



基于级联加性噪声模型的因果结构学习算法

乔杰¹, 蔡瑞初¹, 郝志峰²

(1. 广东工业大学 计算机学院, 广州 510006; 2. 佛山科学技术学院 数学与大数据学院, 广东 佛山 528000)

摘要: 现有级联非线性加性噪声模型可解决隐藏中间变量的因果方向推断问题, 然而对于包含隐变量和级联传递因果关系的因果网络学习存在全局结构搜索、等价类无法识别等问题。设计一种面向非时序观测数据的两阶段因果结构学习算法, 第一阶段根据观测数据变量间的条件独立性, 构建基本的因果网络骨架, 第二阶段基于级联非线性加性噪声模型, 通过比较骨架中每个相邻因果对在不同因果方向假设下的边缘似然度进行因果方向推断。实验结果表明, 该算法在虚拟因果结构数据集的不同隐变量数量、平均入度、结构维度、样本数量下均表现突出, 且在真实因果结构数据集中的F1值相比主流因果结构学习算法平均提升了51%, 具有更高的准确率和更强的鲁棒性。

关键词: 因果结构学习; 加性噪声模型; 级联加性噪声模型; 因果发现; 函数式因果模型

开放科学(资源服务)标志码(OSID):



中文引用格式: 乔杰, 蔡瑞初, 郝志峰. 基于级联加性噪声模型的因果结构学习算法[J]. 计算机工程, 2022, 48(1): 93-98.

英文引用格式: QIAO J, CAI R C, HAO Z F. Causal structure learning algorithm based on cascade additive noise model[J]. Computer Engineering, 2022, 48(1): 93-98.

Causal Structure Learning Algorithm Based on Cascade Additive Noise Model

QIAO Jie¹, CAI Ruichu¹, HAO Zhifeng²

(1. School of Computer, Guangdong University of Technology, Guangzhou 510006, China;

2. School of Mathematics and Big Data, Foshan University, Foshan, Guangdong 528000, China)

[Abstract] The existing cascade nonlinear Additive Noise Model (ANM) can infer the causal direction of hidden intermediate variables, but fail to deal with global structure search and equivalence class recognition in the case of causal network learning that includes hidden variables and cascade causality transferring. This paper presents a two-stage causal structure learning algorithm for non-chronological observation data. In the first stage, a basic causal network skeleton is constructed based on the conditional independence between the observation data variables. In the second stage, by using a cascaded nonlinear ANM, the causal direction of the edge likelihood under the assumptions of different causal directions is inferred by comparing each adjacent causality in the skeleton. The experimental results show that the algorithm has outstanding performance on the virtual causal structure dataset for a varying number of hidden variables, average in-degree, structural dimension, and number of samples. Furthermore, the F1 value of this algorithm on the real causal structure dataset improved by 51% on average compared with mainstream causal structure learning algorithms, displaying a higher accuracy and robustness.

[Key words] causal structure learning; Additive Noise Model (ANM); Cascade Additive Noise Model (CANM); causal discovery; functional causal model

DOI: 10.19678/j.issn.1000-3428.0060176

0 概述

因果网络是一种描述多维数据间因果关系的网络^[1-3], 在地球系统学^[4-5]、生物基因学^[6]等领域均具有广泛应用。随机试验是因果关系推断的黄金标准^[7], 然而通常存在实现成本高、实际操作困难等问

题。因此, 如何从观测数据中推断出因果结构是一个重要且具有挑战性的课题。目前, 研究人员已提出一系列从数据中发现因果关系的方法^[8], 但仅能观测到整个因果结构的部分变量, 还有研究人员针对存在隐变量的因果结构学习提出FCI^[1]、RF-CI^[9]、GF-CI^[10]等算法, 然而这类算法具有大量不

基金项目: 国家自然科学基金(61876043, 61976052)。

作者简介: 乔杰(1993—), 男, 博士研究生, 主研方向为数据挖掘、机器学习; 蔡瑞初(通信作者), 郝志峰, 教授、博士生导师。

收稿日期: 2020-12-03 修回日期: 2021-01-23 E-mail: cairuichu@gmail.com

可识别的因果边。对于某个不存在 v-结构^[6]的因果对,如果其结果变量无法观测,只能观测到结果变量的子代,那么就存在一条级联式传递的因果关系,而这样的因果关系通常会破坏非线性函数式因果模型的可识别性。CAI等^[11]在2019年提出级联加性噪声模型(Cascade Additive Noise Model, CANM),该模型在数据服从非线性加性噪声假设下级联结构仍是可识别的,但只能识别两个结点的因果对,无法学习因果结构。本文针对包含隐变量的级联加性噪声模型,结合基于约束的因果骨架学习算法以及级联函数式因果模型,提出一种混合因果结构学习算法。

1 相关工作

从观测数据中学习因果网络主要包括面向时序序列(如格兰杰因果^[12]、基于神经网络的联合识别方法^[13]等)以及非时序序列两类算法。本文主要针对非时序数据,从非时序观测数据中学习因果网络分为以下两类方法:第一类是基于独立性的方法,包括基于约束^[1]、基于评分^[14]等方法,此类方法存在无法有效区分马尔科夫等价类的问题^[15],导致无法推断部分因果结构;第二类是基于函数的方法,此类方法通常假设原因变量与结果变量间存在某种特定的函数映射关系,使得结果变量是由相互独立的原因变量与噪声变量经过映射得到的。通过判断该噪声与原因变量的独立性,在某些条件下能够推断出因果关系的方向,将这类可以推断因果方向的模型称为可识别模型,典型的函数式因果模型一般只考虑两个结点的因果对关系,主要包括线性非高斯无环模型(Linear Non-Gaussian Acyclic Model, LiNGAM)^[16-17]、非线性加性噪声模型(Additive Noise Model, ANM)^[18]、后非线性(Post-Nonlinear, PNL)因果模型^[19]。

2 基于CANM的因果结构学习框架

定义因果结构为图 $G=\{X, E\}$, 其中, $X=\{X_1, X_2, \dots, X_n\}$ 表示 n 个因果变量结点, $E=\{(i, j)|X_i \rightarrow X_j\}$ 表示结点间的因果边集合, 数据集为 $D=\{x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\}_{i=1}^m$, 样本量大小为 m 。假设 X_i 是 X_j 的父母, 若满足直接因果关系则 $X_i \rightarrow X_j$, 假设 X_i 是 X_j 的祖先, 若存在间接因果边则 $X_i \rightarrow X_k \rightarrow X_j$ 或 $X_i \leftarrow X_k \leftarrow X_j$ 。定义 $X_i \perp_G X_j | S$ 为给定条件集 S , X_i 与 X_j 在图 G 上条件独立。同时, 给出因果忠诚性假设^[1]、因果充分性假设等常见的基本假设。因果忠诚性假设保证了真实数据分布与因果结构独立的一致性, 这一假设使得通过条件独立性进行因果骨架学习成为可能。因果充分性假设保证了任意两个可观测的因果结点不存在一个共同的隐变量父亲根结点, 这一假设使得在隐变量下利用级联因果模型进行方向推断成为可能。

本文使用一种典型因果结构^[20]来表达隐因果结构网络。简而言之, 对于因果路径, 若 $X_i \rightarrow X_k \rightarrow X_j$ 且 X_k 是隐变量, 则在典型因果结构上认为 $X_i \rightarrow X_j$, 对于因果路径 $X_i \leftarrow X_k \rightarrow X_j$ 且 X_k 是隐变量, 同时 X_k

没有任何可观测的父亲或祖先结点, 则认为 $X_i \leftrightarrow X_j$, 反复迭代直到结构不再变化。在因果充分性假设下, 典型因果结构并不存在形如 $X_i \leftrightarrow X_j$ 的双向边。

本文的目标是学习典型因果结构网络。图1给出了基于级联非线性加性噪声模型的因果结构学习算法SCANM框架。该框架分为两个阶段: 第一阶段是将输入的观测数据通过因果骨架来学习因果变量间的骨架; 第二阶段利用级联非线性加性噪声模型, 针对存在隐变量的数据进行因果方向推断。

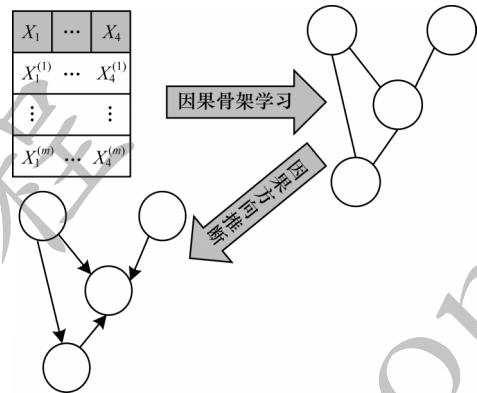


图1 因果结构学习框架

Fig.1 Causal structure learning framework

3 基于CANM的因果结构学习算法

3.1 因果骨架学习

因果骨架学习的基本流程为: 首先初始化一个完全图作为因果结构; 其次使用条件独立性检验对独立边进行删边, 直到无法再删除边为止。

● 算法1 因果骨架学习

输入 样本集 $D = \{x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\}_{i=1}^m$

输出 因果骨架

1. 将 G 初始化为完全图
2. $\text{length} = -1$
3. Repeat
4. $\text{length} = \text{length} + 1$
5. Repeat
6. 选择所有在 G 中相邻且满足条件 $|\text{adj}(G, X_i) \setminus \{X_j\}| \geq \text{length}$ 的边 (X_i, X_j)
7. Repeat
8. 选择所有满足 $|Y| = \text{length}$ 的集合 $Y \subseteq \text{adj}(G, X_i) \setminus \{X_j\}$, 对于每个集合, 如果 X_i 和 X_j 在给定条件 Y 下条件独立, 则在 G 中删除边 (X_i, X_j)
9. Until X_i 和 X_j 之间不再有边, 或者已经遍历所有 Y
10. Until 所有边都被遍历一遍
11. Until 已经找不到满足条件 $|\text{adj}(G, X_i) \setminus \{X_j\}| \geq \text{length}$ 的边
12. Return 因果骨架 G

在算法1中有3层关键的循环: 第1层循环(第3行)的目的是从小到大遍历不同长度的条件集以测试条件独立性, 条件集长度从0开始, 即初始条件集为空集。这么做是因为如果条件越多, 那么条件独立性的判断会越不准确, 所以倾向于从小到大遍历。第2层循环(第5行)的目的是找到每条满足给定条

件集的边, 此外, 对于某条边 (X_i, X_j) 而言, 其条件集应该只会出现在 X_i 的邻居中, 即属于集合 $Y \subseteq \text{adj}(G, X_i) \setminus \{X_j\}$, 通过该方法可以加快遍历条件集的速度。第3层循环(第7行)的目的是遍历所有满足集合长度等于 length 的条件集, 即 $|Y| = \text{length}$, 并使用遍历的条件对结点 X_i 和 X_j 进行独立性检验, 如果发现存在一个条件使得它们独立, 则删除这条边。通过以上步骤, 可以学习到典型因果骨架图 G 。

3.2 因果方向推断

通过因果骨架学习找到典型因果骨架后, 需要对每条边进行因果定向。由于隐变量的存在且在因果充分性假设下, 因此会遇到 $X \rightarrow Z \rightarrow Y$ 和 $X \rightarrow Y \leftarrow Z$ 两类存在隐变量的结构, 其中 Z 均是不可观测的。

3.2.1 结构级联非线性加性噪声模型

对于第1类结构已在非线性级联加性噪声模型中得到解决。对于第2类结构, CANM 是没有考虑的, 即 $X \rightarrow Y \leftarrow Z^*$, 在此使用 Z^* 加以区分。该结构满足 ANM 模型, 即 $Y = f(X, Z^*) + \varepsilon$, 且 Z^* 是隐变量。因此, 本文认为非线性级联加性噪声模型可以应用于该结构, 通过对第1类结构的边缘分布进行变换使其得到等价于第2类结构的形式:

$$\begin{aligned} \log_a \prod_{i=1}^m \int p_\theta(x^{(i)}, y^{(i)}, z) dz &= \\ \log_a \prod_{i=1}^m \int p_\theta(x^{(i)}) p_\theta(y^{(i)} | z) p_\theta(z | x^{(i)}) dz &= \\ \log_a \prod_{i=1}^m \int p(x^{(i)}) p_\theta(\varepsilon^{(i)} = y^{(i)} - f(x, n_i)) p(n_i) dn_i &= \\ \log_a \prod_{i=1}^m \int p(x^{(i)}) p_\theta(\varepsilon^{(i)} = y^{(i)} - \hat{f}(x, z^*)) p(z^*) dz^* &= \\ \log_a \prod_{i=1}^m \int p_\theta(x^{(i)}, y^{(i)}, z^*) dz^* \end{aligned} \quad (1)$$

其中: n_i 是在第1类结构分布下 $X \rightarrow Z$ 的噪声, 在第2个等式中使用一般的 f 表达了中间传递过程, 即 $Y = f_y(f_x(X) + N_z) = f(X, N_z)$ 。关键在于第3个等式, 如果令噪声分布与 Z^* 的分布相等, 则可以得到等价于第2类结构的边缘分布。换言之, 可以使用 CANM 框架对第2类结构进行建模。在下文中将给出 CANM 的变分下界以及基于变分自编码机的优化方案。

3.2.2 变分下界

根据式(1)可进一步将单个隐变量推广到多个, 并且它们的噪声用向量 \mathbf{n} 表示, 因此非线性级联加性噪声模型的边缘似然度可表示如下:

$$\begin{aligned} \log_a \prod_{i=1}^m \int p_\theta(x^{(i)}, y^{(i)}, \mathbf{n}) d\mathbf{n} &= \\ \log_a \prod_{i=1}^m \int p(x^{(i)}) p_\theta(\varepsilon^{(i)} = y^{(i)} - f(x, \mathbf{n})) p(\mathbf{n}) d\mathbf{n} &= \\ \log_a \prod_{i=1}^m \int p_\theta(x^{(i)}, y^{(i)}, \mathbf{n}) d\mathbf{n} \end{aligned} \quad (2)$$

基于式(2)可以给出在样本点 $(x^{(i)}, y^{(i)})$ 下的边缘似然度的变分下界:

$$\begin{aligned} \log_a \int p_\theta(x^{(i)}, \varepsilon^{(i)}, \mathbf{n}) d\mathbf{n} &\geq \\ E_{\mathbf{n} \sim q_\phi(\mathbf{n} | x^{(i)}, y^{(i)})} \left[-\log_a q_\phi(\mathbf{n} | x^{(i)}, y^{(i)}) + \log_a p_\theta(x^{(i)}, \varepsilon^{(i)}, \mathbf{n}) \right] &= \\ \log_a p(x^{(i)}) - \text{KL}(q_\phi(\mathbf{n} | x^{(i)}, y^{(i)}) || p_\theta(\mathbf{n})) + \\ E_{\mathbf{n} \sim q_\phi(\mathbf{n} | x^{(i)}, y^{(i)})} \left[\log_a p(\varepsilon^{(i)} = y^{(i)} - f(x^{(i)}, \mathbf{n}; \theta)) \right] \end{aligned} \quad (3)$$

其中: $q_\phi(\mathbf{n} | x^{(i)}, y^{(i)})$ 是变分分布的, 其作用是近似后验分布 $p_\theta(\mathbf{n} | x^{(i)}, y^{(i)})$, 当且仅当 $\text{KL}(q_\phi(\mathbf{n} | x^{(i)}, y^{(i)}) || p_\theta(\mathbf{n} | x^{(i)}, y^{(i)})) = 0$ 时, 上述下界的等号成立, 即只要能够找到一个足够近似的变分分布, 就能近似最大化模型的边缘似然度。

3.2.3 变分自编码器

为更好地优化边缘似然度的同时近似后验分布 $p_\theta(\mathbf{n} | x^{(i)}, y^{(i)})$, 使用基于变分自编码器 (Variational Auto-Encoder, VAE)^[21] 的求解方案。具体地, 使用多层感知机 (Multi-Layer Perceptron, MLP) 来分别作为编码器 $q_\phi(\mathbf{n} | x^{(i)}, y^{(i)})$ 和解码器 $p_\theta(\mathbf{n} | x^{(i)}, y^{(i)})$ 的全局近似器^[22]。同时, 针对期望项中的梯度求解问题, 使用一种重参数化技术, 假设先验分布 $p(\mathbf{n}) \sim N(0, I)$ 服从高斯分布, 使得式(3)改写如下:

$$\begin{aligned} \log_a \int p_\theta(x^{(i)}, \varepsilon^{(i)}, \mathbf{n}) d\mathbf{n} &\geq \\ \log_a p(x^{(i)}) - \text{KL}(q_\phi(\mathbf{n} | x^{(i)}, y^{(i)}) || p_\theta(\mathbf{n})) + \\ E_{u \sim N(0, I)} \left[\log_a p(\varepsilon^{(i)} = y^{(i)} - f(x^{(i)}, \mu_\phi^{(i)} + \sigma_\phi^{(i)} u; \theta)) \right] \end{aligned} \quad (4)$$

其中: $\mu_\phi^{(i)}, \sigma_\phi^{(i)}$ 分别是2个独立的 MLP, 用于推断变分高斯分布的均值、方差, 输入为 $x^{(i)}, y^{(i)}$ 。在期望中, u 服从标准高斯分布, 因此该期望值可以直接使用蒙特卡洛抽样进行估计。基于式(4), 通过比较各个方向上的边缘似然度对各类存在隐变量的因果关系进行因果方向推断。

3.3 算法描述

在得到各种隐变量结构下仍然适用的因果方向推断方法后, 将这种结构级联非线性加性噪声模型应用到典型因果骨架上, 得到一种基于级联非线性加性噪声模型的混合因果结构学习算法。

算法2 混合因果结构学习

输入 样本集 $D = \{x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\}_{i=1}^m$

输出 因果结构

1. 使用算法1学习出典型因果骨架 G
2. For 每个因果对 $\langle X_i, X_j \rangle \in G$ then
3. 针对 $\langle X_i, X_j \rangle$ 使用 VAE 对式(4)进行优化得到 $L_{X_i \rightarrow X_j}$
4. 针对 $\langle X_j, X_i \rangle$ 使用 VAE 对式(4)进行优化得到 $L_{X_j \rightarrow X_i}$
5. If $L_{X_i \rightarrow X_j} > L_{X_j \rightarrow X_i}$ then
6. G 增加边 $X_i \rightarrow X_j$
7. Else if $L_{X_i \rightarrow X_j} < L_{X_j \rightarrow X_i}$ then
8. G 增加边 $X_i \leftarrow X_j$
9. Else
10. $\langle X_i, X_j \rangle$ 标记为不可识别
11. End if
12. End For
13. Return 典型因果结构 G

算法2分为两个阶段:第一阶段是使用算法1学习出典型因果骨架(第1行);第二阶段是使用级联非线性加性噪声模型对每个因果边进行方向推断(第2~12行)。最终算法的输出是典型因果结构(第13行),该结构揭示了变量间存在隐变量下的因果结构关系。

4 实验与结果分析

使用虚拟因果结构与真实结构数据集对模型进行评估,将本文SCANM算法与以下主流因果结构学习算法进行对比:1)基于盲源分离方法在线性非高斯数据下进行因果结构学习的LiNGAM算法^[16];2)基于爬山法与贝叶斯评分进行因果结构搜索的HC算法^[23];3)将约束与贝叶斯评分两者集成进行因果结构搜索的MMHC算法^[24]。为验证隐变量建模的有效性,在因果结构上的每一个因果对之间均会引入额外的隐变量,该隐变量满足级联加性噪声的形式。在实验中,每组参数都至少运行80次以上,并采用F1值(F)作为评价指标,计算公式如下:

$$F = 2 \times \frac{\text{正确边数量}}{\text{拟合结构边数量} + \text{真实结构边数量}} \quad (5)$$

4.1 虚拟因果结构数据集实验

在虚拟因果结构数据集中设计4个变量控制实验,每个实验都会随机生成不同的虚拟因果结构,具体设置为:隐变量数量为0、1、2、3、4,结构维度为6、10、15、20,平均入度为1.0、1.5、2.0,样本数量为250、500、1 000、2 000、3 000、4 000、5 000,其中加粗数据表示在各控制实验中的默认设置。

图2给出了不同隐变量数量下的实验结果。由图2可以看出:一方面,SCANM算法在0个和1个隐变量下F1值都在0.87附近,验证了隐变量实验的有效性;另一方面,随着隐变量数量的增加,在隐变量数量为4时,SCANM算法的F1值下降至0.73,这是因为隐变量的增加会导致信噪比的降低,从而影响模型学习效果。

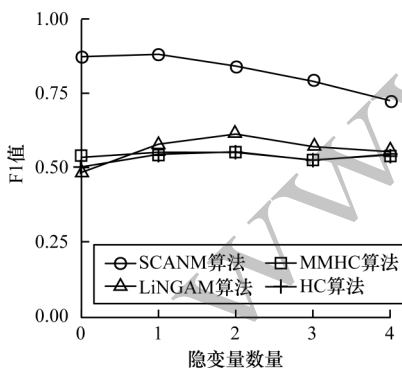


图2 不同隐变量数量下的实验结果

Fig.2 Experimental results under different number of hidden variables

图3给出了不同结构维度下的实验结果,可以看出在不同的结构维度下SCANM算法的F1值在0.83附近波动,这意味着级联非线性加性噪声模型在不同的结构维度下具有较强的鲁棒性。

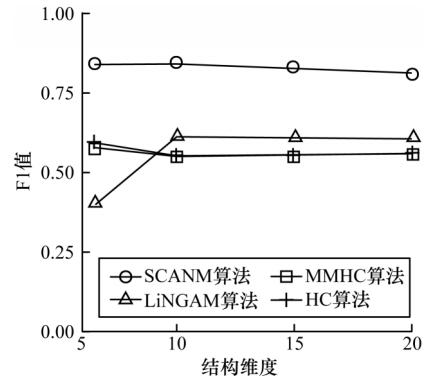


图3 不同结构维度下的实验结果

Fig.3 Experimental results under different structural dimensions

图4给出了不同平均入度下的实验结果。由图4可以看出,HC与MMHC算法对平均入度较为敏感,而且随着平均入度的增加F1值下降比较明显,尤其在平均入度在1.0~1.5时,F1值下降了近0.1,其原因是当边数较多时,若算法无法识别存在的隐变量,则出错的边数会增多,从而使得F1值下降,而SCANM算法由于对于隐变量的方向也可识别,因此对平均入度不敏感。

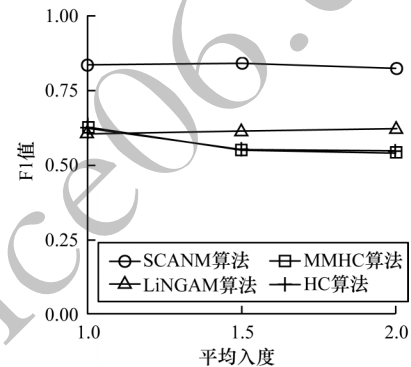


图4 不同平均入度下的实验结果

Fig.4 Experimental results under different average in-degree

图5给出了不同样本数量下的实验结果。由图5可以看出,所有算法对于样本数量都较为敏感,样本数量是一个较为重要的变量。尽管在样本数量较少时,所有算法的F1值均不太理想,但SCANM算法的F1值仍至少比其他算法高0.15,体现出其性能的优越性。

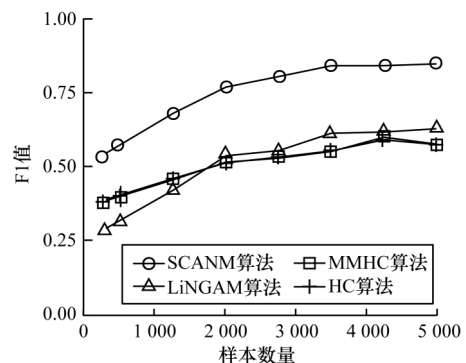


图5 不同样本数量下的实验结果

Fig.5 Experimental results under different number of samples

4.2 真实因果结构数据集实验

选择4种真实结构进行测试,真实结构数据来自<https://www.bnlearn.com/bnrepository/>,具体信息如表1所示。

表1 真实因果结构数据集信息				
Table 1 Real causal structure dataset information				
真实结构	结点数	平均入度	边数	最大入度
survey	6	2.00	6	2
asia	8	2.00	8	2
sachs	11	3.09	17	3
child	20	1.25	25	2

在真实因果结构数据集实验中使用F1值、准确率、召回率作为评价指标,计算公式如下:

准确率 = $\frac{\text{正确边数量}}{\text{预测结构边数量}}$ (6)

召回率 = $\frac{\text{正确边数量}}{\text{真实结构边数量}}$ (7)

在真实因果结构数据集中,使用与虚拟因果结构数据集实验相同的设置,实验结果如图6~图8所示,总体而言,SCANM算法在不同真实结构的平均F1值为0.82,相比其他算法提升了51%。特别地,可以看出:在不同的真实结构中SCANM算法均取得了最好的效果,尤其在sachs结构中,其原因是sachs结构的平均入度更大,级联非线性加性噪声模型在该情况下具有更好的效果;对比算法在不同的真实结构中召回率偏高而准确率偏低,尤其在child结构中最为明显,主要因为没有考虑隐变量的存在,容易学习出冗余边。上述结果验证了SCANM算法的有效性。

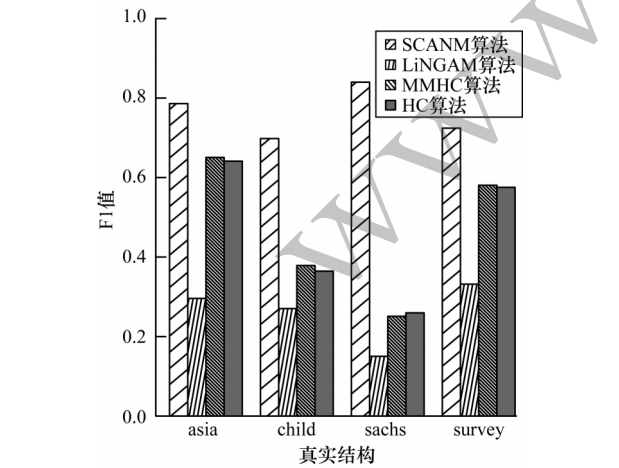


图6 真实因果结构数据集中的算法F1值比较

Fig.6 Comparison of F1 values for algorithms in real causal structure dataset

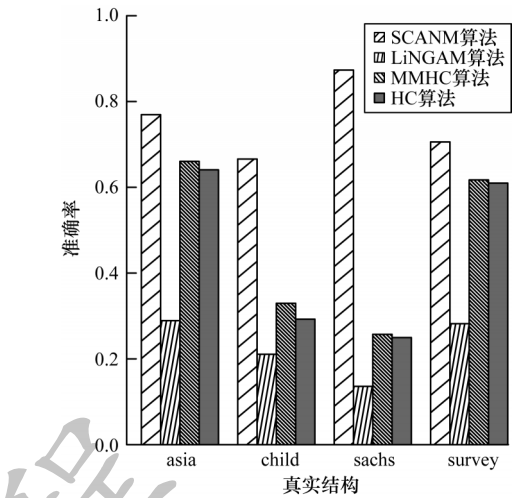


图7 真实因果结构数据集中的算法准确率比较
Fig.7 Comparison of precision for algorithms in real causal structure dataset

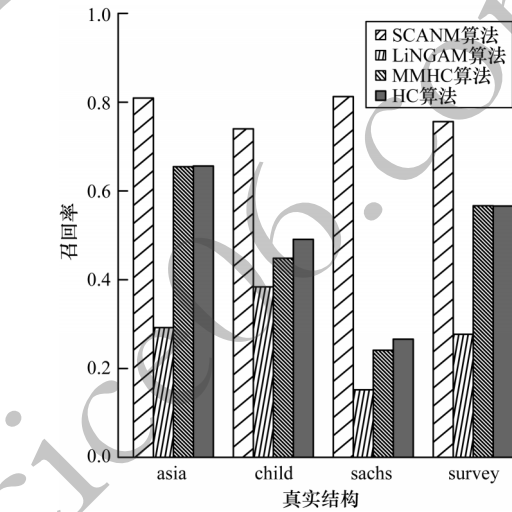


图8 真实因果结构数据集中的算法召回率比较
Fig.8 Comparison of recall for algorithms in real causal structure dataset

5 结束语

本文提出一种基于结构级联非线性加性噪声模型的因果结构学习算法,通过联合传统基于独立性的因果骨架学习算法以及级联加性噪声模型,解决了存在隐变量的因果结构学习问题。实验结果表明,在虚拟结构数据集和真实结构数据集下,该算法相比主流因果结构学习算法准确率更高、鲁棒性更强。后续将研究不满足因果充分性假设下的因果结构学习算法,进一步扩展级联非线性加性噪声模型的适用范围。

参考文献

[1] SPIRITES P, GLYMOUR C, SCHEINES R. Causation, prediction, and search [M]. Cambridge, USA : MIT Press, 2001.

- [2] PEARL J. Causality: models, reasoning and inference[M]. Cambridge, UK: Cambridge University Press, 2009.
- [3] 蔡瑞初,陈薇,张坤,等. 基于非时序观察数据的因果关系发现综述[J]. 计算机学报, 2017, 40(6): 1470-1490.
- CAI R C, CHEN W, ZHANG K, et al. A survey on non-temporal series observational data based causal discovery[J]. Chinese Journal of Computers, 2017, 40(6): 1470-1490. (in Chinese)
- [4] CAI R C, ZHANG Z J, HAO Z F, et al. Understanding social causalities behind human action sequences[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(8): 1801-1813.
- [5] RUNGE J, BATHIANY S, BOLLT E, et al. Inferring causation from time series in earth system sciences[J]. Nature Communications, 2019, 10: 2553-2567.
- [6] CAI R C, ZHANG Z J, HAO Z F. Causal gene identification using combinatorial v-structure search[J]. Neural Networks, 2013, 43: 63-71.
- [7] HERNÁN M A, ROBINS J M. Causal inference[M]. Boca Raton, USA: CRC Press, 2010.
- [8] MOOIJ J M, PETERS J, JANZING D, et al. Distinguishing cause from effect using observational data: methods and benchmarks[J]. Journal of Machine Learning Research, 2016, 17(1): 1103-1204.
- [9] 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.
- [10] OGARRIO J M, SPIRITES P, RAMSEY J. A hybrid causal search algorithm for latent variable models [C]// Proceedings of the 8th International Conference on Probabilistic Graphical Models. Lugano, Switzerland: [s. n.], 2016: 368-379.
- [11] CAI R C, QIAO J, ZHANG K, et al. Causal discovery with cascade nonlinear additive noise model[EB/OL]. [2020-10-08]. <https://arxiv.org/abs/1905.09442v2>.
- [12] BRESSLER S L, SETH A K. Wiener-Granger causality: a well established methodology[J]. Neuro Image, 2011, 58(2): 323-329.
- [13] 张义杰,李培峰,朱巧明. 面向事件时序与因果关系的联合识别方法[J]. 计算机工程, 2020, 46(7): 65-71.
- ZHANG Y J, LI P F, ZHU Q M. Joint identification method for temporal and causal relations of events[J]. Computer Engineering, 2020, 46(7): 65-71. (in Chinese)
- [14] LAM W, BACCHUS F. Learning Bayesian belief networks: an approach based on the MDL principle[J]. Computational Intelligence, 1994, 10(3): 269-293.
- [15] ANDERSSON S A, MADIGAN D, PERLMAN M D. A characterization of Markov equivalence classes for acyclic digraphs[J]. The Annals of Statistics, 1997, 25(2): 505-541.
- [16] 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.
- [17] 姜枫,朱辉生,汪卫. 含隐变量非高斯无环因果模型的估计算法[J]. 计算机工程, 2010, 36(9): 178-180.
- JIANG F, ZHU H S, WANG W. Estimation algorithm for non-Gaussian acyclic causal model with latent variables[J]. Computer Engineering, 2010, 36(9): 178-180. (in Chinese)
- [18] HOYER P, JANZING D, MOOIJ J M, et al. Nonlinear causal discovery with additive noise models [C]// Proceedings of the 22nd Annual Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2008: 689-696.
- [19] ZHANG K, HYVÄRINEN A. On the identifiability of the post-nonlinear causal model[C]// Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. [S. l.]: AUAI Press, 2009: 647-655.
- [20] HOYER P O, SHIMIZU S, KERMINE 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.
- [21] KINGMA D P, WELING M. Auto-encoding variational Bayes[EB/OL]. [2020-11-04]. <https://dare.uva.nl/search?identifier=cf65ba0f-d88f-4a49-8ebd-3a7fce86edd7>.
- [22] HORNIK K, STINCHCOMBE M, WHITE H. Multilayer feedforward networks are universal approximators[J]. Neural Networks, 1989, 2(5): 359-366.
- [23] GÁMEZ J A, MATEO J L, PUERTA J M. Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood[J]. Data Mining and Knowledge Discovery, 2011, 22(1/2): 106-148.
- [24] 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.

编辑 陆燕菲