arXiv:2305.03662v1 [cs.AI] 5 May 2023

基于阶段变量的健康时间序列因果发现

Bharat Srikishan,Samantha Kleinberg计算机
科学系 史蒂文斯理工学院 新泽西州霍博肯
{bsrikish, sa
mantha.kleinberg}@stevens.edu

摘要

当无法进行随机实验时（例如在医疗保健领域），利用观察性数据学习因果关系至关重要。当因果关系会随着疾病进程而改变时（例如，某些药物在疾病早期或针对重症疾病患者最有效），从时间序列健康数据中发现因果关系就更具挑战性。妊娠周数、疾病阶段或糖化血红蛋白等生物标志物等阶段变量，可能影响哪些因果关系适用于患者。然而，由于数据量有限，通常无法在每个阶段内进行因果推断；而将所有数据合并，又可能导致推断错误或遗漏。为此，我们提出基于阶段变量的因果发现（CDSV），利用阶段变量对多个时间序列中的数据重新加权，同时考虑各阶段中不同的因果关系。在模拟数据中，与基线方法相比，CDSV发现的原因更多、假发现更少；在eICU中，其错误发现率低于基线方法；在MIMIC-III中，它发现了更多具有临床相关性的高血压原因。

1 引言

因果关系往往会随时间改变。在个体层面，促使你儿时积极活动的因素，可能与成年后的动机不同。在社会层面，新法律会改变企业和公民的行为。尽管当需要发现的关系不断变化时，因果推断似乎不可能实现，但这些关系往往会系统性地变化，尤其是在健康领域。许多疾病（例如癌症或慢性肾病 - CKD）会分期，因此可以将患者划分为严重程度相近的簇，并为其提供相似的治疗方法。同样，糖化血红蛋白（HbA1c）能够反映一段时间内的平均血糖；孕周用于帮助了解妊娠进程；简化急性生理学评分（SAPS II）则用于确定患者入重症监护病房（ICU）时的病情严重程度。我们将这类变量称为阶段变量，定义如下。

定义 1.1（阶段变量）。其取值随时间变化（无论是否单调），并且可能改变数据生成过程底层因果结构的变量。

这些指标可能会随时间单调变化（例如妊娠周数），也可能以更复杂的方式变化（例如糖化血红蛋白）。其中一些指标可能由医疗专业人员确定（如癌症分期），另一些则可能通过实验室检查识别（如肾功能）。它们的共同点是，个体的阶段可能发生变化，而且不同阶段的因果关系可能有所不同。例如，某一阶段有效的治疗方法在另一阶段可能无效；比如，相较于晚期阶段，某些药物在控制早期慢性肾病症状方面可能更有效 [Qaseem 等人，2013]。

目前尚无专门针对这类场景设计的因果关系学习方法。然而，忽略阶段意味着，最好的情况下，我们将无法知晓某种因果关系在何时最为有效（例如，不知道某种药物可能对哪些人有帮助）；最坏的情况下，则可能完全遗漏某个原因（例如，当某种因果关系仅在短时间内成立时）。相反，利用阶段信息可以用更少的数据实现更准确的因果推断。我们的目标是学习分阶段数据中的因果关系。

以及它们适用的具体阶段。此外，通过利用有关各阶段之间关系的现有知识，我们可以在数据难以获取或获取成本高昂的阶段进行因果推断。
从观测性时间序列数据中学习因果关系是一个广受研究的领域，包括格兰杰因果关系 [Granger、 1969], 动态贝叶斯网络（DBNs） [Murphy， 2002], 以及基于逻辑的方法[Kleinberg， 2013]等方法。尽管许多方法假设平稳性（即因果结构和/或参数不会随时间变化），但这对现实世界的应用造成了很大限制。此外，虽然动态贝叶斯网络已扩展到非平稳数据 [Husmeier 等，2010，Robinson 等， 2010], 但这无法发现阶段依赖关系，也不能跨阶段共享数据。然而，癌症治疗干预的成功与癌症的进展程度密切相关 [Greene 和 Sobin， 2008]。在 I期，癌症局限于局部，手术成功的可能性很高；而在 III期，癌症已扩散至淋巴结，因此更常采用化疗等治疗方法。尽管干预措施的疗效会随阶段发生系统性变化，现有方法却无法学习关系适用的条件，也无法利用其他阶段的数据（如果因果推断是在各阶段内分别进行的）。对于难以或成本高昂地收集数据、且每个时间范围内可用证据有限的应用而言，这一点至关重要。

为此，我们提出基于阶段变量的因果发现（CDSV）。该方法能够推断阶段依赖的因果关系，同时根据用户指定的加权函数在不同阶段之间迁移信息。我们利用关于不同阶段相似性的先验医学知识，通过加权函数引导信息迁移。因此，CDSV 尤其适用于医疗保健领域：某些阶段的样本很多，而其他阶段的样本很少（例如，生存率低或易于治疗、很少发展到后期的疾病）。我们证明，CDSV 在四个模拟数据集上的表现优于最先进方法；我们发现，它能够在 MIMIC-III 重症监护病房数据集中找出医学上合理的高血压（血压）原因 [Johnson et al.，2016], ；此外，我们还表明，使用 eICU 数据集查找低氧饱和度的原因时，CDSV 的错误发现率较低 [Pollard et al.，2018]。

我们的主要贡献如下：

- 一种用于分阶段时间序列数据因果发现的新方法，可学习阶段间阶段依赖的关系。
- 通过实证展示了在模拟和真实世界健康数据中捕捉阶段信息以减少假阳性的必要性。
- 应用于真实世界的 MIMIC-III 和 eICU 数据，分别发现不同重症监护室病情严重程度评分患者高血压和低氧饱和度的原因。

2 相关工作

我们从三个关键领域综述相关工作：基于观察性数据的因果推断、因果可迁移性以及阶段感知预测。

2.1 因果发现

许多因果模型学习方法基于贝叶斯网络（BN） [Pearl， 2000], 即有向无环图，其中节点表示变量，边表示因果关系。PC 算法 [Spirtes and Glymour， 1991] 是从数据中学习 BN 的广泛使用的方法，并且在具有稀疏关系的高维数据上表现良好 [Kalisch and Bühlman， 2007]。该算法通过 PCMCI [Runge et al.，2019] 扩展到时间序列数据，并通过 PCMCI+ [Runge， 2020], 增加了对同时发生关系的支持。FCI [Spirtes et al.， 2000]及其时间扩展 tsFCI [Entner and Hoyer， 2010], 同样通过重复进行条件独立性检验来识别因果结构，但不同于 PCMCI，它们允许存在潜变量。CD-NOD [Huang et al.， 2020] 将 PC 算法与非参数核结合，用于在非平稳数据中进行因果发现。该方法假设因果结构的变化由一个隐藏变量驱动，并使用基于核的条件独立性检验 [Zhang et al.， 2012], 其关于时间步数量的计算复杂度为立方级。尽管这些方法为因果发现提供了理论保证，但它们没有显式考虑阶段，并且大多数方法都假设平稳性。因此，它们会发现错误的因果关系或因果强度。

BN 也通过 DBN 扩展到了时间序列数据领域 [Murphy， 2002], 这类模型使用一系列跨时间连接的 BN，来表示某一时间的变量如何影响稍后时间的变量。然而，在 BN 和 DBN 中进行精确推断是不可行的，因此必须使用 MCMC 抽样和变分推断等近似方法。这些方法假设底层因果结构保持不变，而对于阶段性数据而言，这一假设并不成立。可以将阶段作为一个变量纳入模型，但这可能会违反既有假设。非平稳 DBN [Robinson and Hartemink， 2008]允许因果关系随时间变化，并已扩展用于支持连续数据、潜变量和变点检测 [Grzegorczyk andHusmeier, 2009, Shafiee Kamalabad and Grzegorczyk， 2020]。尽管这些

方法允许因果结构随时间变化，但图之间的转换受到限制：只能添加或删除单条边，或合并两张图。相比之下，我们允许阶段之间的因果结构发生任意变化，并进行精确推断。

时间序列因果推断的其他方法包括格兰杰因果关系和基于逻辑的方法。如果 X 的过去值能够提高对 Y 未来值的预测能力，则称 X 格兰杰导致 Y [格兰杰， 1969]。格兰杰因果关系已扩展至多变量数据，并支持非线性关系 [Ancona 等人，2004 年，Marinazzo 等人，2011]。然而，如果在每个阶段内分别进行推断，那么当某些阶段的数据较少时，该方法可能会遗漏原因；而在跨阶段合并数据时，则可能同时产生假阳性和假阴性。最后，Kleinberg [2013] 提出了一种方法，将因果关系表示为概率计算树逻辑公式 [Hansson 和 Jonsson， 1994], ，并利用因果显著性度量来区分因果关系与相关性。这使得该方法能够处理复杂关系，例如变量在一段时间内持续为真，或多个因素共同作用而产生某种效应。该方法为每种关系使用一个时间窗口（即 x 在 10-20 个时间单位内导致 y），而非单一的时间滞后，并且能够自动发现这一时间窗口。尽管该方法不支持非平稳关系，但我们将其扩展以构建 CDSV，并将在方法部分进一步讨论。

2.2 因果迁移

另一个相关领域是因果可迁移性，它定义了将一个数据集中的因果关系应用于另一个数据集所需满足的理论条件。对于包含阶段变量的数据，可以将数据按阶段划分为独立的数据集，并使用因果可迁移性方法推断跨阶段的因果关系。该领域的主要研究 [Bareinboim and Pearl, 2013, 2014, Lee and Honavar, 2013] 定义了将一个数据集中的治疗效应应用于另一个数据集所需的充分必要条件。Pearl 和 Bareinboim [2014] 引入了选择变量，这些变量作为结构因果模型（SCM）的补充，用于描述其方程在不同总体之间如何变化。在不同总体上使用结构因果模型时，需要对这些选择变量的不同取值进行条件化。指定这些变量需要完整地指定结构因果模型，而这又要求事先了解哪些变量会在不同数据集中发生变化。此外，这些方法并不直接适用于时间序列数据。相反，我们提出了一种仅需指定阶段变量的方法，而无需了解哪些变量及因果关系会在不同阶段之间发生变化。

2.3 分阶段健康预测

虽然阶段变量尚未应用于因果推断，但 StageNet [Gao 等人 2020] 将其引入健康风险预测模型，并用于 MIMIC-III 和终末期肾病数据集上的患者分型。这种基于深度学习的架构采用结合阶段感知卷积模块的长短期记忆网络（LSTM）来预测患者死亡风险。尽管与基线方法相比，该模型具有良好的预测性能，但它无法揭示预测结果背后的原因，而且需要大量样本，而在我们的研究场景中，某些阶段的数据量可能要少得多（例如晚期癌症）。

3 方法

我们介绍了一种用于时间序列数据因果发现的新方法，称为基于阶段变量的因果发现（CDSV）。CDSV 允许利用来自多个阶段的数据来学习因果结构，同时仍能够推断阶段依赖关系（例如，学习妊娠早期与妊娠后期饮食的不同原因）。我们通过在针对每个目标阶段进行因果推断时，对来自不同阶段的数据赋予不同权重来实现这一点。这使得信息能够在不同阶段之间迁移，有助于克服噪声的影响（例如，当阶段的划分模糊或不确定时），同时提供具有阶段依赖性的输出。

3.1 问题设定

我们首先考虑观测数据中的多变量数据$\mathbf{X}$，其中包含长度可变的时间序列 $\mathbb{R}^{p\times d_p\times t_p}$，其中 d_p 是变量的数量， t_p 是时间序列的时间步数 p。此外，我们假设已知每位患者的阶段变量$\mathbf{S}$。一个人可能在整个时间序列中都处于同一阶段，也可能随时间发生变化。

3.2 Kleinberg方法

接下来，我们旨在确定该时间序列中一组结果（效应）的成因。图 3 展示了这一设置的示意图。为此，我们扩展了 Kleinberg [2013] 的方法。在这种方法中，因果关系由

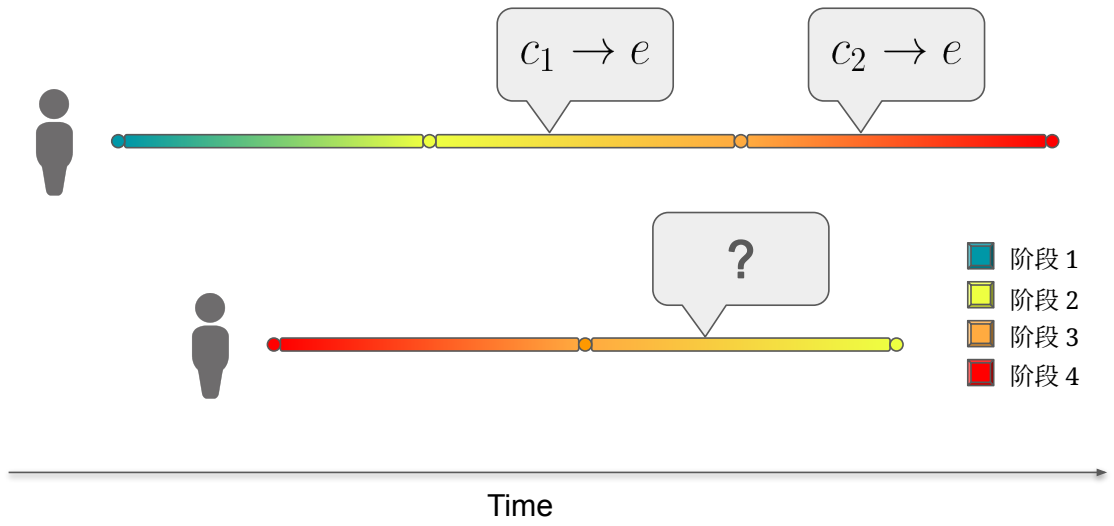


图 1：基于阶段的因果推断与迁移示例。每个人进入和结束的阶段各不相同。对于上方的人，我们对在阶段 2 和阶段 3 中学习到的因果关系很有把握，因此 CDSV 使用加权函数将这些知识迁移给第二个人，以便在阶段 3 中进行推断。圆圈表示阶段转换。

概率计算树逻辑（PCTL） [Hansson 和 Jonsson， 1994] 公式：[1]

$$c \rightsquigarrow_{m}^{\geq k, \leq l} e. \tag{1}$$

该公式表示，原因 c 会导致结果 e 在 k 到 l 个时间单位后发生，概率为 m。

例如，考虑这样一个问题：确定不同治疗方法对处于 4 个不同癌症阶段的患者的癌症缓解结果有何影响。每位患者 p 都有自己的时间序列，记录癌症缓解或进展的结果。每条时间序列包含的数据点数量不一，因为每位患者开始治疗的时间可能不同，而且每位患者以不同的速度经历不同的癌症阶段。

c对 e 的因果显著性定义如下：

$$\epsilon_{avg}(c, e) = \frac{\sum_{x \in X} P(e|c \wedge x) - P(e|\neg c \wedge x)}{|X \setminus c|} \tag{2}$$

$$P(e|c \wedge x) = \sum_{t \in T} \frac{\#(c_t, x_{i:j}, e_{k:l})}{\#(c_t, x_{i:j})} \tag{3}$$

请注意，上方的 # 表示其参数在指定时间窗口内出现的次数，其中 c在时间 t出现， x 在 e之前 i 到 j 个时间单位内出现，而 e 在 t之后 k 到 l 个时间单位内出现。 X 是潜在原因（会提高 e发生概率的变量）的集合。通过对数据中的所有实例，计算原因与结果同时发生的基于频率的概率，并减去没有原因时结果发生的基于频率的概率，该方法得出所有潜在原因的平均因果显著性。因果显著性ϵ_{avg} 高于某个阈值的潜在原因被视为显著原因。

Thi该方法需要满足两个主要假设，以确保所发现原因的正确性：

1. 观测变量之间的因果关系不会随时间变化。 2. 所有变量对的共同原因都必须包含在数据中。

在本文中，我们扩展了 Kleinberg [2013]，使其能够处理阶段依赖关系，同时允许跨阶段进行信息共享。此外，由于关系可能随阶段变量而变化，我们降低了平稳性要求。

[1]PCTL wa s，并为 Kleinberg 的导向公式添加了时间下界 [2013]。我们在此使用该版本 .

算法1基于阶段变量的因果显著性

对 $c \rightsquigarrow^{\geq k, \leq l} e$ 对于患者 p 在阶段 s **执行**
  对 每个阶段 $o \in S$ **执行**
    $\epsilon_o(c, e) = c$ 在阶段 o中的因果显著性
    $\epsilon_{avg}(c, e, s) += \epsilon_o(c, e) \cdot w(s, o)$
  结束循环
  如果 $\epsilon_{avg}(c, e, s) \geq \epsilon$ **则**
    $R[c, e, s] = \epsilon_{avg}(c, e, s)$
  结束条件。
结束循环
return R

3.3 CDSV 假设

为确保 CDSV 找到正确的因果关系，时间序列 $\mathbf{X}$ 中的因果关系必须在每个阶段内保持平稳。请注意，这一假设比要求整个时间序列都保持平稳更宽松。我们方法的一项主要优势是，因果关系可以在不同阶段之间发生变化。其次，我们要求满足因果忠实性，这意味着观测数据在每个阶段都遵循潜在的因果结构。第三，我们必须有一个由用户定义的加权函数，用于确定不同阶段的数据应占多大权重。最后，我们假设已知时间序列中每个时点对应的阶段变量值。

3.4 使用阶段变量的因果显著性

在 Kleinberg [2013] 的基础上，我们提出 CDSV 算法。CDSV 的输入包括时间序列数据 $\mathbf{X}$、阶段变量（例如癌症阶段、糖化血红蛋白或 SAPS II 评分）$\mathbf{S}$、一组变量（潜在原因与结果）、最小时间滞后、最大时间滞后、显著性阈值，以及加权函数 $w(s, o)$。最小和最大时间滞后定义了用于检验因果关系的时间窗口范围，而加权函数则定义了阶段之间如何关联以便在阶段间迁移。显著性阈值定义了因果显著性必须达到的大小，才能将其视为真实的因果关系。CDSV 返回时间序列中的显著原因与结果集合，以及相应的时间窗口和每个原因的强度。

CDSV 分为两个步骤。首先，它会考察每个阶段内的每一对原因与结果，同时排除其他潜在原因，并计算在给定原因及其他每个潜在原因的条件下，结果发生概率的平均差值。这一步与 Kleinberg [2013] 中的方法相同。随后，CDSV 将所得的平均差值乘以加权函数 $w(s, o)$，该函数返回阶段 s 与阶段 o 之间的权重，从而得到给定因果关系的因果显著性。

3.5 推断与计算复杂度

我们通过对下文等式所定义的相应时间窗口内相关事件的数量求和，来计算方法中的概率。此处，时间窗口定义为：c 发生于时间 t，x 发生于 e之前 i 至 j 个时间单位内，且 e 发生于 t之后 k 至 l 个时间单位内。

$$P_o(e|c \wedge x) = \sum_{t \in T} \frac{\#_o(c_t, x_{i:j}, e_{k:l})}{\#_o(c_t, x_{i:j})} \tag{4}$$

等式 下文的式（5）定义了我们计算的核心，而算法1定义了逐步的方法 d.

$$\epsilon_{avg}(c, e, s) = \sum_{o \in S} \frac{\sum_{x \in X} P_o(e|c \wedge x) - P_o(e|\neg c \wedge x)}{|X \setminus c|} \cdot w(s, o) \tag{5}$$

式（5）给出了阶段 s内原因 c 对效应 e 的重要性。此处，X 是 s 中 e的全部潜在原因集合（即所有会提高 e 概率的变量）。c 和所有其他 $x \in X$ 都具有如下形式的因果关系：$x \rightsquigarrow^{\geq i, \leq j} e$。

我们在实验中使用了两个加权函数：

$$w(s, o) = 1 - \frac{|s - o|}{T} \tag{6}$$

$$w(s,o)=\begin{cases}\left(1-\frac{|s-o|}{T}\right)\cdot\frac{N_o}{N_s} & \text{if } N_s\le\tau\\ 1 & \text{if } N_s>\tau, s=o\\ 0 & \text{if } N_s>\tau, s\neq o\end{cases} \tag{7}$$

加权函数 6 根据不同阶段之间的绝对差值，对其他阶段的数据进行线性加权。这里T表示阶段的数量，而来自其他阶段的数据 o 距离当前阶段 s越远，其权重就越低。例如，对于 3 个阶段， $w(1,1) = 1, w(1,2) = 2/3,$和 $w(1, 3) = 1/3$。

函数 7 与函数 6 类似，但会乘以额外的 $\frac{N_o}{N_s}$比率，该比率为阶段 o中的样本数除以阶段 s中的样本数。乘以该比率可确保样本较多的阶段比样本较少的阶段具有更高的权重。此外，我们定义了一个阈值 τ，当阶段 s 拥有 τ个数据点时，函数会收敛为 1 或 0。这保证了阶段 s中的样本数超过 τ时，只会使用该阶段的数据。有关无限数据极限下数据一致性的证明，请参见附录A。

请注意，上述加权函数适用于我们的实验，但我们的方法也可以使用任何加权函数。我们建议领域专家针对医疗保健中的不同问题确定合适的加权函数。有关如何确定加权函数的更多信息，请参见附录F。

我们的方法的计算复杂度为 $O(N^3T)$，其中有 N个变量， T为时间序列的长度。加权不会增加时间复杂度，因此这与 Kleinberg [2013]中的情况相同。由于每种关系都可以独立于其他关系进行检验，我们对实现进行了并行化，从而显著提升了性能。

3.6 寻找显著关系

在算法1的第6行中需要注意，只有当潜在因果关系大于或等于某个阈值 ϵ时，它才具有显著性。为了以一种有依据的方式选择该阈值，我们为每个因果显著性得分计算p值。根据Kleinberg [2013], , 我们知道，如果因果显著性值集合 R 服从均值为0的正态分布，那么就不会发现显著的因果关系。因此，为了计算 R中每个值的p值，我们对这些数据拟合一个正态分布。利用这一经验零分布，我们计算每种因果关系的p值，然后在错误发现率水平为0.05的条件下，使用Benjamini-Hochberg校正控制多重假设检验。

4 实验

我们在六个数据集上开展实验，以评估 CDSV。首先，我们在四个模拟数据集上进行评估：一个包含三个阶段和三个变量的简单结构；另一个包含两个阶段和三个变量，其中一个原因从正向作用转为负向作用；一个包含三个阶段和 10 个变量的复杂结构；以及最后一个在阶段 1 样本较少的相同复杂结构。在模拟数据上，我们使用受试者工作特征曲线下面积（AUROC）、假阳性率（FPR）和假阴性率（FNR），将推断结果与生成的真实因果关系及其真实发生时间进行比较，从而评估我们的方法与最先进方法（tsFCI、PCMCI 和 PCMCI+）。由于 CD-NOD 的时间复杂度为立方级，不适用于我们的长时间序列数据，因此我们不与之比较。作为真实世界测试，我们使用两个大型医院和重症监护病房数据集：eICU 和 MIMIC-III。在 MIMIC-III 上，我们将 CDSV 与基线方法进行比较，以找出高血压的原因；在 eICU 上，我们旨在找出低氧饱和度的原因，并与 PCMCI 和 PCMCI+进行比较。

4.1 模拟数据

4.1.1 简单模拟

该数据集用于测试因果关系随阶段推进（阶段 1 到阶段 2）而减弱，并最终在阶段 3 消失的情形。这种情况在现实中确实存在，例如慢性肾病（CKD）：在疾病早期，患者会接受药物治疗以缓解水肿并控制血压，但随着疾病进展，药物的疗效会逐渐减弱。到了阶段 5，慢性肾病药物已不再有效，需要进行透析或肾移植。

我们生成一个包含 3 个二元变量的三阶段数据集： x_1、 x_2, 以及 y。每个阶段生成 5 个时间序列，总计 15 个时间序列，每个时间序列包含 1000 个时间点。在这些数据中， x_2 是一个不与其他变量相互作用的噪声变量。在阶段 1， x_1 以 0.9 的概率在两个时间步后导致 y。在阶段 2，该概率降至 0.52。最后，在阶段 3 x_1 完全不会导致 y，而 y 则以 0.5 的概率自发出现。

我们通过对阶段进行线性加权来评估 CDSV，并将其与 Kleinberg [2013], 的基础方法进行比较；该方法在各个阶段内分别进行因果推断，且不在阶段之间共享信息。我们预期 CDSV 的表现最佳

方法	简单			反转关系			复杂的			复杂小样本		FNR
	曲线下面积	假阳性率	假阴性率	曲线下面积	假阳性率	假阴性率	曲线下面积	假阳性率	假阴性率	曲线下面积	假阳性率	
Kleinberg [2013]	0.57	0.16	1.00	**1.00**	**0.00**	**0.00**	0.52	0.07	0.80	0.00	1.00	1.00
CDSV（本研究）	**1.00**	**0.00**	**0.00**	**1.00**	**0.00**	**0.00**	**0.55**	**0.02**	0.74	**0.55**	**0.02**	**0.78**
PCMCI	0.50	0.67	1.00	0.29	0.57	1.00	0.00	1.00	1.00	0.00	1.00	1.00
PCMCI+	0.50	0.67	1.00	0.29	0.57	1.00	0.40	0.75	**0.61**	0.00	1.00	1.00
tsFCI	0.50	**0.00**	1.00	0.50	**0.00**	1.00	0.49	0.15	1.00	0.00	1.00	1.00

表1：模拟数据上的结果。曲线下面积（AUC）是受试者工作特征曲线下面积，数值 0.00 表示该方法未发现有效的因果关系。Kleinberg [2013]是基线方法，在每个阶段内分别推断原因。每列中的最佳值以粗体标示。请注意，我们在评估中同时考虑因果关系及其时间安排。

因为在第二阶段进行推断时，如果因果关系较弱，它可以借用第一阶段的数据（即利用阶段 1 的证据来补充阶段 2 的观测数据）。我们将其与 tsFCI、PCMCI 和 PCMCI+进行比较。选择这些方法是因为它们与我们的方法最为接近，且支持时间序列数据。为公平地与 CDSV 比较，我们将阶段变量作为另一项输入提供给所有基线方法。

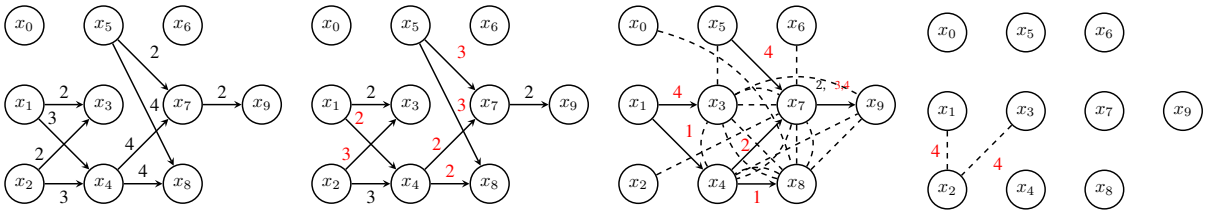
如表1所示，与 CDSV 相比，tsFCI、PCMCI 和 PCMCI+的表现较差。当错误发现率为 0.05 时，两个 PCMCI 基线方法都错误地发现了 x_1 在 3 个时间步后导致 y 的因果关系。它们还发现每个阶段都会导致 y，这是不正确的。tsFCI 的假阳性率低于 PC方法，但假阴性率仍然较高。作为进一步的基线，我们使用 Kleinberg [2013], 的因果推断方法，利用每个阶段的数据独立进行因果推断。在错误发现率为 0.05 时，该方法的表现略优于基线方法，正确发现了在阶段 1 中 x_1在 2 个时间步后导致 y，其 AUROC 为 0.57；然而，该方法未能发现 x_1 导致 y，尽管这是一种有效的因果关系。与 PCMCI 和 PCMCI+ 相比，CDSV、Kleinberg [2013], 和 tsFCI 的假阳性率都较低，但只有 CDSV 同时具有 0 的假阳性率和假阴性率。与 PCMCI+相比，这些方法在选择潜在关系作为真实关系时更加保守。tsFCI 之所以保守，是因为它假设可能存在潜变量，因此要求数据中具有较强的条件独立性。
CDSV 的表现最佳：它以 0.05 的错误发现率发现了阶段 1 和阶段 2 中所有正确的因果关系，且没有在阶段 3 中发现任何错误的因果关系。这些结果证实了我们的假设：CDSV 通过对阶段 1 的数据重新加权，使其更容易学习阶段 2 中较弱的因果关系。

4.1.2 翻转双阶段模拟

现在，我们考察某个原因在不同阶段中从正向作用转为负向作用时的表现。例如，二甲双胍是糖尿病的常见早期治疗方法，但根据患者所处的疾病阶段，它可能使糖尿病患者的血糖降低或升高 [Gormsen 等人， 2019]。为检验这种情况，我们生成一个包含 3 个二元变量的双阶段数据集： x_1、 x_2 和 y。在阶段 1 中， x_1 以 0.9 的概率使 y 取值为 1；而在阶段 $2\ x_1$ 中， y 以 0.9 的概率使 y 取值为 0。我们仍将 x_2 设为噪声变量，不影响其他变量。我们在每个阶段生成 5 条时间序列，每条包含 1000 个时间点。
如表1所示，CDSV 和 Kleinberg [2013] 均能正确发现阶段依赖关系，而其他方法则发现了错误的关系。PCMCI+ 和 tsFCI 未能发现任何正确的因果关系，因为它们将不同阶段的数据视为等同的数据，导致阶段 1 和阶段 2 因反转关系而相互抵消。由于 Kleinberg [2013] 方法分别推断各个阶段中的原因，因此也能够发现正确的关系。尽管 CDSV 采用了信息共享机制，但它仍然能够发现正确的关系，因为它对来自同一阶段的数据赋予了最高权重。

4.1.3 ComplexThreeStage模拟

为了测试我们的方法如何扩展，我们生成了另一个三阶段数据集，其中有 10 个变量，变量之间的因果关系也更多。我们创建 3 层，每层包含 3 个变量，最后一层包含一个结果变量。我们随机设置这些层之间的因果边以及每种关系的时间滞后。其结构如图 2(a) 所示。每个阶段中因果关系的权重均随机设置。阶段 1 的因果权重随机设置在 $[1, 2]$, , 阶段 2 的权重在 $[0.1, 0.5]$, , 阶段 3 的权重在 $[0.01, .05]$, , 以模拟随时间推移而减弱的关系。在向前生成时间序列时，我们使用逻辑函数 $\sigma(x) = 1/(1 + e^{-x})$ 将这些权重转换为概率，该函数会将所有输入映射到 $[0, 1]$ 概率空间。与实验1不同，阶段 3 中仍存在关系，但现在非常微弱。同样，我们令阶段 1 中的因果关系最强，阶段 2 中较弱，阶段 3 中最弱。为进一步增加复杂性，我们引入更多 FINAL



(a) 真实结构 (b) CDSV 结果 (c) PCMCI+结果 (d) tsFCI 结果

图2：我们复杂（10 个变量）模拟的真实结构 (a) 和推断出的因果关系 (b、c、d)。边上的注释表示每个关系的时间。实线表示正确的关系，虚线表示错误的关系。边上的红色数字表示推断出的时间不正确（尽管节点之间存在真实关系）。值得注意的是，CDSV 的所有错误都出现在时间判断上，而所有关系均被恢复，且没有假阳性。

关系之间的时间安排各不相同，在 $[1, 4]$ 中随机生成窗口。我们在每个阶段生成 5 个时间序列，每个时间序列包含 1000 个时间步，最终得到一组 15 个时间序列。

表1中的 AUROC、FPR 和 FNR 结果表明，该结构中的弱关系、更复杂的时间安排以及更多的因果关系，使所有方法都面临更大的挑战。尽管如此，CDSV 仍优于基线方法。PCMCI 在错误发现率水平为 0.05 时未能找到任何正确关系。PCMCI 表现不佳的一个原因是，它将阶段变量视为原因，这与我们的问题设定不符；此外，由于每个阶段的因果关系都会发生变化，PCMCI 的平稳性假设也不成立。我们的方法更为成功，因为它允许因果关系在各阶段之间发生变化，同时在阶段间迁移已学习的关系。我们还测试了不含阶段变量的 PCMCI+ 。虽然这样可以避免将阶段识别为原因，但仍会识别出许多虚假关系。

图 4.1.3 展示了真实结构、CDSV 的输出、未使用阶段信息的 PCMCI+ 输出，以及 tsFCI 的输出。CDSV 找到了所有正确的边，只有部分关系的时间判断存在小误差。我们严格评估 AUROC，将这类关系视为假发现；但值得注意的是，CDSV 找到了所有实际原因与结果之间的关系，且没有错误的边。相比之下，未使用阶段信息的 PCMCI+ 发现了许多虚假关系，并且未能恢复真实关系（例如 $x_2 \to x_3$）。这可能是因为 PCMCI+ 合并了不同阶段的数据，因而发现了更大比例的弱因果关系，导致识别出错误的因果边。tsFCI 比 PCMCI+更保守，但代价是未能找到任何正确关系。在假阳性率方面，CDSV 的值最低，而 Kleinberg [2013] 和 tsFCI 的假阳性率也较低。与 PCMCI+相比，这些方法在将潜在关系判定为真实关系时更加保守。请注意，CDSV、Kleinberg [2013], 和 tsFCI 的假阴性率高于 PCMCI+ ，因为当一种方法的假阳性率较低时，这是一种常见的统计权衡。

4.1.4 复杂小样本模拟

现在，我们测试一种情况：某个阶段的样本数量较多，而另一个阶段的样本数量较少。这种情况在医疗保健领域十分常见，尤其是在数据稀缺、获取成本高昂或难以收集的情况下。例如，由于 IV期癌症患者的预期寿命较短，癌症数据中 IV期患者的数量将少于 I期患者。

我们生成了一个数据集，其结构与图 4.1.3 相同，其中阶段 1 包含一个时间序列，阶段 2 和阶段 3 各包含五个时间序列。每个时间序列均生成 1000 个时间点，并且再次在阶段 1 中设置较强的因果关系，其强度在后续各阶段中逐渐降低。在本实验中，我们使用函数 7 作为一种与数据一致的加权函数，使样本数量较多的阶段中的数据获得更高的权重。

如表1所示，只有 CDSV 能够在错误发现率水平为 0.05 时，从该数据中发现任何真实关系。这体现了 CDSV 的一项关键优势：通过利用不同阶段之间的信息，克服单个阶段内的小样本量问题。

4.2 ICU 中低血氧饱和度的原因

4.2.1 数据

接下来，我们在 eICU 数据集 [Pollard et al., 2018], 上评估 CDSV、PCMCI 和 PCMCI+ 。该多中心数据库包含 20 多万例重症监护病房入院记录。对于住院时间为 1 – 10 天的每位重症监护病房患者，我们创建了按小时记录的时间序列数据。我们的目标是找出低氧（O_2）饱和度的原因，输入变量包括血管加压药、氧合治疗、心率、呼吸频率等。我们使用急性生理学与慢性健康状况评价（APACHE）IV 评分 [Zimmerman et al., 2006] 作为阶段变量。APACHE IV 是一种经过医学验证的评分系统，用于评估

因果关系	Time	阶段
胶体液快速输注	2至4小时低SAPS评分	
心率过低	1至2小时 低SAPS评分	

表2：CDSV在MIMIC数据中发现的导致血压升高的原因，错误发现率=0.05

重症监护病房患者的病情严重程度，并且该评分可在 eICU 数据集中获取。我们根据 APACHE 评分将患者分为两个区间：低（$0-70$）和高（≥ 70）。APACHE 评分越高，意味着患者病情越严重；我们再次使用经验分位数来划分评分类别。最终数据包含 70,151 名患者，其中最长的时间序列为 579 小时，最短的为 24 小时。实验详情请参见附录 D。

4.2.2 结果

在以错误发现率水平为0.05检验1至4小时之间的关系时，CDSV未发现低氧饱和度的显著原因。尽管我们假设氧合治疗会对未来的低氧饱和度产生负面影响，但由于该治疗在数据中出现频率较低，CDSV无法确信这一原因。在CDSV中不对p值进行阈值筛选而对因果显著性排序时，我们发现，在两个APACHE评分阶段中，正常氧饱和度都会在1至4小时后对低氧饱和度产生负面影响。
PCMCI和PCMCI+ 分别发现了65个和62个导致低氧饱和度的原因，错误发现率为0.05。这些原因大多不合理，但两种基线方法都发现，正常氧饱和度会在未来1至4小时对低氧饱和度产生负面影响。

虽然正常 O_2 并不会直接阻止低 O_2 饱和度，但由于数据中有许多患者的 O_2 正常，且从未出现低 O_2 饱和度，因此我们会看到这一结果。因此，CDSV 和基线方法都推断正常 O_2 对低 O_2, 有负面影响，这意味着正常 O_2 饱和度会降低未来 1-4 小时内出现低 O_2 的可能性。然而，CDSV 使用 FDR校正确认这一因果关系并不显著，而两种基线方法都认为这是显著的因果关系，这是不正确的。这表明，在相同的错误发现率水平下，CDSV 比其他方法更保守。

4.3 MIMIC-III 中高血压的原因

4.3.1 数据

作为第二项现实世界评估，我们在 MIMIC-III [Johnson 等人， 2016], 这一公开可用的重症监护数据集上测试所有方法。在这项实验中，我们旨在根据机械通气、血管加压药、胶体液快速输注、晶体液快速输注、无创通气、心率和呼吸频率这些输入变量，找出高血压（血压）的原因。我们选择这些变量，是因为它们会直接影响血压，且在 MIMIC-III 中缺失率较低。我们使用简化急性生理学评分（SAPS-II） [Le Gall 等人， 1993] 作为阶段变量。这是一种经过医学验证的重症监护病房患者病情严重程度评分，综合了 15 个变量，并在患者入院时计算。我们将每位患者分为 3 个阶段：SAPS-II评分较低（0-22）、SAPS-II评分中等（22-48）和 SAPS-II评分较高（48-163）。SAPS-II评分越高，表示患者的病情越严重。这些范围是根据所有患者评分分布的经验分位数确定的。

MIMIC-Extract项目 [Wang et al., 2020] 提供按小时记录的实验室测量值、生命体征以及包括血管加压药和机械通气在内的干预措施。MIMIC代码仓库项目 [Johnson et al., 2018] 同样提供用于标准化严重程度评分等概念的代码。在本实验中，我们结合了SAPS-II评分和上述项目的数据。最终提取的数据包含28,283名患者，每名患者对应29个变量。最长的时间序列包含240小时的数据，最短的时间序列则为24小时。

我们选择研究较为透彻的高血压结果，将 CDSV 与基线方法进行比较。我们假设，血管加压药和推注液体应在 1 到 4 小时内升高血压。我们使用 SAPS-II 作为阶段变量，以实验室检查结果和生命体征作为数据，并将高血压作为结果，应用 CDSV 和基线方法来学习因果关系。实验详情请参见附录 C。

4.3.2 结果

CDSV 在表 2 中以 0.05 的错误发现率水平找到了导致高血压的两个原因。找到的第一个原因是，胶体液推注会在 2 到 4 小时内导致高血压。胶体液快速输注是一种液体溶液，当患者循环系统中流失过多液体时会对其进行输注。快速输注会增加液体容量，从而升高血压并使身体恢复正常。

因果关系	影响时间/有效期	
高心率	1-4小时	+ ✓
低SAPS评分	1–4 小时	+ ✗
高呼吸频率	1-3 小时	+ ✓
正常血氧饱和度	1-4 小时	+ ✗
血管加压药	1–4 小时	- ✗
苯肾上腺素	1-4 小时	- ✗
去甲肾上腺素	1-4 小时	- ✗

表 3：在 MIMIC 数据中，PCMCI 和 PCMCI+ 在 FDR=0.05 下发现的导致血压升高的原因。负向关系会导致血压正常或降低。

恢复稳态。这种关系得到了医学文献的充分支持 [Lewis et al., 2018]。所识别出的第二种关系，即心率过低导致高血压，可能出现在患者接受β受体阻滞剂治疗[Rothwell et al., 2010] 或心肌组织增厚 [Karam et al., 1989]的情况下。

相比之下，tsFCI 没有发现高血压的原因。它只发现，给予肾上腺素是 4 小时后给予未来肾上腺素的原因。在数据中，它没有发现其他原因或结果。由于 tsFCI 假设可能存在隐藏的混杂因素，如果一种关系可以由一个共同父变量解释，则不会将其作为有效关系返回。MIMIC 中有许多条件独立的变量，因此 tsFCI 没有发现任何原因。

PCMCI 和 PCMCI+ 均在错误发现率水平为 0.05 时识别出 88 个原因。两种算法发现的得分最高的原因如表 3 所示。虽然其中一些原因具有合理性，但也存在许多错误的原因，例如血管加压药对高血压（BP）具有负面影响，而这一结论并未得到医学文献的支持。此外，由于基线方法将阶段变量视为一种可能的原因，因此它们得出低SAPS评分会导致高血压（BP）的结论，这在医学上并不合理。

PCMCI 和 PCMCI+ 将阶段变量视为潜在原因，这导致这些方法对许多可能无效的关系赋予过强的因果强度，即使在错误发现率阈值为 0.05 时也是如此。另一方面，CDSV 在相同的错误发现率水平下更加保守，识别出的原因集合规模更小，但在医学上更有依据。CDSV 也不假设因果关系在不同阶段之间保持稳定，而是能够学习每个阶段各自不同的因果关系。

5 局限性

CDSV 的主要局限在于需要准确的用户定义阶段变量。在许多数据集（如 MIMIC）中，确实存在这类变量，用于表示渐进性或分级条件（例如癌症）或其他事件（例如妊娠），但这些阶段信息可能稀疏、缺失或不精确。例如，糖化血红蛋白的测量可能并不频繁，或者部分患者缺少相关数据。缺乏准确的阶段信息时，CDSV 可能会发现无效的原因。未来需要研究如何自动获取此类信息，以及如何从数据中学习加权函数。虽然我们展示了线性加权函数具有良好的表现，但使用特定领域的函数或许可以进一步提升效果。最后，尽管我们的方法在真实世界数据上表现良好，但由于缺乏真实结构，这些数据给评估带来了挑战。未来，我们计划与临床医生合作，验证我们发现的因果关系。

6 结论

在许多健康应用中，因果关系会随阶段而变化，例如妊娠和重症监护室病情严重程度评分。然而，现有基于观测时间序列数据进行因果推断的方法，无法可靠地推断此类阶段依赖关系。有些方法会发现虚假因果关系，另一些则无法发现真实的因果关系。为弥补这一不足，我们提出基于阶段变量的因果发现（CDSV），该方法能够在阶段之间共享信息的同时，学习阶段依赖的因果关系。这样一来，既能发现更多真实的因果关系，也能避免虚假推断。

我们通过多种模拟证明，CDSV 的性能优于最先进方法（PCMCI、PCMCI+ 和 tsFCI）。在 eICU 数据集中发现低血氧饱和度的原因时，与 PCMCI 和 PCMCI+ 相比，CDSV 更善于消除假发现，而后两者均发现了错误的抑制性原因。在使用阶段变量 SAPS-II 评分对 MIMIC-III 进行真实世界应用，以寻找高血压的原因时，CDSV 找到了更多在医学上合理的原因，而 PCMCI 和 PCMCI+ 则发现了许多缺乏文献支持的原因，并且

tsFCI 未能找到血压的任何原因。CDSV 在因果关系会随疾病进程发生变化的现实世界健康应用中，能够实现更准确的因果发现。

参考文献

Nicola Ancona、Daniele Marinazzo 和 Sebastiano Stramaglia。利用径向基函数研究时间序列的非线性格兰杰因果关系。*Physical Review E*，70(5):056221，2004。Elias Bareinboim 和 Judea Pearl。判定实验结果可迁移性的一般算法。*Journal ofCausalInference*，1(1):107–134，2013。Elias Bareinboim 和 Judea Pearl。在有限实验条件下，多个环境之间的可迁移性：完备性结果。*NeurIPS*，2014。T Dahhan、M Jamil、A Al-Tarifi、N Abouchala 和 M Kherallah。APACHE IV评分系统在重症脓毒症患者中的验证及与APACHE II系统的比较。重症监护，13:1–2，2009。Doris Entner 和 Patrik O Hoyer。使用 FCI 从时间序列数据中进行因果发现。概率图模型，2010。Junyi Gao、Cao Xiao、Yasha Wang、Wen Tang、Lucas M. Glass 和 Jimeng Sun。StageNet：用于健康风险预测的阶段感知神经网络。载于*2020*年 *The Web Conference* 论文集，WWW '20，第530–540页，美国纽约州纽约市，2020年。计算机协会。ISBN 9781450370233。doi: 10.1145/3366423.3380136。URL https://doi.org/10.1145/3366423.3380136。Lars C Gormsen、Esben Søndergaard、Nana L Christensen、Kim Brøsen、Niels Jessen 和 Søren Nielsen。二甲双胍可增加非糖尿病个体及近期发病的2型糖尿病个体的内源性葡萄糖生成。*Diabetologia*，62(7):1251–1256，2019。Clive WJ Granger。利用计量经济学模型和互谱方法研究因果关系。*Econometrica:Journal of the Econometric Society*，第424–438页，1969年。Frederick L Greene 和 Leslie H Sobin。癌症分期：回顾性与前瞻性评估。*CA: a cancerjournal for clinicians*，58(3):180–190，2008。Marco Grzegorczyk 和 Dirk Husmeier。非平稳连续动态贝叶斯网络。载于*NeurIPS*，2009。Hans Hansson 和 Bengt Jonsson。关于时间与可靠性的推理逻辑。*Formal aspectsof computing*，6(5):512–535，1994。Biwei Huang、Kun Zhang、Jiji Zhang、Joseph Ramsey、Ruben Sanchez-Romero、Clark Glymour 和 Bernhard Schölkopf。从异质/非平稳数据中进行因果发现。*The Journal of Machine Learning Research*，21 (1):3482–3534，2020。Dirk Husmeier、Frank Dondelinger 和 Sophie Lebre。非齐次动态贝叶斯网络中的跨时间片段信息共享。载于*NeurIPS*，2010。Alistair EW Johnson、Tom J Pollard、Lu Shen、H Lehman Li-Wei、Mengling Feng、Mohammad Ghassemi、Benjamin Moody、Peter Szolovits、Leo Anthony Celi 和 Roger G Mark。MIMIC-III：一个可免费访问的重症监护数据库。*Scientific Data*，3(1):1–9，2016。Alistair EW Johnson、David J Stone、Leo A Celi 和 Tom J Pollard。MIMIC代码库：助力重症监护研究的可复现性。*JAMIA*，25(1):32–39，2018。Markus Kalisch 和 Peter Bühlman。使用 PC算法估计高维有向无环图。*JMLR*，8(3)，2007。Roger Karam、Harry M Lever 和 Bernadine P Healy。高血压性肥厚型心肌病，还是伴高血压的肥厚型心肌病？一项针对78名患者的研究。*Journal of the AmericanCollegeof Cardiology*，13(3):580–584，1989。Samantha Kleinberg。因果性、概率与时间。剑桥大学出版社，2013年。Jean-Roger Le Gall、Stanley Lemeshow 和 Fabienne Saulnier。基于欧洲/北美多中心研究的新型简化急性生理学评分（SAPS II）。*JAMA*，270(24):2957–2963，1993。Sanghack Lee 和 Vasant Honavar。m-可迁移性：因果效应在多个环境间的可迁移性。载于*AAAI*，2013。Sharon R Lewis、Michael W Pritchard、David JW Evans、Andrew R Butler、Phil Alderson、Andrew F Smith 和 Ian Roberts。胶体液与晶体液用于危重症患者液体复苏的比较。*Cochrane*系统综述数据库，8(8)，2018。

Daniele Marinazzo、Wei Liao、Huafu Chen 和 Sebastiano Stramaglia。基于格兰杰因果关系的非线性连通性。*Neuroimage*，58(2):330–338，2011。Kevin Patrick Murphy。《动态贝叶斯网络：表示、推断与学习》。博士论文，加利福尼亚大学伯克利分校，2002。Judea Pearl。《因果关系：模型、推理与推断》。剑桥大学出版社，英国剑桥，2000。Judea Pearl 和 Elias Bareinboim。外部效度：从 do演算到总体间的可迁移性。*StatisticalScience*，29(4):579–595，2014。Tom J Pollard、Alistair EW Johnson、Jesse D Raffa、Leo A Celi、Roger G Mark 和 Omar Badawi。eICU协作研究数据库：一个可免费使用的重症监护研究多中心数据库。*Scientific Data*，5 (1):1–13，2018。Amir Qaseem、Robert H Hopkins Jr、Donna E Sweet、Melissa Starkey 和 Paul Shekelle。1至3期慢性肾病的筛查、监测与治疗：美国内科医师学会的临床实践指南。*Annals of Internal Medicine*，159(12):835–847，2013。Joshua Robinson 和 Alexander Hartemink。非平稳动态贝叶斯网络。*Advances in Neural Information Processing Systems*，21，2008。Joshua W Robinson、Alexander J Hartemink 和 Zoubin Ghahramani。学习非平稳动态贝叶斯网络。*JMLR*，11(12)，2010。Peter M Rothwell、Sally C Howard、Eamon Dolan、Eoin O'Brien、Joanna E Dobson、Bjorn Dahlöf、Neil R Poulter、Peter S Sever 等。β 阻滞剂和钙通道阻滞剂对个体内血压变异性及卒中风险的影响。*The Lancet Neurology*，9(5):469–480，2010。Jakob Runge。在自相关非线性时间序列数据集中发现同期及滞后因果关系。载于*UAI*，2020。Jakob Runge、Peer Nowack、Marlene Kretschmer、Seth Flaxman 和 Dino Sejdinovic。在大型非线性时间序列数据集中检测并量化因果关联。*Science Advances*，5(11):eaau4996，2019。Mahdi Shafiee Kamalabad 和 Marco Grzegorczyk。具有边级顺序耦合参数的非齐次动态贝叶斯网络。*Bioinformatics*，36(4):1198–1207，2020。Peter Spirtes 和 Clark Glymour。一种快速恢复稀疏因果图的算法。*Social Science ComputerReview*，9(1):62–72，1991。Peter Spirtes、Clark N Glymour 和 Richard Scheines。《因果、预测与搜索》。麻省理工学院出版社，2000。Shirly Wang、Matthew BA McDermott、Geeticka Chauhan、Marzyeh Ghassemi、Michael C Hughes 和 Tristan Naumann。MIMIC-Extract：面向MIMIC-III的数据提取、预处理与表示流程。载于*ACM*健康、推断与学习会议（*CHIL*）论文集，第222–235页，2020。Kun Zhang、Jonas Peters、Dominik Janzing 和 Bernhard Schölkopf。基于核的条件独立性检验及其在因果发现中的应用。*arXiv*预印本 *arXiv:1202.3775*，2012。Jack E Zimmerman、Andrew A Kramer、Douglas S McNair 和 Fern M Malila。急性生理与慢性健康评估（APACHE）IV：当今危重患者的医院死亡率评估。*Critical Care Medicine*，34(5):1297–1310，2006。

CDSV 数据一致性证明

定理 1 (加权函数数据一致性)。 给定原因 c、结果 e，以及 M 个阶段（每个阶段的样本数为 $\{N_1, \ldots, N_M\}$），加权函数 $\mathcal{T}$ 将使阶段 s 中的因果显著性：$\epsilon_{avg}(c, e, s)$ 收敛到其真实的未加权值 $\epsilon_s(c, e)$，当 $\{N_1, \ldots, N_M\} \to \infty$ 时。

证明。 根据 CDSV 算法1，阶段 o 的未加权因果显著性为：

$$\epsilon_o(c, e) = \frac{\sum_{x \in X} P(e|c \wedge x) - P(e|\neg c \wedge x)}{|X \setminus c|}$$

首先，计算每个阶段的未加权因果显著性：

$$\epsilon_1(c, e), \epsilon_2(c, e), \ldots, \epsilon_M(c, e)$$

现在计算阶段 s的 CDSV 平均因果显著性:

$$\epsilon_{avg}(c,e,s) = \epsilon_1(c,e)w(s,1) + \epsilon_2(c,e)w(s,2) + \\ \ldots + \epsilon_M(c,e)w(s,M)$$

$$w(s,o) = \begin{cases} \left(1 - \frac{|s-o|}{T}\right) \cdot \frac{N_o}{N_s} & \text{if } N_s \leq \tau \\ 1 & \text{if } N_s > \tau, s = o \\ 0 & \text{if } N_s > \tau, s \neq o \end{cases}$$

取该平均因果显著性的极限，其中 $\{N_1, \ldots, N_M\} \to \infty$:

$$\begin{aligned} &\lim_{\{N_1,\ldots,N_M\}\to\infty} \epsilon_{avg}(c,e,s) \\ =& \lim_{\{N_1,\ldots,N_M\}\to\infty} \epsilon_1(c,e)w(s,1) + \epsilon_2(c,e)w(s,2) \\ &+ \ldots + \epsilon_M(c,e)w(s,M) \\ =& \epsilon_1(c,e) \lim_{N_1,N_s\to\infty} w(s,1) + \epsilon_2(c,e) \lim_{N_2,N_s\to\infty} w(s,2) \\ &+ \ldots + \epsilon_M(c,e) \lim_{N_M,N_s\to\infty} w(s,M) \end{aligned}$$

注意，对于任意阶段 s和不同阶段 o，当 $N_s \to \infty$, $w(s,o) = 0$时，因为 $N_s > \tau$。当 $s = o$时，则$w(s,o) = 1$，因为 $N_s \to \infty$。结果极限收敛到 $\epsilon_s(c,e)$:

$$\begin{aligned} &\epsilon_1(c,e) \lim_{N_1,N_s\to\infty} w(s,1) + \epsilon_2(c,e) \lim_{N_2,N_s\to\infty} w(s,2) \\ &+ \ldots + \epsilon_M(c,e) \lim_{N_M,N_s\to\infty} w(s,M) \\ =& \epsilon_1(c,e) \cdot 0 + \epsilon_2(c,e) \cdot 0 + \ldots + \epsilon_s(c,e) \cdot 1 \\ &+ \ldots + \epsilon_M(c,e) \cdot 0 = \epsilon_s(c,e) \end{aligned}$$

□

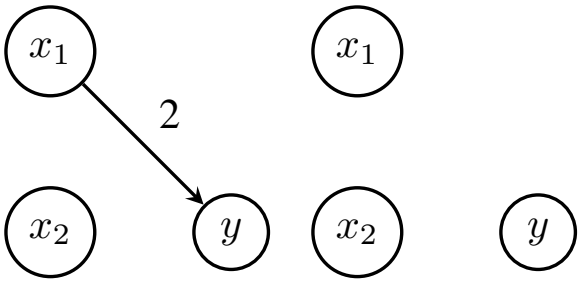
B 方法参数

B.1 简单模拟

- 结构：图 3
- 数据：每个阶段包含 5 个时间序列，共 1000 个时间点
- 原因 = $\{x_1, x_2\}$
- 结果 = $\{y\}$
- 阶段 = $\{1, 2, 3\}$
- 时间窗口 = (1、4)
- 错误发现率水平 = 0.05
- PCMCI条件独立性检验：偏相关检验

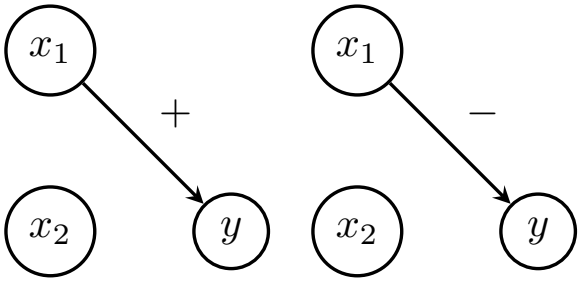
B.2 翻转两阶段模拟

- 结构：图4
- 数据：每个阶段包含5条时间序列，每条有1000个时间点
- 原因 = $\{x_1, x_2\}$
- 结果 = $\{y\}$
- 阶段 = $\{1, 2\}$
- 时间窗口 = (1,4)
- 错误发现率水平 = 0.05
- PCMCI条件独立性检验：偏相关检验



(a) 阶段 1 和阶段 2 (b) 阶段 3

图 3：简单三阶段模拟结构



(a) 阶段 1 (b) 阶段 2

图4：翻转双阶段模拟结构

B.3 复杂三阶段模拟

- 结构：图5
- 数据：每个阶段包含 5 个时间序列，每个时间序列有 1000 个时间点
- 原因 $= \{x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$
- 结果 $= \{x_3, x_4, x_7, x_8, x_9\}$
- 阶段 $= \{1, 2, 3\}$
- 时间窗口 = (1，4)
- 错误发现率水平 = 0.05
- PCMCI条件独立性检验：偏相关检验

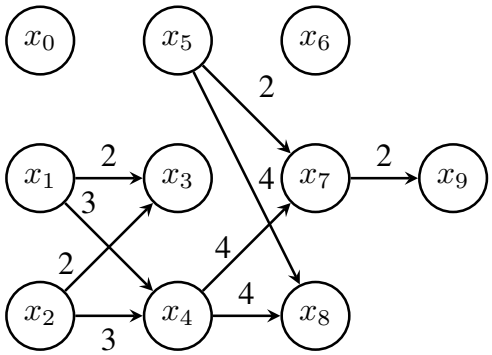


图5：复杂三阶段模拟结构。该模拟在三个阶段中具有相同的结构，其中阶段 1 的关系最强，阶段 2 较弱，阶段 3 最弱。

B.4 ComplexSmallSample 仿真

- 结构：图5
- 数据：阶段 1 有 1 条时间序列，包含 1000 个时间点；阶段 2 有 5 条时间序列，包含 1000 个时间点；阶段 3 有 5 条时间序列，包含 1000 个时间点
- 原因 $= \{x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$
- 结果 $= \{x_3, x_4, x_7, x_8, x_9\}$
- 阶段 $= \{1, 2, 3\}$
- 时间窗口 = (1,4)
- 错误发现率水平 = 0.05
- PCMCI条件独立性检验：偏相关检验

C MIMIC 高血压实验详情

C.1 原因、结果和参数

- 原因 = {呼吸机、血管加压药、腺苷、多巴酚丁胺、多巴胺、肾上腺素、异丙肾上腺素、米力农、去甲肾上腺素、苯肾上腺素、加压素、胶体液快速输注、晶体液快速输注、无创通气、高心率、心率正常、心率过低、高呼吸频率、呼吸频率正常、呼吸频率过低、血氧饱和度正常、低氧饱和度}
- 结果 = 高血压
- 阶段 = 高SAPS-II评分、中SAPS-II评分、低SAPS-II评分
- 时间窗口 = (1,4) 小时
- 错误发现率水平 = 0.05
- PCMCI条件独立性检验：偏相关检验

C.2 高血压实验变量类别

为了对本实验中的每个变量进行分类，我们使用经验分位数。我们将 $0-20\%$ 的数据定义为低值， $20-80\%$ 定义为正常/中等值， $80-100\%$ 定义为高值。下面给出了每个感兴趣变量的取值范围：

- 平均血压： $0-60, 60-100, >100$
- SAPS-II评分： $0-22, 22-48, 48-163$
- 心率： $0-60, 60-110, >110$
- 呼吸频率： $0-12, 12-21, >21$

D eICU低氧饱和度实验详情

D.1 原因、结果和参数

- 原因 = {血管加压药、氧合、高中心静脉压、中心静脉压正常、低中心静脉压、高心率、心率正常、心率过低、高呼吸频率、呼吸频率正常、呼吸频率过低、血氧饱和度正常、高体温、体温正常、低体温、高血压、血压正常、低血压}
- 结果 = 低氧饱和度
- 阶段 =高 APACHE IV评分，低 APACHE IV评分
- 时间窗口 = (1,4) 小时
- 错误发现率水平 = 0.05
- PCMCI条件独立性检验：偏相关检验

D.2 低氧饱和度实验变量类别

为了对本实验中的各变量进行分类，我们使用经验分位数。我们将 $0-20\%$ 的数据定义为低， $20-80\%$ 定义为正常/中等， $80-100\%$ 定义为高。下面提供了各个关注变量的范围：

- APACHE IV评分： $0-70, \geq 70$
- 中心静脉压： $0-6, 6-20, > 20$
- 心率： $0-70, 70-101, > 101$
- 呼吸频率： $0-15, 15-25, > 25$
- 血氧饱和度： $0-94.5, \geq 94.5$
- 温度： $0-36.6, 36.6-37.9, \geq 37.9$
- 总血压： $0-163, 163-211, > 211$

E 相关工作比较

	阶段依赖 因果关系	阶段之间的迁移 阶段	精确 推断	先验概率 不要求	因果图 非必需。
SCM	✗	✗	✓	✗	✗
DBN	✗	✗	✗	✗	✗
nsDBN	✓	✓	✗	✗	✗
PCMCI	✗	✗	✓	✓	✓
PCMCI+	✗	✗	✓	✓	✓
tsFCI	✗	✗	✓	✓	✓
Kleinberg 2013	✗	✗	✓	✓	✓
CDSV	✓	✓	✓	✓	✓

表 4：相关方法及其特点的比较。CDSV 是唯一符合我们期望的方法。

在表 4 中，我们将 CDSV 与相关方法进行了比较。请注意，只有 CDSV 支持阶段依赖的因果关系，同时也满足我们的其他要求。

F 加权函数设计

在确定加权函数时，我们建议遵循以下准则：

1. 查阅有关所研究健康状况/疾病的文献，寻找经医学验证的状态测量指标。 2. 根据上述测量指标的计算方法，确定该状况/疾病的动态特征。 3. 一旦了解疾病的动态特征及其相应指标，即可选择适当的加权函数。

上述过程在医疗保健领域通常很直接，因为孕周、糖化血红蛋白、癌症阶段等许多健康指标都已得到广泛研究和验证。

例如，在真实数据实验中，我们研究了重症监护病房中的不良事件，例如高血压。由此，我们找到了 SAPS II评分和 APACHE IV评分，这两者都是经过临床验证的重症监护室病情严重程度评分。由于这些评分已用于评估患者病情严重程度和死亡风险，并经过校准[Dahhan 等人， 2009],我们将患者划分为不同的评分组，并用线性加权函数将其关联起来。