 Minimum and maximum causality of different discretization methods

表 4 不同离散化方法的最小与最大因果关系

变量	模式离散化		结构离散化		模式与结构离散化	
	最小因果关	最大因果关系	最小因果	最大因果关系	最小因果	最大因果关系
$X_2[t]$	2←1	2←1	2←1	2←1	2←1	2←1
$X_3[t]$	3←2,1	3←2,1	3←2,1	3←2,1	3←2,1	3←2,1
$X_4[t]$	4←3,1	4←3,2,1	4←3,1	4←3,1	4←3,1	4←3,2,1
$X_5[t]$	5←4,3	5←4,3,1	5←4,3,1	5←4,3,1	5←4,3	5←4,3,1
$X_6[t]$	6←4,3,1	6←4,3,1	6←4,1	6←5,4,3,1	6←4,1	6←5,4,3,1
$X_7[t]$	7←4,1	7←6,5,4,1	7←3,1	7←6,4,3,1	7←1	7←6,5,4,3,1
$X_8[t]$	8←4,1	8←6,5,4,1	8←5,1	8←7,5,4,1	8←1	8←7,6,5,4,1
$X_9[t]$	9←4,3	9←8,7,6,4,3	9←4,3	9←8,6,4,3	9←4,3	9←8,7,6,4,3
$X_{10}[t]$	10←9,3	10←9,3,2	10←9,3	10←9,3,2	10←9,3	10←9,3,2
$X_{11}[t]$	11←8	11←10,8	11←8	11←8,3,2	11←8	11←10,8,3,2
$X_{12}[t]$	12←9,1	12←9,8,6,1	12←9,1	12←9,8,4,1	12←9,1	12←9,8,6,4,1
$X_{13}[t]$	13←11	13←11,10,5	13←11,2	13←11,2	13←11	13←11,10,5,2
$X_{14}[t]$	14←12,4	14←12,10,7,4,3,1	14←12,5	14←12,7,6,5,4	14←12	14←12,10,7,6,5,4,3,1
$X_{15}[t]$	15←12,3,1	15←12,3,1	15←12,1	15←12,6,4,1	15←12,1	15←12,6,4,3,1
$X_{16}[t]$	16←15,6	15,14,7,6,1	16←15	16←15,7,4,1	16←15	15,14,7,6,4,1

虽然不同的离散化方法使学习得到的因果关系具有不同的实际含义,但有一些因果关系比较稳定,它们不易受到离散化方法的影响,我们可以将这些因果关系称为通用因果关系,使用这些因果关系也可以进行新的离散化方法因果关系学习的推断,相对于模式与结构离散化方法,学习得到的稳定因果关系网络如图 4 所示.

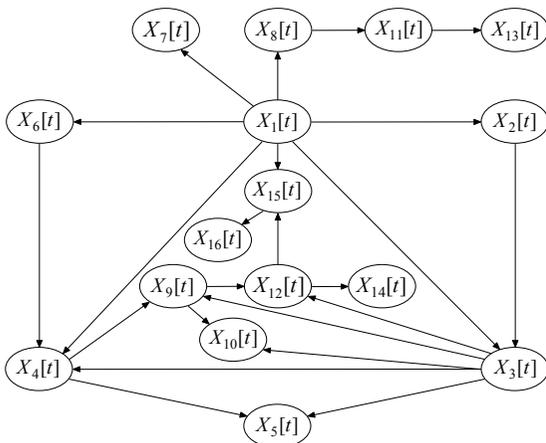


Fig.4 Stable causal network

图 4 稳定的因果关系网络

(2) 离散化方法对时滞因果关系的影响

同样使用表 4 的时间序列数据集和离散化方法进行时滞因果关系学习,学习得到的时滞因果关系情况如表 5 所示.为表述上的方便,用 $t \leftarrow$ 表示不含有因果关系, $t \leftarrow t - i_1, t - i_2, \dots, t - i_k$ 表示因果关系

$$X_j[t] \leftarrow X_j[t - i_1], X_j[t - i_2], \dots, X_j[t - i_k].$$

Table 5 Influence of different discretization methods on delay causality

表 5 不同离散化方法对时滞因果关系的影响

变量	增减	转折	波动	突变	等距	等频	密度	聚类
$X_1[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow$	$t \leftarrow t-1$	$t \leftarrow$	$t \leftarrow t-7$
$X_2[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow$	$t \leftarrow t-2$	$t \leftarrow t-1, t-2$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$
$X_3[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1, t-3$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$
$X_4[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow t-1$	$t \leftarrow$
$X_5[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow$	$t \leftarrow t-1, t-2$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$
$X_6[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1, t-8$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow t-9$
$X_7[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-2$	$t \leftarrow t-1, t-4$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow t-7$
$X_8[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow t-1, t-4, t-9$	$t \leftarrow t-1, t-8$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow t-1, t-8$
$X_9[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-2$	$t \leftarrow t-1, t-3$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow t-5$
$X_{10}[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$
$X_{11}[t]$	$t \leftarrow t-1$	$t \leftarrow t-1$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow t-1, t-4$	$t \leftarrow t-2$	$t \leftarrow$	$t \leftarrow t-1$
$X_{12}[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow t-7$
$X_{13}[t]$	$t \leftarrow t-1$	$t \leftarrow t-1$	$t \leftarrow t-1$	$t \leftarrow t-1, t-6$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$
$X_{14}[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1, t-2$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$
$X_{15}[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1, t-6$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow t-3$
$X_{16}[t]$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1$	$t \leftarrow t-1, t-2$	$t \leftarrow t-1, t-7$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$	$t \leftarrow$

表5显示,离散化方法对时滞因果关系影响较大,尤其是结构离散化方法,使学习得到的因果关系基本没有实际意义(离散化没有考虑时序变化因素),在模式离散化中,因果关系的变化也比较大,从模式离散化的整体来看,只有 $t \leftarrow t-1$ 相对比较稳定,因此,对于新的模式离散化方法,除 $t \leftarrow t-1$ 之外不宜做其它的因果关系推断.

4 结论和进一步的工作

本文通过时间序列预处理、转换数据集构建、时序变量的排序、局部贪婪打分-搜索和最大似然参数估计等给出了时间序列的因果关系学习方法,在此基础上,再将时间序列段的因果关系结构学习、因果关系结构数据集构建、因果关系变量的排序、局部贪婪打分-搜索、最大似然参数估计和最大后验参数估计等相结合建立了元因果关系学习方法,从而可实现时间序列数据集两个层次的因果关系学习.

在转换数据集中,把纵向时滞依赖转化为横向非时滞依赖,这样便可以忽略记录之间存在的时序影响,从而实现了时滞与非时滞信息的整合与统一,使基于转换数据集能够有效地进行时滞、非时滞和混合因果关系学习,以及因果推理计算;结合滑动平均初始化、最大似然树与Gibbs 抽样的丢失数据迭代修复,可使学习得到的因果关系更加可靠;基于时间序列段模式的离散化,使学习得到的因果关系更具有实际意义、可理解性和针对性;将专家经验与变量删除法相结合的时序变量排序,一方面实现了主观与客观知识的结合与互补,并可避免现有的只使用专家知识(变量多时排序困难)或数据(结果不便于解释)排序变量所带来的问题,而在变量排序基础上的因果关系学习也具有更高的效率和可靠性;通过因果关系结构数据集的构建,时间序列段的因果关系结构被转化为数据集中的记录,因果关系变成随机变量,从而使建立在因果关系结构数据集之上的元因果关系能够进行因果关系的不确定性研究.

进一步的工作是提高因果关系和元因果关系的学习效率与可靠性,在因果关系结构数据集中考虑混合因

果关系变量,为因果关系变量引入时滞因素,并将因果关系用于低频宏观经济指标时间序列的定性与定量因果分析,以及将元因果关系用于高频金融指标时间序列的因果关系演化分析和未来发展趋势判断等。

References:

- [1] Ferreira C. Debt and Economic Growth in the European Union: A Panel Granger Causality Approach. *International Advances in Economic Research*, 2016, 22(2): 131-149
- [2] Fredrik A, Katarzyna B, Sonja O. Lending for growth? A Granger causality analysis of China's finance-growth nexus. *Empirical Economics*, 2016, 51(3): 897-920
- [3] Chang T Y, Deale D, Gupta R et al. The causal relationship between coal consumption and economic growth in the BRICS countries: Evidence from panel-Granger causality tests. *Energy Sources, Part B: Economics, Planning, and Policy*, 2017, 12(2): 138-146
- [4] Kristofer M, Ghazi S, Pär S. A new ridge regression causality test in the presence of multicollinearity. *Communications in Statistics: Theory and Methods*, 2014, 43(2): 235-248
- [5] David C, David S L, Zhuan P et al. Inference on causal effects in a generalized regression kink design. *Econometrica*, 2015, 83(6): 2453-2483
- [6] Ryutah K, Yuya S. On using linear quantile regressions for causal inference. *Econometric Theory*, 2017, 33(3): 664-690
- [7] Luo W, Zhu Y Y, Ghosh D. On estimating regression-based causal effects using sufficient dimension reduction. *Biometrika*, 2017, 104(1): 51-65
- [8] Rubin D. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, 1978, 6(1): 34-58
- [9] Pearl J. Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Mateo, California: Morgan Kaufmann, 1988, 383-408
- [10] Heckman J J. The scientific model of causality. *Sociological Methodology*, 2005, 35(1): 1-97
- [11] John A, Rafael L. Counterfactuals and Causal Inference: Methods and Principles for Social Research By Stephen L. Morgan & Christopher Winship. *Structural Equation Modeling*, 2011, 18(1): 152-159
- [12] Kullback C, Weiss S. Representation of expert knowledge for consultation: The CASNET and EXPERT projects. *Artificial Intelligence in medicine*, 1982, 1(4): 21-55
- [13] Cheng J, Greiner R, Kelly J. Learning Bayesian networks from data: An efficient approach based on information-theory. *Artificial Intelligence*. 2002, 137 (1-2): 43-90
- [14] Liu X Q, Liu X S. Swamping and masking in Markov boundary discovery. *Machine Learning*, 2016, 104(1): 25-54
- [15] Parviainen P, Kaski S. Learning structures of Bayesian networks for variable groups. *International Journal of Approximate Reasoning*, 2017, 88(5): 110-127
- [16] Xiao C, Jin Y, Liu J et al. Optimal Expert Knowledge Elicitation for Bayesian Network Structure Identification. *IEEE Transactions on Automation Science & Engineering*, 2018, PP(99): 1-15
- [17] Heckerman D, Geiger D, Chickering D M. Learning Bayesian networks: the combination of knowledge and statistical data, *Machine Learning*, 1997, 20(3): 197-243
- [18] Suzuki J. A theoretical analysis of the BDeu scores in Bayesian network structure learning. *Behaviormetrika*, 2016, 1(1): 1-20
- [19] Gheisari S, Meybodi M R, Dehghan M et al. Bayesian network structure training based on a game of learning automata. *International Journal of Machine Learning and Cybernetics*, 2017, 8(4): 1093-1105
- [20] Liu X, Liu X. Structure learning of Bayesian networks by continuous particle swarm optimization algorithms. *Journal of Statistical Computation & Simulation*, 2018, 88(9): 1-29
- [21] Friedman N, Murphy K P, Russell S. Learning the structure of dynamic probabilistic networks. In: *Proceedings of the 14th International Conference on Uncertainty in Artificial Intelligence*, Madison, USA, 1998, 139-147
- [22] Wang Shuang-Cheng, Leng Cui-Ping, Li Xiao-Lin. Learning Bayesian network structure from small data set. *Acta Automatica Sinica*, 2009, 35(8): 1063-1070 (in Chinese)

[23] Lähdesmäki H, Shmulevich I. Learning the structure of dynamic Bayesian networks from time series and steady state measurements. *Machine Learning*, 2008, 71(2-3): 185-217

附中文参考文献:

[22] 王双成, 冷翠平, 李小琳. 小数据集中的贝叶斯网络结构学习. *自动化学报*, 2009, 35(8): 1063-1070