arXiv:2602.02623v1 [cs.LG] 2 Feb 2026

学习一致因果抽象网络

Gabriele D'Acunto[1,2],*Paolo Di Lorenzo*[1,2],*Sergio Barbarossa*[1]

[1]罗马第一大学信息工程、电子与电信系，意大利罗马[2]国家大学间电信联盟（CNIT），意大利帕尔马 电子邮箱：{gabriele.dacunto, paolo.dilorenzo,sergio.barbarossa}@uniroma1.it

摘要

因果人工智能旨在借助结构因果模型（SCM）提升人工智能的可解释性、可信度和鲁棒性。在这一探索中，近期进展对因果知识的网络层和余层进行了形式化。沿着这一方向，我们致力于学习一致的因果抽象网络（CAN）。这是一种层理论框架，其中*(i)*结构因果模型服从高斯分布，*(ii)* 限制映射是遵循语义嵌入原理的构造性线性因果抽象（CA）的转置，并且*(iii)*边茎在忽略置换的情况下对应于更详细结构因果模型的节点茎。我们的问题表述分解为针对各条边的局部黎曼问题，并避免了非凸目标函数。我们提出了一种高效的搜索过程，使用 SPECTRAL 求解局部问题。SPECTRAL 是一种具有闭式更新的迭代方法，适用于正定和半正定协方差矩阵。在合成数据上的实验表明，该方法在 CA 学习任务中表现出竞争力，并成功恢复了多种 CAN 结构。

索引词— 因果抽象、因果人工智能、网络层、斯蒂费尔流形、结构因果模型。

1. 引言

因果人工智能 [1] 是人工智能中一种新兴范式，旨在实现可解释性、可信性以及对领域偏移和分布偏移的鲁棒性。其核心是结构因果模型（SCM，[2]），其中因果变量通过因果机制加以描述，即来自其他变量的传入因果关系。借助同时支持干预与反事实推理，结构因果模型支撑了三个层次，即*Pearl*因果层级 [3]——观察、行动与想象——从而超越纯粹预测性和特定任务的人工智能。相应地，结构因果模型在因果变量上诱导出一组观测、干预与反事实概率测度，下文称为因果知识（CK）。

在这一追求中，并考虑到对智能体人工智能至关重要的协作维度，[4], D'Acunto和Battiloro [5]提出"结构因果模型是对世界主观且不完美的表征，不能与其所置身的关系网络相分离"。这将标准设定加以推广：在该设定中，主体——例如人工智能智能体、个人、卫星——被假定与真实值结构因果模型一一对应，并能直接访问它。此外，结构因果模型内在地绑定于一种视角，该视角由其在关系网络中的位置决定。主观因果知识通过干预一致因果抽象（CA）[6, 7] 映射及其对偶在网络中传播，最终引出相对因果知识的定义：换言之，"主体B视角下主体A的因果知识"。

上述构造通过函子性构造形式化，这些构造称为*CK* 的网络层与余层 [5]。这里，CK——以概率测度的凸空间表示 [8]——被关联到网络的节点和边，而仿射可测映射则对应于节点-边关联关系。前者称为节点茎和边茎，后者称为限制映射（用于层）和延拓映射（用于余层）。这一新颖的函子性框架有望应用于人工智能系统：其中，全局 *CK*会随着因果人工智能智能体的交互而归纳地涌现；反过来，智能体层面的局部 CK 也可以由系统演绎地塑造。我们的工作朝这一方向推进，重点是学习该框架的一个特定实例，即因果抽象网络（CAN，[9]）。

相关工作。 我们在以下设定下开展研究：[10], 具体而言，我们不假设：*(i)* 结构因果模型的完整规格；*(ii)* 已知其因果图；*(iii)* 已知其函数形式；*(iv)* 可获得干预数据；或 *(v)*联合采样的观察数据。与 [5], 一样，我们将CA视为结构因果模型之间的映射，并采用α-抽象框架 [7]。具体而言，我们聚焦于构造性线性因果抽象(CLCA，[6, 11])，并遵循语义嵌入原理(SEP，[10]其非正式地指出，嵌入低层级时必须保留高层级的CK。我们的问题表述基于2.1中的定理 [10],，该定理聚焦于层理论框架之外的CLCA学习。最后，[9]是本文的扩展版本，其中引入了CAN框架并研究其理论性质。

贡献。 给定一组来自主观结构因果模型的高斯测度，我们提出了一致的CAN学习问题。我们的问题表述可拆分为针对每条边的局部问题，对应于CLCA学习任务。针对这些任务，我们提出一种替代 [10], 的方法，该方法避免了非凸目标函数，并明确刻画最优CLCAs的集合。我们的解决方案是一种高效的搜索过程，基于定理2.1 以及CLCA的合成性。该过程采用黎曼方法SPEC-TRAL，通过闭式更新迭代求解局部问题。与[10], 中的方法不同，SPECTRAL适用于正定和半正定协方差矩阵。后者对于全局截面尤为重要：全局截面是赋给网络各节点的概率测度的一致赋值——无论是观察性还是干预性——且不会违反局部规则 [5]。在合成数据上的实验表明，在利用正定协方差学习限制映射时，SPECTRAL的表现与 [10] 中的方法相当。此外，当观察到全局截面时，也就是说，当输入为半正定协方差时，我们的搜索过程能够成功恢复不同的CAN结构。

我们的工作朝着实用地学习因果抽象网络迈出了一步，并有望应用于智能体型人工智能系统和协作式人工智能系统。

2. 因果抽象网络简介

本节介绍一致的因果抽象网络的记号和关键概念。有关全面的论述，请参阅 [9]。

记号。 从 1 到 n的整数集合为 $[n]$。大小为 n的零向量为 $\mathbf{0}_n$。大小为 $n \times n$的单位矩阵为 $\mathbf{I}_n$。$\mathrm{R}^{n\times n}$上的正定矩阵集合为 $\mathcal{S}^n_{++}$。斯蒂费尔流形为 $\mathrm{St}(\ell, h) := \{\mathbf{V} \in \mathrm{R}^{\ell\times h} \mid \mathbf{V}^\top\mathbf{V} = \mathbf{I}_h\}$. 哈达玛积为 $\odot$。$\mathbf{A}$ 是$\mathbf{A}$按列向量化的结果。

一致因果抽象网络。我们考虑一组 N 主观结构因果模型，记为 $\mathcal{K}$，它们以不同的粒度层次对同一现象进行建模。例如，不同的神经科学家使用不同的脑图谱研究有效连接 [12]。每个结构因果模型M^ℓ 所包含的 ℓ 因果变量 $\mathcal{X}^\ell = \{X_1, \ldots, X_\ell\}$数量各不相同——这些变量是由图谱划分出的感兴趣区域（ROI）——并且我们假设它诱导出一个观测零均值高斯测度 $\chi^\ell \sim N(\mathbf{0}_\ell,\ \mathbf{\Sigma}^\ell)$——即来自各感兴趣区域信号的联合概率密度。因此，我们称该结构因果模型为高斯模型。

我们假设这些结构因果模型通过 CLCA 相互关联。具体而言，考虑两个结构因果模型 M^ℓ 和 M^h，前者为低层级（细粒度）模型，后者为高层（粗粒度抽象）模型。CLCA 是一对映射，记为 $\langle \mathbf{B}_{h\ell}, \mathbf{V}_{h\ell}\rangle$，其中 （$i$） $\mathbf{B}_{h\ell} \in \{0,1\}^{h\times\ell}$对低层级变量$\mathcal{X}^\ell$到高层 $\mathcal{X}^h$的结构性不相交划分进行编码； （ii）$\mathbf{V}_{h\ell} \in \mathrm{R}^{h\times\ell}$是一个线性映射，将低层级变量的结果映射为高层结果，其非零项由$\mathbf{B}_{h\ell}$中的 1 决定。这里我们关注性质特别良好的 CA，即满足 SEP 的 CA。SEP 意味着，从高层模型 M^h 到低层模型 M^ℓ，再抽象回 M^h，能够完美重构 CK。对于 CLCA，SEP 与斯蒂费尔流形的几何性质相关，这意味着$\mathbf{V}_{\ell h} \in \mathrm{St}(\ell,\ h)$，其中$\mathbf{V}_{\ell h} = \mathbf{V}^\top_{h\ell}$。 此外，它还为高斯测度导出一个必要条件，这是奥斯特罗夫斯基定理在矩形矩阵上的应用 [13]。

定理2.1 （引自 [10]）。 设 $\chi^\ell \sim N\left(\mathbf{0}_\ell,\ \mathbf{\Sigma}^\ell\right)$， $\chi^h \sim N\left(\mathbf{0}_h,\ \mathbf{\Sigma}^h\right)$，其中 $\mathbf{\Sigma}^\ell \in \mathcal{S}^\ell_{++}$ 且 $\mathbf{\Sigma}^h \in \mathcal{S}^h_{++}$。用 $0 < \lambda_1 \leq \ldots \leq \lambda_\ell$ 表示 $\mathbf{\Sigma}^\ell$的特征值，用 $0 < \kappa_1 \leq \ldots \leq \kappa_h$ 表示 $\mathbf{\Sigma}^h$的特征值。若存在一个符合*SEP*的*CLCA* $\mathbf{V}$，从 χ^ℓ 到 χ^h，则

$$\lambda_i \leq \kappa_i \leq \lambda_{i+\ell-h}, \quad \forall\, i \in [h]\,. \tag{1}$$

由 $\mathcal{K}$出发，我们构建一个偏序集，其定义为 $\ell \leqslant h$ 当且仅当 M^h 是M^ℓ 的一个符合SEP的CLCA。鉴于因果抽象的组合性 [9]，，为求简约，我们考虑该偏序集的传递约简，仅保留不可约 关系，即 $\ell \leqslant h$，使得不存在满足$\ell \leqslant m \leqslant h$的 m。CK的动力学通过一个一致的CAN来建模，同时刻画抽象与嵌入。

定义2.2 （一致因果抽象网络 [9]）。 一致 CAN G 是一个定向无向图，具有： （i）对应于由高斯结构因果模型 $\mathrm{M}^\ell \in \mathcal{K}$ 导出的概率测度凸空间的节点；以及 （ii）对于每条由 $\ell \leqslant h$ 确定的边 $e_{\ell h}$，当从 M^h 转移到 M^ℓ 时，存在一个线性嵌入映射 $\mathbf{V}_{\ell h} \in \mathrm{St}(\ell,\ h)$；当从 M^ℓ 抽象到 M^h 时，存在一个 CLCA $\mathbf{V}_{h\ell} = \mathbf{V}^\top_{\ell h}$。 G 的定向遵循嵌入方向。

之所以称 CAN 为一致，是因为$\mathbf{V}_{\ell h} \in \mathrm{St}(\ell,\ h)$意味着沿每个有向环 $\mathrm{M}^h \xrightarrow{\text{嵌入}} \mathrm{M}^\ell \xrightarrow{\text{抽象}} \mathrm{M}^h$ [9]的限制映射复合满足交换性。结合相关组合拉普拉斯算子的非空核，一致性是全局截面 [9] 存在的必要条件。这一结果对全局截面的统计性质有直接影响；全局截面是概率测度 $\chi := \{\chi_1, \ldots, \chi_N\}$ 的一致赋值，将其赋给网络中的每个节点，且不违反局部规则 [5, 9]。

推论 2.3。 设 χ 为一个全局截面。则，其元素 $\chi_i \sim N(\mathbf{0}_{d_i}, \mathbf{\Sigma}^{d_i})$、$i \in [N]$的协方差矩阵具有相同的 $k \leq$ 最小$_i\, d_i = h$ 个非零特征值。

由于篇幅所限，证明从略，但可见 [9]。类似于图信号处理，定义相对于一致的 G 的 [14], 平滑性 。直观而言，若对于每条边 $e_{\ell h}$： $\ell \sim h$，较粗粒度的 M^h 近似为较精细的 M^ℓ的 CLCA，则一组结构因果模型相对于 G 是平滑的。按照 [10], ，平滑性可通过低层级概率测度 χ^ℓ 的抽象对高层 χ^h的逼近程度来衡量。具体而言，记$D^{\mathrm{KL}}_\alpha(\mu||\nu)$为 μ 与由映射 α 作用于 ν所诱导的推前测度之间的 Kullback–Leibler（KL）散度，则将$D^{\mathrm{KL}}_{\mathbf{V}_{h\ell}}\left(\chi^h||\chi^\ell\right)$对 $e_{\ell h} \in \mathcal{E}$ 求和，作为平滑性的度量。

3. 学习因果抽象网 RKS

我们的输入是一组 $\chi = \{\chi_1, \ldots, \chi_N\}$ 高斯概率测度，按维数降序排列，由N 假设关于$\mathbb{G}$光滑的结构因果模型所诱导。根据定义2.2，学习一致的CAN相当于学习*(i)* 决定其边集的CA 关系 $\mathcal{E}$，以及*(ii)* 每个 $\{\mathbf{V}^\top_{\ell h}\}$ 对应的CLCA $e_{\ell h} \in \mathcal{E}$，其中$\mathbf{V}_{\ell h} \in \mathrm{St}(\ell, h)$。令 $\mathcal{F}$ 为满足定理2.1中CLCA必要条件的边集。那么，我们的学习问题为

$$\min_{\{\mathbf{V}_{\ell h} \in \mathrm{St}(\ell,h)\}} \sum_{e_{\ell h} \in \mathcal{F}} D^{\mathrm{KL}}_{\mathbf{V}^\top_{\ell h}}\left(\chi^h||\chi^\ell\right)\,. \tag{P1}$$

值得注意的是，使局部KL散度为零的CLCA集合$\mathbf{V}^\top_{\ell h}$ 确定了所学得的$\mathbb{G}$的边集。根据构造$\mathbf{V}_{\ell h}$ 只有相对于某个CA才有意义，因此（P1）的解就是后者对应的CLCA集合。问题（P1）可按边分解，且当 [10] 和 $\mathbf{\Sigma}^\ell$ 均为 $\mathbf{\Sigma}^h$ 正定时，可使用2.3提出的算法求解每个局部问题。然而，还必须考虑具有半正定协方差的全局截面（参见推论P1），因此现有方法必须加以推广。此外，即使在正定的情形下，（P1）仍因若干原因而颇具挑战性。首先，对 $\mathcal{F}$的搜索需要求解多个CA学习问题，至多 $N(N-1)/2$个。随着 N 增大，这会带来高昂的计算开销；此外，每个局部问题都是非凸的，必须使用不同的随机初始化多次求解，才能获得可靠结果 [10]。其次，考虑零均值高斯分布，并以 r_h 表示 $\mathbf{\Sigma}^h$的秩，则KL散度

$$\begin{aligned} D^{\mathrm{KL}}_{\mathbf{V}^\top_{\ell h}}\left(\chi^h||\chi^\ell\right) &= \mathrm{Tr}\left\{\left(\mathbf{V}^\top_{\ell h}\mathbf{\Sigma}^\ell\mathbf{V}_{\ell h}\right)^\dagger \mathbf{\Sigma}^h\right\} + \\ &+ \log \mathrm{gdet}\left\{\mathbf{V}^\top_{\ell h}\mathbf{\Sigma}^\ell\mathbf{V}_{\ell h}\right\} - \log \mathrm{gdet}\left\{\mathbf{\Sigma}^h\right\} - r_h\,, \end{aligned} \tag{2}$$

在$\mathbf{V}_{\ell h}$ 中是非凸的，且优化成本高昂 [10]。第三，斯蒂费尔流形对$\mathbf{V}_{\ell h}$施加了非凸的正交约束。

为应对这些挑战，我们提出了一种高效的搜索策略，利用 CA 的合成性 [9]。此外，基于定理2.1，我们为局部问题引入了一种替代性问题形式 [10]，以避免非凸且成本高昂的目标函数，并提出了适用于正定和半正定情形的相关 SPECTRAL 算法。全文中，我们假设已充分了解 CA 结构$\mathbf{B}_{h\ell}$，尽管这仍然是一个具有挑战性的设定 [10]。

搜索过程。 考虑 χ如上所述。对每个 $\mathbf{\Sigma}^{d_i},\ i \in [N]$，计算特征值。我们将可能的 CA 关系根据定理 2.1 编码为一个严格下三角二元矩阵 $\mathbf{P} \in \{0,1\}^{N\times N}$。我们并非应用定理 2.1 $N\ (N-1)\ /2$次来填充 $\mathbf{P}$，而是利用 CA 的合成性。具体而言，我们首先将定理 2.1 应用于 $N-1$ 个 SCM 对，即第一条次对角线上的相邻项，也就是链 $\chi_1 - \chi_2 - \ldots - \chi_N$中的相邻节点。然后，我们将 1 赋给$\mathbf{P}$ 中对应于满足

定理 2.1 的那些对，并将其余条目留空。随后，我们将 $\mathbf{P}$ 设为其传递闭包。接下来，我们转到第二条次对角线。在此阶段，最多只需检验 $N-2$ 个 SCM 对，因为某些条目可能已在上一步通过传递闭包被填充。更新之后，我们再次应用传递闭包。我们对所有剩余的次对角线重复此过程。

一旦构造出$\mathbf{P}$，我们就求解与其非零元素对应的局部问题。在这里，也可以利用传递闭包来减少需要求解的子问题数量。考虑一个初始为空的 $N \times N$ 二值矩阵$\mathbf{A}$。我们首先求解$\mathbf{P}$第一条次对角线上的非零元素所对应的局部问题；每当 CLCA 成立，也就是满足局部问题的停止准则时，就将$\mathbf{A}$ 中相应的元素设为 1。随后计算$\mathbf{A}$ 的传递闭包，并从$\mathbf{P}$ 中移除已由该闭包蕴含的元素。然后对后续次对角线迭代应用同样的步骤。根据构造，学习得到的$\mathbf{A}$ 是组合性所蕴含的所有可能 CA 关系的传递约简。

局部问题。现在我们转向单条边 $e_{\ell h}$对应的 CA 学习局部问题的另一种表述。由于 $e_{\ell h} \in \mathcal{F}$，因此定理 2.1 成立，并且 M^{ℓ} 与 M^{h} 之间可能存在 CLCA。令 $\mathrm{V}_{\ell h} = \mathrm{B} \odot \mathrm{V}$，其中 $\mathrm{B} = \mathrm{B}^{\top}_{h\ell}$。由于我们考虑的是零均值高斯分布，输入为 $\mathbf{\Sigma}^{\ell}$ 和 $\mathbf{\Sigma}^{h}$，它们通常是半正定的，秩分别为 r_ℓ 和 r_h。若 CLCA 存在，则任意满足以下条件的 $\mathrm{V} \in \mathrm{St}(\ell, h)$

$$(\mathbf{B} \odot \mathbf{V})^{\top} \mathbf{\Sigma}^{\ell} (\mathbf{B} \odot \mathbf{V}) = \mathbf{\Sigma}^{h}\,, \tag{3}$$

对于边 $e_{\ell h}$是最优的，也就是说，它使式（2）中的局部 KL 散度为零。考虑特征分解

$$\mathbf{\Sigma}^{\ell} = \mathbf{U}_{\ell,+} \Lambda_{\ell,+} \mathbf{U}^{\top}_{\ell,+}\,, \text{ and } \mathbf{\Sigma}^{h} = \mathbf{U}_{h,+} \Lambda_{h,+} \mathbf{U}^{\top}_{h,+}\,. \tag{4}$$

此外，定义矩阵 $\mathbf{A} = \mathbf{U}_{\ell,+} \Lambda^{\frac{1}{2}}_{\ell,+} \in \mathbb{R}^{\ell \times r_\ell}$和$\mathbf{C} = \mathbf{U}_{h,+} \Lambda^{\frac{1}{2}}_{h,+} \in \mathbb{R}^{h \times r_h}$。将式（4）代入式（3），并在左侧乘以 $\mathbf{C}^{\top}$、在右侧乘以 $\mathbf{C}$，得到条件

$$\mathbf{T}^{\top}\mathbf{T} = \mathbf{I}_{r_h}\,, \text{ with } \mathbf{T} = \mathbf{A}^{\top}(\mathbf{B} \odot \mathbf{V})\mathbf{C}\,. \tag{5}$$

因此，任何满足结构 $\mathbf{B}$ 且满足式（5）的 $\mathbf{V} \in \mathrm{St}(\ell, h)$ 都会产生一个最优的 CLCA。此外，由 $\mathbf{T}$ 的表达式可直接看出，只要 $\mathbf{\Sigma}^{\ell}$经 CLCA $(\mathbf{B} \odot \mathbf{V})^{\top}$投影后的主轴与 $\mathbf{\Sigma}^{h}$的主轴对齐，式（5）就成立。因此，我们处理以下可行性问题：

$$\begin{aligned} &\text{find} \quad \mathbf{V}, \mathbf{T} \in \mathrm{St}(\ell, h) \times \mathrm{St}(r_\ell, r_h) \\ &\text{such that} \quad \mathbf{T} - \mathbf{A}^{\top}(\mathbf{B} \odot \mathbf{V})\mathbf{C} = \mathbf{0}_{r_\ell \times r_h}\,. \end{aligned} \tag{SP1}$$

SPECTRAL 方法。处理（SP1）中的等式约束颇具挑战，因为 $\mathbf{V}$ 和 $\mathbf{T}$ 分别位于维数不同的斯蒂费尔流形上，而这些流形是非凸的。为此，我们利用正交约束拆分（SOC，[15]）背后的原理。具体而言，我们将 $\mathbf{V}$ 放宽至欧几里得空间 $\mathbb{R}^{\ell \times h}$，并通过拆分变量 $\mathbf{Y} \in \mathrm{St}(\ell, h)$施加其正交性，满足

$$\mathbf{Y} - \mathbf{B} \odot \mathbf{V} = \mathbf{0}_{\ell \times h}\,. \tag{6}$$

结合方程 (5) 和 (6)，并引入缩放对偶变量$\mathbf{\Psi} \in \mathbb{R}^{\ell \times h}$ 和 $\mathbf{\Upsilon} \in \mathbb{R}^{r_\ell \times r_h}$，得到 (SP1) 的增广拉格朗日函数：

$$\begin{aligned} \mathcal{L}(\mathbf{V}, \mathbf{Y}, \mathbf{T}, \mathbf{\Upsilon}, \mathbf{\Psi}) = \frac{1}{2}\|\mathbf{B} \odot \mathbf{V} - \mathbf{Y} + \mathbf{\Psi}\|^2_{\mathrm{F}} + \\ + \frac{1}{2}\left\|\mathbf{T} - \mathbf{A}^{\top}(\mathbf{B} \odot \mathbf{V})\mathbf{C} + \mathbf{\Upsilon}\right\|^2_{\mathrm{F}}. \end{aligned} \tag{7}$$

随后，我们通过交替方向乘子法 [16]迭代地最小化方程 (7) 中关于原始变量的部分，并最大化关于对偶变量的部分。值得注意的是，原始变量可以通过闭式更新。对偶变量则是原始残差的标准累积和。

对 $\mathbf{V}$ 的更新。考虑迭代 k。需要求解的子问题为

$$\begin{aligned} \mathbf{V}^{k+1} = \arg\min_{\mathbf{V}} \frac{1}{2}\left\|\mathbf{B} \odot \mathbf{V} - \mathbf{Y}^{k} + \mathbf{\Psi}^{k}\right\|^2_{\mathrm{F}} + \\ + \frac{1}{2}\left\|\mathbf{T}^{k} - \mathbf{A}^{\top}(\mathbf{B} \odot \mathbf{V})\mathbf{C} + \mathbf{\Upsilon}^{k}\right\|^2_{\mathrm{F}}. \end{aligned} \tag{8}$$

解可以用闭式向量化形式表示。

引理 3.1。设 $\mathcal{B} := \{i \in [\ell h] \mid b_i = 1, b_i \in B\}$, $\mathbf{K} = \mathbf{C} \otimes \mathbf{A}$，且 $\mathbf{b} = \mathbf{Y}^{k} - \mathbf{\Psi}^{k} \pm \mathbf{A}\mathbf{T}^{k}\mathbf{C}^{\top} - \mathbf{A}\mathbf{\Upsilon}^{k}\mathbf{C}^{\top}$。则，

$$\underline{\mathbf{V}}_{\mathcal{B}} = \left(\mathbf{I}_{|\mathcal{B}|} + \mathbf{K}_{\mathcal{B}}\mathbf{K}^{\top}_{\mathcal{B}}\right)^{-1} \mathbf{b}_{\mathcal{B}}\,. \tag{9}$$

证明。对式（7）关于$\mathbf{V}$ 求导，并利用克罗内克积进行向量化，可得驻点条件

$$\mathbf{0}_{\ell h} = \left(\mathbf{D} + \mathbf{D}\mathbf{K}(\mathbf{D}\mathbf{K})^{\top}\right)\underline{\mathbf{V}} - \mathbf{D}\mathbf{b}\,, \tag{10}$$

其中 $\mathbf{D} = \mathrm{diag}(\mathbf{B})$。只有满足 $\mathbf{D}_{ii} = 1$ 的项在 $\mathbf{V}$中处于激活状态；其余项均为零。将范围限制在 $\mathcal{B}$ 中的索引，可得式（9）。 □

对 $\mathbf{Y}$ 的更新。需要求解的子问题为

$$\begin{aligned} \mathbf{Y}^{k+1} = \arg\min_{\mathbf{Y}} \frac{1}{2}\left\|\mathbf{B} \odot \mathbf{V}^{k+1} - \mathbf{Y} + \mathbf{\Psi}^{k}\right\|^2_{\mathrm{F}} = \\ = \mathrm{prox}_{\mathrm{St}(\ell,h)}(\mathbf{B} \odot \mathbf{V}^{k+1} + \mathbf{\Psi}^{k})\,; \end{aligned} \tag{11}$$

其中，$\mathrm{prox}^{\mathrm{St}(\ell,h)}(\mathbf{S})$ 由$\mathbf{U}$ 给出，它是$\mathbf{S}$ [10]的极分解中的因子。

对$\mathbf{T}$的更新。与$\mathbf{Y}$类似，我们得到

$$\begin{aligned} \mathbf{T}^{k+1} = \arg\min_{\mathbf{T}} \frac{1}{2}\left\|\mathbf{T} - \mathbf{A}^{\top}(\mathbf{B} \odot \mathbf{V}^{k+1})\mathbf{C} + \mathbf{\Upsilon}^{k}\right\|^2_{\mathrm{F}} = \\ = \mathrm{prox}_{\mathrm{St}(r_\ell, r_h)}(\mathbf{A}^{\top}(\mathbf{B} \odot \mathbf{V}^{k+1})\mathbf{C} + \mathbf{\Upsilon}^{k})\,. \end{aligned} \tag{12}$$

按照 [16], ，通过原始残差和对偶残差评估经验收敛；当 $k \to \infty$ 时，这些残差必须趋于零。详情参见 [9] 。

计算成本比较。 SPECTRAL 的主要优势在于能够进行闭式更新。从渐近角度看，每次迭代中成本最高的步骤是式（$\mathbf{V}$ （9）中对$\mathcal{O}(\ell^3)$次浮点运算的更新。相比之下， [10] 中的方法成本为 $\mathcal{O}(K\ell^3)$，其中 K 是相应的内迭代次数，每次迭代渐近需要 $\mathcal{O}(\ell^3)$ 次运算，以求解$\mathbf{V}$子问题。

4. 实证评估

为了通过实证评估所提出学习方法的性能，我们首先测试局部求解器 SPECTRAL，然后探讨不同 CAN结构的学习。

局部问题上的性能。给定低层和高层模型作为输入，即 M^{ℓ} 和 M^{h}，目标是在已知 $\mathbf{B}$的全部先验知识的条件下，恢复 CLCA矩阵 $\mathbf{V}^{\top}$。我们将 SPEC-TRAL 与近期在 [10]中提出的 CLinSEPAL、LinSEPAL-ADMM 和 LinSEPAL-PG 方法进行比较。这些基线方法针对正定矩阵的情形，最小化式（2）中的 KL散度，并采用不同的优化框架（参见 [10]）。

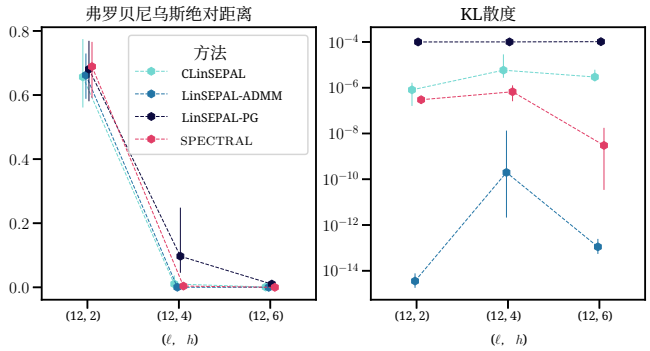


图 1. 在所有设置下求解局部问题的合成结果 $(\ell,\ h)$ 来自 [10]。

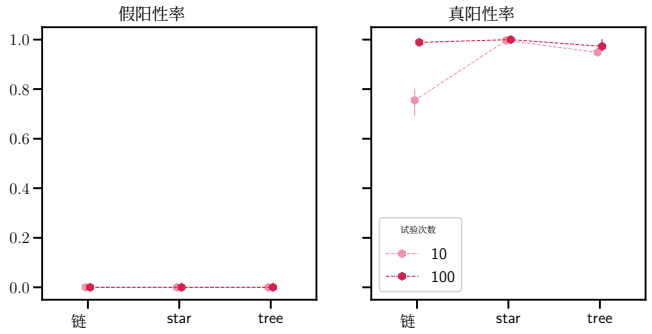


图 2。所提出的搜索过程在三种 CAN结构上的假阳性率（左）和真阳性率（右）。

我们在 [10] 全先验设置中使用的同一合成数据集上测试了这些方法。该数据集包含三种配置（$\ell,\ h$）$\in$ {(12，2)、(12，4)、(12，6)}，分别对应高、中高和中等抽象层级。有关数据生成过程的详细信息，请参阅 [10]。为评估性能，我们考察：(i) 构造性，即所学习的 CA 中具有构造性的比例；*(ii)* $D^{\mathrm{KL}}_{\widehat{\mathbf{V}}^\top}\left(\chi^h||\chi^\ell\right)$，用于量化对齐；*(iii)* 估计值 $\mathbf{V}$ 与真实值 $\mathbf{V}^\star$之间经归一化的绝对 Frobenius 距离；以及*(iv)* 结构干预一致性的 F1分数。

图 1展示了各测试方法的性能——由于篇幅所限，我们省略了构造性和 F1分数——。竖线表示 30 次模拟的四分位距。所有方法在所有模拟中均始终得到 CLCA，且 F1分数均为 1，因此恢复了由 $\mathbf{B}$ 编码的真实值结构。弗罗贝尼乌斯距离表明，在粗粒化程度较高时，识别 $\mathbf{V}^\star$（符号不计）更具挑战性，因为满足（$\mathbf{V}$）的集合会扩大。最后，这些方法也实现了良好的对齐。

总而言之，这些结果表明 SPECTRAL 在学习 CLCA 方面的性能与 [10] 中的方法相当，其输入为正定协方差，同时避免了非凸且代价高昂的目标函数，并提供闭式更新。

CAN 学习性能。接下来，我们转向 CAN 的学习。我们设定 $N = 10$，并为每个 $\mathbf{M}^\ell$分配随机维度 $\ell \in [2, 2N]$。随后，我们将结构因果模型附加到三种图结构的节点上：*(i)* 其完整 CAN（传递闭包）为全连接图的链；*(ii)*与完整 CAN 相同的星形图；*(iii)* 其完整 CAN 会形成更稠密树的树。对于传递闭包中的每个 CLCA $\ell \leqslant h$，我们随机生成 $\mathbf{B}_{\ell h} \in \{0,1\}^{\ell\times h}$，并相应地采样$\mathbf{V}_{\ell h} \in \mathrm{St}(\ell,\ h)$。

所有考虑的结构都存在全局截面，可从最粗的 χ^h（参见 [9]）开始迭代生成。对于每种结构，我们生成 $S = 30$ 个全局截面，并使用 SPECTRAL 求解每个局部问题，运行第 3 节所述的搜索过程。由于存在非凸性，每个局部问题都通过使用不同的初始化运行 SPECTRAL 至多 ntrials $\in \{10, 100\}$ 次来求解。如果在.....内实现了残差收敛

经过 ntrials 次运行后，我们停止并将 $\mathbf{A}$ 中对应的元素设为 1。在实验中，我们将残差容差设为 10^{-3}，并将最大迭代次数限制为 $K = 1000$。性能根据相对于完整 CAN 的假阳性率 （FPR）和真阳性率 （TPR）进行评估。根据推论 2.3，节点 χ_1、.....、χ_9 的协方差均为半正定，且具有相同的非零特征值，这些特征值对应于 χ_{10} 的谱。因此，只有 SPECTRAL 能够解决局部 CA 学习问题。这也是一种极端情形

因为定理 2.1 的必要条件对所有 $N(N-1)/2$ SCM 对都成立。结果，矩阵 $\mathbf{P}$ 的严格下三角部分中的所有元素均为 1，因此无法提供任何信息来减少局部 CA 学习问题的数量。图 2 展示了结果。ntrials 的不同取值以红色色调表示，竖线表示 S 次模拟的四分位距。尽管定理 2.1 成立，但 FPR（左图）证实，当真实值中不存在 CLCA 时，SPECTRAL 从未将其识别出来。这一结论适用于所有设置和 ntrials 取值。相反，TPR（右图）取决于 ntrials。正如预期的那样，由于非凸性，初始化很重要，因为残差可能无法在最大 K 次迭代内收敛。总体而言，结果证实，在所有设置下，当 ntrials $= 100$ 时，采用 SPECTRAL 的搜索过程都能实现近乎最优的 TPR。

5. 结论与未来工作

本文朝着协作式因果人工智能系统的原则性基础迈进了一步。我们研究了一致的CAN的学习，这是一个层理论框架，其中主观高斯结构因果模型通过对CLCA的转置进行对齐并粘合在一起。我们的问题表述避免了对非凸目标函数的优化，其解是一种高效的搜索过程，使用具有闭式更新的局部黎曼方法SPECTRAL。与以往方法不同，SPECTRAL既适用于正定协方差矩阵，也适用于半正定协方差矩阵，从而即使在我们观测到全局截面时也能恢复CAN结构。在合成数据上的实验证实了我们方法的有效性，其在CA学习中的表现与现有方法相当，并成功检索了CAN结构。

由于篇幅所限，本文省略了许多理论结果，但这些结果详见本文扩展版 [9]。值得注意的是，这些结果凸显了拓扑的影响：在某些情况下，拓扑会导致全局截面的存在受到阻碍。针对给定拓扑，开发能够恢复出诱导相同 CA 集合的最稀疏结构，并对其进行最小修改以支持全局截面的方法，是未来的一项重要工作。另一个重要方面是真实 CLCA 的可识别性。在 [10], 的一般设定下，我们只能通过结构先验保证所学习映射的结构干预一致性。未来，确保反事实一致性 [17] 也十分重要，尽管这可能需要提供干预数据和反事实数据。最后，超越合成设置和混合设置，对于充分评估 CAN 在现实场景中的潜力至关重要。这也是层理论框架 [5]普遍面临的一个关键问题，尤其是因为现实应用仍然十分匮乏。

致谢。本研究得到智能网络与服务联合事业项目 6G-GOALS [18] 的资助。该项目由欧盟地平线计划提供支持，资助协议编号为 101139232；同时，本研究还得到华为法国技术简化股份有限公司资助，资助编号为 Tg20250616041；以及欧盟下一代欧盟意大利国家复苏与韧性计划下的“未来电信”合作项目（PE00000001—RESTART 项目）的资助。

6. 参考文献

[1] Elias Bareinboim，《因果人工智能》，见 https://causalai-book.net/，2025，草稿版本。（引用于 1）[2] Judea Pearl，《因果关系》，剑桥大学出版社，2009。（引用于 1）[3] Judea Pearl 和 Dana Mackenzie，《为什么之书：因果关系的新科学》，Basic Books，2018。（引用于 1）[4] Deepak Bhaskar Acharya、Karthigeyan Kuppan 和 B. Divya，“智能体人工智能：面向复杂目标的自主智能——综合综述”，IEEE Access，第 13 卷，18936，2025，18912，1，第 – 页。（引用于 ）[5] Gabriele D’Acunto 和 Claudio Battiloro，“因果知识的相对性”，载于《第四十一届人工智能不确定性会议论文集》，Silvia Chiappa 和 Sara Magliacane 编，2025 年 7 月 21—25 日，机器学习研究论文集第 286 卷，第 863–881 页，PMLR。（引用于 1、2、4）[6] Sander Beckers 和 Joseph Y Halpern，“抽象因果模型”，载于《AAAI人工智能会议论文集》，2019 年，第 33 卷，第 2678–2685 页。（引用于 1）[7] Eigil Fjeldgren Rischel，《因果模型的范畴论》，硕士论文，哥本哈根大学，2020 年。（引用于 1）[8] Tobias Fritz，“凸空间 I：定义与示例”，arXiv 预印本 arXiv:0903.5522，2009。（引用于 1）[9] Gabriele D’Acunto、Paolo Di Lorenzo 和 Sergio Barbarossa，“因果抽象网络：层理论框架”，arXiv 预印本 arXiv:2509.25236，2026。（引用于 1、2、3、4）

[10] Gabriele D’Acunto、Fabio Massimo Zennaro、Yorgos Felekis 和 Paolo Di Lorenzo，“基于语义嵌入原理的因果抽象学习”，载于第四十二届国际机器学习会议，2025 年。（引用于 1、2、3、4）[11] Riccardo Massidda、Sara Magliacane 和 Davide Bacciu，“线性结构因果模型的因果抽象学习”，载于《第四十届人工智能不确定性会议论文集》，Negar Kiyavash 和 Joris M. Mooij 编，2024 年 7 月 15—19 日，机器学习研究论文集第 244 卷，第 2486–2515 页，PMLR。（引用于 1）[12] Karl J Friston，“功能连接与有效连接：综述”，Brain Connectivity，第 1 卷，第 1 期，第 13–36 页，2011 年。（引用于 2）[13] Nicholas J Higham 和 Sheung Hun Cheng，“修改优化问题中出现的矩阵惯性”，Linear Algebra and its Applications，第 275 卷，第 261–279 页，1998 年。（引用于 2）[14] Antonio Ortega、Pascal Frossard、Jelena Kova˘cevi´c、Jos´e MF Moura 和 Pierre Vandergheynst，“图信号处理：概述、挑战与应用”，Proceedings of the IEEE，第 106 卷，第 5 期，第 808–828 页，2018 年。（引用于 2）[15] Rongjie Lai 和 Stanley Osher，“一种用于正交约束问题的分裂方法”，Journal of Scientific Computing，第 58 卷，第 431–449 页，2014 年。（引用于 3）[16] Stephen Boyd、Neal Parikh、Eric Chu、Borja Peleato、Jonathan Eckstein 等，“通过交替方向乘子法进行分布式优化与统计学习”，Foundations and Trends® in Machine Learning，第 3 卷，122，2011，1，1，3，第 – 页。（引用于 ）[17] Kevin Xia 和 Elias Bareinboim，“神经因果抽象”，载于《AAAI人工智能会议论文集》，2024 年，第 38 卷，第 20585–20595 页。（引用于 4）[18] Emilio Calvanese Strinati、Paolo Di Lorenzo、Vincenzo Scian-calepore、Adnan Aijaz、Marios Kountouris、Deniz G¨und¨uz、Petar Popovski、Mohamed Sana、Photios A Stavrou、Beatriz Soret 等，“6G人工智能原生网络中的目标导向与语义通信：6G-GOALS方法”，载于《2024年欧洲网络与通信联合会议暨6G峰会》（EuCNC/6G峰会）。IEEE，2024 年，第 1–6 页。（引用于 5）