

基于神经网络的混合数据的因果发现

耿家兴¹, 万亚平^{1,2}, 李洪飞¹

(1. 南华大学 计算机学院, 湖南 衡阳 421001;

2. 中核集团高可信计算重点学科实验室, 湖南 衡阳 421001)

摘要:因果推理正在成为机器学习领域一个越来越受关注的研究热点, 现阶段的因果发现主要是在研究某一种假设条件下, 基于纯粹的观测数据推断变量之间的因果方向。然而在现实世界中观察到的数据往往是由一些假设生成, 使得传统因果推断方法的识别率不高、稳定性较差。针对当前的问题, 提出了一种基于神经网络来解决混合数据因果推断的方法。该方法在混合加性噪声模型 (ANM-MM) 的假设下, 使用梯度下降法最优化改进的损失函数得到混合数据的抽象因果分布参数, 然后将分布参数看作是原因变量和结果变量之间的隐变量, 通过比较原因变量和分布参数之间的 HilberSchmidt 独立性来确定二元变量的因果方向。在理论上证明了该方法的可行性, 并通过实验表明该算法在人工数据和真实数据的表现较传统的 IGCI, ANM, PNL, LiNGAM, SLOPE 方法具有较好的准确率和稳定性。

关键词:神经网络; 混合加性噪声; 因果推断; 梯度下降; HilberSchmidt 独立性

中图分类号: TP181

文献标识码: A

文章编号: 1673-629X(2020)05-0026-06

doi: 10.3969/j.issn.1673-629X.2020.05.006

Causal Discovery of Mixed Data Based on Neural Network

GENG Jia-xing¹, WAN Ya-ping^{1,2}, LI Hong-fei¹

(1. School of Computer Science, University of South China, Hengyang 421001, China;

2. CNC Key Laboratory on High Trusted Computing, Hengyang 421001, China)

Abstract: Causal discovery is becoming a research hotspot in the field of machine learning. At present, the causal discovery is mainly to investigate the causal direction between variables based on pure observation data under the study of a certain assumption. However, the data observed in the real world is often generated by some assumptions, which makes the traditional causal inference method less accurate and less stable. Aiming at the current problem, a method based on neural network to solve the causal inference of mixed data is proposed. Under the assumption of additive noise model-mixture model (ANM-MM), the gradient loss method is used to optimize the improved loss function to obtain the abstract causal distribution parameters of the mixed data, and then the distribution parameters are regarded as hidden variable between the cause variable and the result variable. The hidden variable determines the causal direction of the binary variable by comparing the HilberSchmidt independence between the causal variable and the distribution parameter. The feasibility of the method is proved theoretically. The experiment shows that the proposed algorithm has better accuracy and stability than the traditional methods like IGCI, ANM, PNL, LiNGAM and SLOPE.

Key words: neural network; mixed additive noise; causal inference; gradient descent; HilberSchmidt independence

0 引言

随着机器学习、深度学习等人工智能算法的崛起, 人们逐渐意识到不能单纯地关注数据的相关性。2011年图灵奖得主 Judea Pearl^[1]认为人工智能深陷关联概率的泥潭从而忽视了因果, 同时 Judea Pearl 认为研究

者应该研究因果, 因为这可能是实现真正的人工智能的可能路径。随着因果推断的深入研究, 基于因果推断的应用也越来越多, 例如在 fMRI 数据上的分析和分子发现途径研究; Statnikov 等人^[2]在 2012 提出了基因组学数据中因果分子相互作用的鉴定方法; Ma Sisi

收稿日期: 2019-06-19

修回日期: 2019-10-22

网络出版时间: 2020-01-10

基金项目: 国家自然科学基金(11805093); 中央军委科技委创新特区项目(17-163-15-XJ-002-002-04); 湖南省教育重点项目(17A185); 湖南省自然科学基金资助项目(2019JJ0486)

作者简介: 耿家兴(1995-), 男, 硕士研究生, 研究方向为机器学习、因果关系发现; 万亚平, 教授, 硕导, CCF 会员(14108M), 研究方向为大数据与因果关系。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20200110.1121.032.html>

在2017年的工作中讨论了生物医学中观察数据可用于因果发现^[3],以及在网络性能优化^[4]、网购服装推荐^[5]、社交媒体应用等^[6]领域都有相关应用。因此因果推理正在成为机器学习领域一个越来越受关注的研究方向。现阶段的因果发现主要是在不进行任何干预的情况下确定两个变量之间的因果方向,这是一项具有挑战性的工作。在近阶段的因果发现研究进展中允许在某种假设下,基于纯粹观察数据推断两个变量之间的因果方向。然而在现实世界中观察到的数据往往由多种生成方式产生,这使得传统的因果推断方法存在识别率不高和稳定性较差的现象。

1 相关工作

目前在二元变量上的因果推断方法主要有两种,第一种是基于时序数据的因果方向推断,第二种是基于非时序数据的因果关系发现,现阶段研究比较多的是基于非时序数据的因果推断。

非时序数据上的因果关系发现又分为基于统计分析的因果推断和基于模型假设的因果推断。基于统计分析:Kano 等人^[7]于2003年提出路径分析法用于观测数据的因果结构,该方法通过将隐含协方差矩阵与样本协方差矩阵进行比较,对假设的因果关系进行统计检验,从而得出假设的因果关系的正确性。2003年Comley^[8]提出了一种对称贝叶斯网络通过使用一种简单的决策树算法能够很好地识别连续型和离散型的联合概率分布。Sun 等人^[9]于2006年提出一种使用马尔可夫核函数来选择多变量之间的所有假设因果方向中的最可能的一个方向。基于模型假设:Shimizu^[10]在2006年提出了一种线性非高斯非循环模型LinGAM,该算法假设数据是线性非混淆的,同时干扰变量是非高斯的,该模型使用独立成分分析的方法来进行因果分析。Zhang 等人^[11]在2009年通过考虑原因的非线性效应,内部噪声效应和观测变量中的测量失真效应,提出了后非线性因果模型PNL,并证明了该方法在因果方向的可识别性。在2009年Hoyer 提出^[12]加性噪声模型ANM,该模型能够处理非线性的数据,并成为了一种常用的因果发现方法。Janzing 等人^[13]提出通过二阶指数模型定义一系列平滑密度和条件密度,即通过最大化受第一和第二统计矩影响的条件熵来识别包含离散变量和连续变量的因果结构。2010年Daniusis 等人^[14]提出一种确定两个可逆函数相互关联性来识别确定的因果方向。2011年Jonas^[11]提出了利用加性噪声模型对离散型数据的因果推断算法。2012年Janzing^[15]提出了IGCI算法,该算法通过信息空间中的正交性来定义独立性的方法描述因果变量值之间的边缘概率分布和条件概率分布,

从而通过因果关系的不对称性来确定因果方向。2015年Sgouritsa 等人^[16]提出了无监督的高斯回归过程来估计因果变量之间的条件概率CURE算法。2016年Mooij 等人^[17]总结了使用观测数据区分因果关系的方法和标准。2017年Marx 等人^[18]采用了一种基于柯尔莫戈罗夫复杂度的信息论方法,并利用最小值描述长度原理提供了一种实用的因果关系发现方法SLOPE。以上这些方法在进行二元变量的混合数据的因果推断中表现较差,2018年Hu^[19]提出了一种基于加性噪声算法的混合数据的因果推断方法ANM-MM(additive noise model-mixture model)模型。

ANM-MM模型可以处理二元变量的混合数据的因果推断,但是在求解分布参数的过程中没有考虑到结果变量对分布参数的影响。因此文中在ANM-MM模型的基础上提出了一种使用神经网络学习改进的目标函数从而得到分布参数,原ANM-MM模型的目标函数被看作是该目标函数的一种特例。

2 预备知识

2.1 模型假设

文中提出的方法是在ANM-MM模型的假设下进行的,ANM-MM模型的定义是连续性变量 X 和 Y 由有限ANM模型生成,如图1所示。

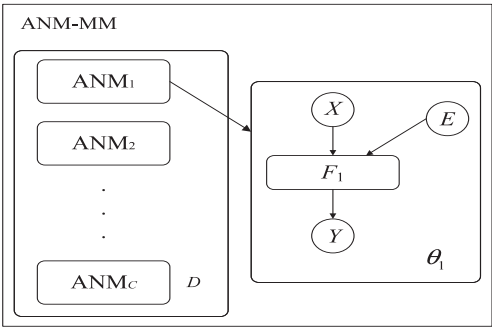


图1 ANM-MM模型

ANM-MM模型满足下面的形式: $Y = f(X; \theta) + \varepsilon$,这里 X 表示原因变量, Y 表示结果变量,噪声 e 与 X 相互独立, f 是非线性函数,由参数 θ 决定,ANM-MM模型之间的差异取决于不同的参数 θ 。在ANM-MM中, θ 的值来源于一个离散分布的有限集,这里的 $a_c > 0, \sum_{c=1}^c a_c = 1, 1_{\theta_c}(\cdot)$ 是 θ_c 的指数函数。

ANM-MM模型中所有的观测数据都是由同一个形式的函数 f 不同的参数 θ 生成,因为在每个独立的试验中,由于某些难以控制的外部因素的影响,数据生成过程略有不同。此外这些因素通常被认为是独立于观测变量的。

2.2 HilberSchmidt 独立性准则

二元变量的独立性测试方法有很多,文中使用的

是 HilberSchmidt 独立性准则。假设 X, Y 是两个一维变量, 定义一个非线性映射 $\varphi(x) \in F, F$ 属于再生核希尔伯特空间 (reproducing kernel Hilbert space) 且 $x \in X$, 再定义一个 $\psi(y) \in G, G$ 同样属于再生核希尔伯特空间且 $y \in Y$, 则相应的核函数分别记为:

$$k(x, x') = \langle \varphi(x), \varphi(x') \rangle, x, x' \in X \quad (1)$$

$$l(y, y') = \langle \psi(y), \psi(y') \rangle, y, y' \in Y \quad (2)$$

对于所有的 $f \in F, g \in G$ 定义:

$$\langle f, C_{xy}g \rangle_F = E_{xy}([f(x) - E_x(f(x))][g(y) - E_y(g(y))]) \quad (3)$$

交叉协方差算子 $C_{xy}: G \rightarrow F$ 且:

$$C_{xy} = E_{xy}[(\varphi(x) - \mu_x) \otimes (\psi(y) - \mu_y)] \quad (4)$$

其中, $\mu_x = E_x \varphi(x), \mu_y = E_y \varphi(y), \otimes$ 为张量积, C_{xy} 可以看作 Hilbert-Schmidt 算子, 然后将 Frobenius 范数扩展到该算子上, 得到 Hilbert-Schmidt 范数, 将该范数平方即可得到 Hilbert-Schmidt 独立性准则:

$$HSIC(F, G, \text{Pr}_{xy}) = \|C_{xy}\|_{\text{HS}}^2 \quad (5)$$

其中, Pr_{xy} 是 X, Y 的概率分布。

3 模型可识别性和参数估计

3.1 模型的可识别性

定理: 在混合的 ANM 中, 如果 $Y \rightarrow X$, 假设存在

反向 ANM, 其中 $\omega \sim p_\omega(\omega) = \sum_{c=1}^C b_c 1_{\omega_c}(\cdot), b_c > 0, \sum_{c=1}^C b_c$ 并且 $\varepsilon \perp Y$, 则相反的方向上应满足常微分方程:

$$\xi'' - \frac{G(X, Y)}{H(X, Y)} \xi' = \frac{G(X, Y)V(X, Y)}{U(X, Y)} - H(X, Y) \quad (6)$$

$$\xi = \log p(X), G(X, Y) = \frac{\partial U}{\partial X}, H(X, Y) = \frac{\partial V}{\partial X}$$

其中, $U(X, Y) = \frac{\partial^2 \pi}{\partial X \partial Y}, V(X, Y) = \frac{\partial^2 \pi}{\partial X^2}$, 这里的 π 为

$$\pi(X, Y) = \log p(X, Y) = \log \sum_{c=1}^C a_c p_c(Y - f_c(X)) + \log p_X(X)$$

简要证明: 假设 ANM-MM 在两个方向上都存在方向, 那么一定存在一个非重叠区域 $D = \{(x_n, y_n)\}_{n=1}^N = D_1 \cup D_2 \cup \dots \cup D_n$, 假设是 D_c 为非重叠区域, 满足 $Y = f(X; \theta) + \varepsilon$, 则 $\theta \sim p_\theta^c(\theta)$ 是有限集合

$\Theta^c \subseteq \Theta$ 上的离散分布, 并且 $X = g(Y; \omega = \omega_c) + \varepsilon$, 为了保证在因果方向上存在一个 ANM-MM, 则反因果方向上存在一个 ANM, 使得其 $(p_X, p_\varepsilon, f, p_\theta)$ 应满足常微分方程的形式(1)。那么后向因果方向的存在需要 C 常微分方程同时满足得到(4), 所以在混合的 ANM-MM 模型中不存在反向 ANM 的情况。如图 2 所示, 将分布参数看作是原因变量和结果变量之间的隐

变量。

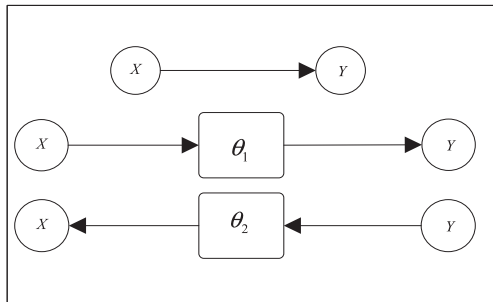


图 2 因果推断示意图

根据定理可以通过研究假设原因与相应函数参数之间的独立性, 推断出模型的因果方向, 如果它们在因果方向上是独立的, 那么它们很可能在反因果方向上是相互依赖的。因此在实际中两变量间的因果推断的方向是变量和中间参数之间独立性更大的方向。

3.2 参数估计

在 ANM-MM 模型中利用目标函数的对数似然极大化通过梯度下降来寻找潜在的分布参数的表示形式, 然而 θ 不能直接通过极大似然估计进行求解, 所以 ANM-MM 需要添加 X 和分布参数 θ 之间的 HSIC 独立性到损失函数得到:

$$\argmin_{\Theta, \Omega} J(\Theta) = \argmin_{\Theta, \Omega} [-L(\Theta | X, Y, \Omega) + \lambda \log \text{HSIC}_b(X, \Theta)] \quad (7)$$

由于在求解分布参数的过程中没有考虑到结果变量对分布参数的影响, 文中修改了 ANM-MM 模型的损失函数(7), 另外添加 Y 和分布参数 θ 之间的 HSIC 独立性到损失函数得到:

$$\argmin_{\Theta, \Omega} J(\Theta) = \argmin_{\Theta, \Omega} [-L(\Theta | X, Y, \Omega) + \lambda \log \text{HSIC}_b(X, \Theta) + \alpha(1 - \lambda) \log \text{HSIC}_b(Y, \Theta)] \quad (8)$$

其中, λ 控制原因变量 X 和分布参数之间的 Hilbert-Schmidt 独立性, α 用来调节结果变量 Y 和分布参数的 Hilbert-Schmidt 独立性, 当 λ 增大时即增强 X 和分布参数的独立性, 由于结果变量前又有 $(1 - \lambda)$ 进行约束, 则 Y 和分布参数独立性势必会减小, 因此并未违反因果逻辑性。当 $\alpha = 0$ 时, 文中使用的目标函数(8)和 ANM-MM 的目标函数(7)具有一致性。由于该目标函数包含两个方向上的独立性测试, 因此文中算法记作 Dual-ANMM, 另外当 $\alpha = 1$ 时目标函数改写为:

$$\argmin_{\Theta, \Omega} J(\Theta) = \argmin_{\Theta, \Omega} [-L(\Theta | X, Y, \Omega) + \lambda \log \text{HSIC}_b(X, \Theta) + (1 - \lambda) \log \text{HSIC}_b(Y, \Theta)] \quad (9)$$

其中, 目标函数(8)引入 α 的目的是控制数据和参数之间 HilbertSchmidt 独立性的强弱。文中在实验中使用神经网络来学习, 采取梯度下降法最优化的 $\alpha = 1$ 的目标函数(9)得到混合数据的抽象因果分布参数。

4 算法过程

文中使用了三层网络结构来学习混合数据的分布参数,第一层是 Normlization layer,该层网络主要将输入数据正则化、规范化;第二层是 Full-Connect 全连接层,文中采用的是 10 个神经元;最后一层是激活层,文中使用的是 LRelu 激活。最后将通过梯度下降法求得目标函数(3)的最小值从而得到混合数据的分布参数。再通过 HilberSchmidt 独立性测试确定因果方向,如图 3 所示。

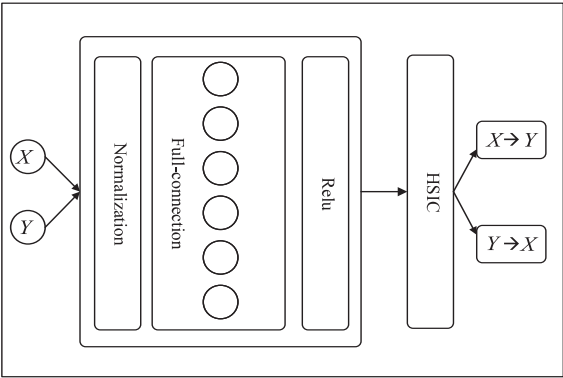


图 3 网络架构

当 $HSIC(X, \theta_1) > HSIC(Y, \theta_2)$ 时, $X \rightarrow Y$;
当 $HSIC(X, \theta_1) < HSIC(Y, \theta_2)$ 时, $Y \rightarrow X$;
当 $HSIC(X, \theta_1) = HSIC(Y, \theta_2)$ 时无法识别,第三种情况在实际情况下很少发生。文中使用的 HilberSchmidt 独立性是 Hu Shoubo 改进的 GPPOM_ HSIC 独立性方法。

Dual-ANMM 算法过程如下:

Algorithm Dual-ANMM(A kind of causal discovery by neural network)
输入: D , $\alpha = 1$, λ //其中 D 是数据, α 和 λ 为超参数
输出: $direct$ //输出方向
1. Standardize observation of each D ;
2. Random initialize weight of NN and other argument
3. Optimize target function by NN in $X \rightarrow Y$ estimate θ_1
4. $D_1 = HSIC(\theta_1, X)$
5. Optimize target function by NN in $Y \rightarrow X$ estimate θ_2
6. $D_2 = HSIC(\theta_2, X)$
7. If $D_1 < D_2$ then $direct = 1$
8. Elif $D_1 > D_2$ then $direct = -1$
9. Else $direct = 0$
10. Return $direct$

5 传统因果推断方法对比

LiNGAM 模型^[10]是一种线性非高斯因果推断模型, LiNGAM 模型假设数据由 $E = \beta C + N$ 产生,其中 β 属于实数域, $C \perp N$ 并且 N 是非高斯的。LiNGAM 仅

适用于非高斯噪声的线性关系,因此如果生成的数据违反了该假设, LiNGAM 模型会表现得极差。如果数据之间的关系是线性函数的关系,识别率会提高,而且计算成本相对较低。在实验中,文中使用的 LiNGAM 是原始版本中基于独立成分分析的算法。

ANM 模型^[12]是一种加性噪声模型, ANM 模型假设 $E = f(C) + N$, 这里的 f 是非线性函数并且 $C \perp N$, 该方法能够处理非线性数据。由于现实世界中的数据大多是非线性数据,该方法得到了广泛的应用。ANM 算法识别因果关系使用原因和结果之间的不对称来检测,原因和结果之间的不对称是通过原因和残差之间独立的假设实现。因此该方法需要拟合回归函数,并对输入与的选择对性能识别率是至关重要的。在实验中,文中使用了 Mooij 等人(2016)提出的一种实现,该实现使用高斯过程回归进行预测,并使用熵估计进行输入和残差之间的关系评估。

PNL 模型^[11]是一种后线性模型,该模型是对 ANM 模型的一种泛化。PNL 模型假设数据的生成机制是由 $E = g(f(C) + N)$ 表示,其中 g 是非线性函数, f 是线性函数,并且 $C \perp N$ 。为了推断出因果关系的方向,在两个可能的方向上进行 PNL 模型测试,并检查输入和扰动之间的独立性。然而这里的扰动不同于回归残差,拟合 PNL 模型比拟合 ANM 要困难得多,准确率较高于 ANM 模型。在实验中,文中使用了基于约束非线性独立分量分析来估计扰动,并使用 HilberSchmidt 独立性准则进行统计独立测试。

IGCI 算法^[15]是一种基于信息准则估计方法的因果推断模型,该模型通过比较一个参考测度和 P_x 之间的 KL 散度大于或小于相同参考测度和 P_y 之间的 KL 散度,来进行因果方向识别。该算法已应用于部分成功的噪声因果关系,但是对于非确定性关系,IGCI 的一般化实际上并不能得到很好的识别,而且计算成本依赖于所使用的信息准则估计方法,一般较低。实验中,文中使用 Mooij 等人提供的一种实现,在该实现中算法测试所有可能的参考度量和信息估计的组合。这些组合表示为 IGCI-ij,其中 i 的取值为 U 和 G 分别表示统一参考测量(标准化 X 和 Y)和高斯参考测量(标准化 X 和 Y),当 $j = 1$ 时使用 Daniu'sis 等人在 2010 年提出的 $\hat{S}(pX) = \psi(m) - \psi(1) + \frac{1}{m-1} \sum_{i=1}^{m-1} \log |x_{i+1} - x_i|$ 作为熵估计器;当 $j = 2$ 时使用方程 $\int \log |f(x)| p(x) dx \approx \frac{1}{m-1} \sum_{i=1}^{m-1} \log \left| \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right|$ 进行积分近似;当 $j = 3$ 时采用 Mooij 等人在 2016 年提出的方程 $\frac{|f(x_2) - f(x_1)|}{x_2 - x_1} \leq \max_{x \in [0,1]} |f'(x)| = s_{\max}$ 进行积分

近似。实验中分别设置 $i=1, j=3$ 作为 IGC1 的算法估计。

SLOPE 算法^[18]是由 Marx 和 Vreeken 在 2017 年提出的,该方法基于最小描述长度原则进行 $\frac{(K(p_X) + K(p_{Y|X}))}{K(p_X) + K(p_Y)}$ 和 $\frac{(K(p_X) + K(p_{Y|X}))}{K(p_X) + K(p_Y)}$ 的估计和比较来识别因果关系。这种方法使用全局和多个局部回归模型来拟合数据,其中拟合回归模型的描述长度和相对于数据的误差描述长度可用于近似 $K(p_{Y|X})$ 和 $K(p_{X|Y})$,在比较实验中,文中使用了作者提供的实现,其参数与他们在实际数据实验中使用的参数相同,回归模型需要根据数据的结构进行匹配,因此计算成本在数据集之间可能有所不同。

6 实验过程

实验在 CentOS 系统上 python3.6 环境中进行,分别在模拟数据集和真实数据集上实验,其他对比实验包括 LiNGAM, IGC1, PNL, ANM 分别按照上面实现方法在 matlab2016a 中进行。SLOPE 实验是 R3.5.2 中进行。首先比较了 λ 对实验准确率的影响,在实验中分别取 λ 为 0.5, 0.75, 1.0, 1.25, 1.5, 其中当 λ 等于 1.0 的时候该模型等价于 ANM-MM 模型使用的损失函数。文中对混合数据类型个数对因果推断的准确率的影响进行了实验,然后是数据量对 ANM-MM, ANM, IGC1, PNL, SLOPE, LiNGAM 的影响进行了实验,最后在 cause-effect pairs 真实数据集上评估了算法的因果推理性能。cause-effect pairs 具有用于测试因果检测算法的不同数据,数据文件一共包含 108 个 txt 文件,每个文件包含两个变量,一个是原因,另一个是效果。对于每个数据文件,都有一个描述性文件,可以在其中找到真实的因果关系以及数据来源描述,实验中 108 组数据中有 9 组被排除在外,因为它们要么由多变量数据组成,要么由分类数据组成(对 47、52、53、54、55、70、71、101 和 105)。每对实验重复 100 次独立实验,记录不同方法的正确推理率。然后计算同一数据集对的平均百分比作为对应数据集的精度,实验结果如下。

如图 4 所示,实验测试了 λ 对实验准确率的影响,其中当 λ 等于 1.0 时表示 ANM-MM 模型对应的目标函数,实验表明 $\alpha=1.0$ 时, $\lambda=1.25$ 时结果会达到最好。

如图 5 所示,实验对比了混合数据类型个数对两种算法的影响。可以看出文中提出的算法是优于 ANM-MM 模型的,在实验中,当只有一种因果机制的时候 ANM-MM 模型的准确率低于 65%,这是在 100 次重复实验中得到的结果,而文中提出的 Dual-

AMMM 在参数 $\alpha=1.0, \lambda=1.25$ 时准确率能达到 77%,即当数据并不是由多种因果机制生成,ANM-MM 识别率低于 Dual-AMMM。在多于一种混合类型的数据时, Dual-AMMM 算法和 ANM-MM 模型结果不相上下。

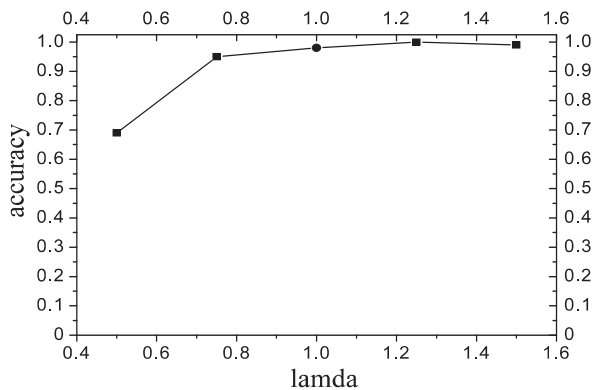


图 4 λ 对实验准确率的影响

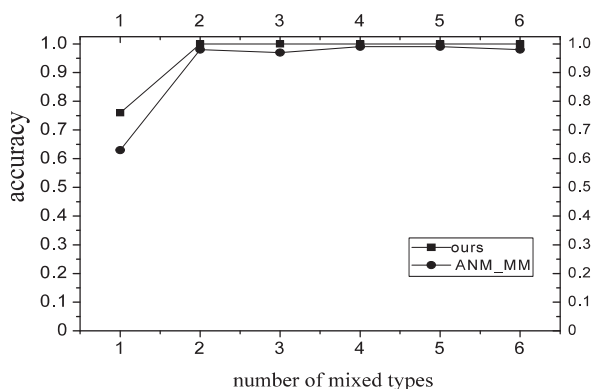


图 5 混合数据类型个数对准确率的影响

该实验对比了几种不同的传统的因果推断算法和 ANM-MM 模型在不同数据量下的模拟数据的准确率,如图 6 所示。实验结果表明,在不同的数据量下,该算法在稳定性上优于传统的方法,其中 IGC1 在数据量增加时结果逼近文中提出的 Dual-AMMM 算法以及 ANM-MM 模型,由上图可以看出 Dual-AMMM 和 ANM-MM 原始算法和在稳定性和准确率上都优于其它方法。

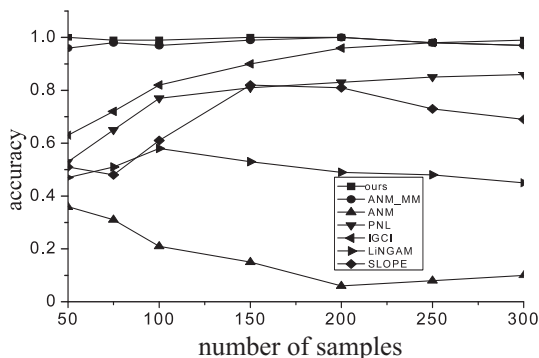


图 6 数据量对模型的影响

该实验在真实因果数据集 cause-effect pairs 上进行,结果如图 7 所示,柱状图的高度表示实验 100 次的

平均准确率,其中 Y 的误差取上下界误差的最大。结果表明文中提出的 Dual-AMMM 算法优于其他传统的方法,平均准确率最高约 85% 左右,ANM-MM 算法在 82% 左右。PNL 也表现很好,其次是 IGCi 算法,但是 IGCi 和 SLOPE 的误差范围较大。

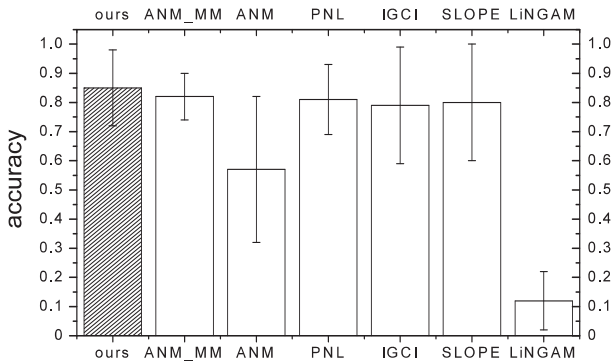


图7 在真实数据集上的对比

7 结束语

在 ANM-MM 的假设基础上提出了一种基于神经网络的因果推断方法。该方法使用梯度下降法最优化损失函数得到混合数据的抽象因果分布参数,然后将分布参数看作是原因变量和结果变量之间的隐变量,通过比较原因变量和分布参数之间的 HilberSchmidt 独立性来确定二元变量的因果方向。由于改进的 ANM-MM 求解分布参数的过程中没有考虑到结果变量对分布参数的影响,在实验上也表现出较其他传统的方法具有较好的稳定性和准确率。但是现阶段的研究主要是在二元变量之间的因果方向推断,因此将该方法推广到多元变量之间的因果图的识别是下一阶段的主要内容。

参考文献:

[1] PETERS J, JANZING D, SCHOLKOPF B. Causal inference on discrete data using additive noise models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2436-2450.

[2] STATNIKOV A, HENAFF M, LYTKIN N I, et al. New methods for separating causes from effects in genomics data [J]. BMC Genomics, 2012, 13: S22.

[3] MA S, STATNIKOV A. Methods for computational causal discovery in biomedicine [J]. Behaviormetrika, 2017, 44(1): 165-191.

[4] 郝志峰,陈薇,蔡瑞初,等. 基于典型因果推断算法的无线网络性能优化[J]. 计算机应用, 2016, 36(8): 2114-2120.

[5] 赵森栋,刘挺. 因果关系及其在社会媒体上的应用研究综述[J]. 软件学报, 2014, 25(12): 2733-2752.

[6] 张晓阳. 基于大数据因果推断的大学生网购服装推荐模型研究[D]. 哈尔滨: 哈尔滨工业大学, 2016.

[7] KANO Y. Causal inference using non-normality[C]//Proceedings of the international symposium on science of modeling, the 30th anniversary of the information criterion. Tokyo, Japan: AIC; 2003: 261.

[8] COMLEY J W, DOWE D L. General Bayesian networks and asymmetric languages[C]//Hawaii international conference on statistics and related fields. Honolulu, Hawaii: [s. n.], 2003: 5-8.

[9] SUN X, JANZING D, SCHÖLKOPF B. Causal inference by choosing graphs with most plausible markov kernels[C]//Ninth international symposium on artificial intelligence and mathematics. Florida: [s. n.], 2006: 1-11.

[10] 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(10): 2003-2030.

[11] ZHANG K, HYVÄRINEN A. On the identifiability of the post-nonlinear causal model[C]//Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence. Canada: AUAI Press, 2009: 647-655.

[12] HOYER P O, JANZING D, MOOIJ J M, et al. Nonlinear causal discovery with additive noise models[C]//Advances in neural information processing systems. New York: Curran Associates, 2009: 689-696.

[13] JANZING D, SUN X, SCHÖLKOPF B. Distinguishing cause and effect via second order exponential models[EB/OL]. 2009. arXiv:0910.5561.

[14] DANIUSIS P, JANZING D, MOOIJ J, et al. Inferring deterministic causal relations[EB/OL]. 2010. arXiv:1203.3475.

[15] JANZING D, MOOIJ J, ZHANG K, et al. Information-geometric approach to inferring causal directions[J]. Artificial Intelligence, 2012, 182-183: 1-31.

[16] SGOURITSA E, JANZING D, HENNIG P, et al. Inference of cause and effect with unsupervised inverse regression[C]//Artificial intelligence and statistics. California: AISTATS, 2015: 847-855.

[17] MOOIJ J, 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.

[18] MARX A, VREEKEN J. Telling cause from effect using MDL-based local and global regression[C]//2017 IEEE international conference on data mining (ICDM). New Orleans: IEEE, 2017: 307-316.

[19] HU S, CHEN Z, NIA V P, et al. Causal inference and mechanism clustering of a mixture of additive noise models[C]//Advances in neural information processing systems. New York: Curran Associates, 2018: 5206-5216.