不对称性标记有向无环图的学习与解释：以 COVID-19 恐惧为例

Manuele Leonelli，西班牙马德里 IE大学科学与技术学院；
Gherardo Varando，西班牙巴伦西亚大学图像处理实验室

arXiv:2301.00629v1 [cs.AI] 2 Jan 2023

2023年1月3日

摘要

贝叶斯网络广泛用于学习和推理离散变量的依赖结构。然而，它们只能对对称条件独立性进行形式化编码，而这种独立性在实践中往往过于严格，难以成立。近期提出的不对称性标记有向无环图，既通过放宽独立性的对称性假设扩展了贝叶斯网络的类别，也能够表示所关注变量之间依赖关系的类型。本文介绍了适用于这一类模型的新型结构学习算法。这些算法不仅效率高，还能直观地解释其底层依赖结构。一项全面的计算研究展示了这些算法的效率。通过使用在意大利收集的COVID-19恐惧量表数据开展真实世界数据应用，本文展示了这些算法在实践中的用途。

关键词： 贝叶斯网络、条件独立性、概率图模型、分阶段树、结构学习。

1 引言

贝叶斯网络（BN）是一种概率图模型，通过有向无环图（DAG）简洁地表示离散变量之间的依赖结构[29, 41]。模型中蕴含的变量间任何条件独立性，都可以通过所谓的 D-分离准则从底层有向无环图中直接读出 [30]。

然而，在实际应用中，人们发现条件独立性的对称性假设往往过于严格，因此需要能够以图形方式描绘非对称独立性的模型。此后，人们定义了各种非对称条件独立性的概念，包括情境特定 [4], 部分 [32]和局部 [7], ，并对其性质开展了正式研究 [11, 12, 36, 43]。尽管已有用于嵌入和表示非对称条件独立性的贝叶斯网络扩展 [16, 22, 31, 33, 42],，但这些扩展往往会失去有向无环图所具有的直观性，而且目前没有可供实际使用的软件。

最近，人们提出了不对称性标记有向无环图（ALDAG），作为有向无环图的扩展，其边会根据任意一对随机变量之间依赖关系的类型进行标记 [45]。构建 ALDAG 时，首先从一种称为分阶段树 [8, 38], 的基于树的概率图模型出发，再应用转换算法构造与之对应的 ALDAG。有向无环图的标准工具（例如前文提到的 D-分离判据）同样适用于 ALDAG。不过，ALDAG 还包含额外信息，既包括相应的边标记，也包括用于构造它们的底层分阶段树。

提出了一种用于稀疏 ALDAG（即边数较少的 ALDAG）的结构学习算法。该算法包含三个步骤：（i）从数据中学习一个父节点数有上限的有向无环图；（ii）将学得的有向无环图细化为一个嵌入非对称条件独立关系的分阶段树

；（iii）将该分阶段树转换为其对应的 ALDAG 表示。然而，这一程序存在各种缺陷：

- 变量的排序取自步骤 (i) 中有向无环图的某个任意拓扑序，但这对于 ALDAG 来说未必最优，因为它只考虑对称形式的独立性；

- 变量的父节点也取自步骤 (i) 中的有向无环图，而这种选择同样可能因忽视非对称依赖关系而受到很大影响。

在此，我们提出了新的 ALDAG 结构学习算法，通过将从数据中学习有向无环图的步骤 (i) 替换为一个同时确定变量排序及最终 ALDAG 中每个变量父节点的过程，克服上述困难。为此，我们利用条件互信息这一概念； [6], 已采用该概念来量化变量间依赖关系的强度。

我们在此提出的算法能够学习连接数量较少的稀疏 ALDAG。稀疏性有两大主要优势。首先，学习算法只需在有限数量的候选 ALDAG 中搜索最优模型，因此效率更高、计算速度也更快。其次，稀疏性使非对称依赖关系能够以直观且易于解读的方式呈现；对于一般的 ALDAG，这种呈现方式并不可行。这与《人工智能》和《机器学习》领域近期的趋势一致，即关注如何解释模型的输出 [19, 26, 27]。

本文结构如下。第 2 节回顾有向无环图、ALDAG 以及对称和非对称条件独立性的各种概念。第 3 节首先回顾构建 ALDAG 所需的分阶段树，随后介绍用于 ALDAG 的新型结构学习程序，这些程序兼具高效性和可解释性。第 4 节介绍一项模拟研究，以考察我们算法的质量。第 5 节呈现

对 ALDAG 在研究影响新冠病毒恐惧的各因素之间关系方面的应用进行全面分析。本文最后以讨论作结。

2 非对称依赖与 ALDAG

2.1 有向无环图与条件独立性

令 $G = ([p], F)$ 为一个有向无环图（DAG），其顶点集为 $[p] = \{1, \ldots, p\}$，边集为 F。令 $\boldsymbol{X} = (X_i)^{i \in [p]}$ 为分类随机变量，其联合质量函数为 P，样本空间为 $\mathbb{X} = \times_{i \in [p]} \mathbb{X}_i$。对于 $A \subset [p]$，我们令 $\boldsymbol{X}_A = (X_i)_{i \in A}$ 和 $\boldsymbol{x}_A = (x_i)_{i \in A}$，其中 $\boldsymbol{x}_A \in \mathbb{X}_A = \times_{i \in A} \mathbb{X}_i$。如果对于 $\boldsymbol{x} \in \mathbb{X}$，满足

$$P(\boldsymbol{x}) = \prod_{k \in [p]} P(x_k | \boldsymbol{x}_{\Pi_k}),$$

其中 Π_k 是 G 中 k 的父节点集，$P(x_k|\boldsymbol{x}_{\Pi_k})$ 是 $P(X_k = x_k|\boldsymbol{X}_{\Pi_k} = \boldsymbol{x}_{\Pi_k})$ 的简写。有序马尔可夫条件蕴含如下形式的条件独立性

$$X_i \perp\!\!\!\perp \boldsymbol{X}_{[i-1]} \,|\, \boldsymbol{X}_{\Pi_i}, \tag{1}$$

等价于

$$P(x_i|\boldsymbol{x}_{[i-1]\setminus\Pi_i}, \boldsymbol{x}_{\Pi_i}) = P(x_i|\boldsymbol{x}_{\Pi_i}), \quad \text{for all } \boldsymbol{x} \in \mathbb{X}. \tag{2}$$

图1显示了由四个随机变量 X_1，……，X_4构成的有向无环图，表示对称条件独立性 $X_3 \perp\!\!\!\perp X_2|X_1$ 和 $X_4 \perp\!\!\!\perp X_1|X_2$，$X_3$。相应的马尔可夫分解为 $P(x_4|x_3,\ x_2)P(x_3|x_2)P(x_2|x_1)P(x_1)$。

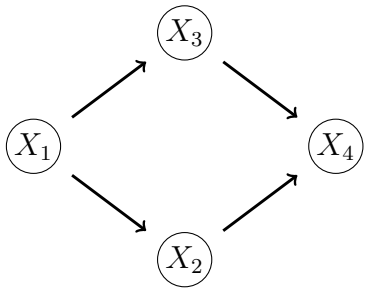

图1：一个包含四个随机变量的有向无环图示例。

2.2 非对称条件独立性

贝叶斯网络只能表示式（1）和式（2）所示形式的对称条件独立性。最常见的条件独立性非对称扩展是所谓的情境特定独立性，通常通过为贝叶斯网络的每个顶点关联一棵树来表示 [4]。设 A、B 和 C 是 $[p]$的三个互不相交的子集。若满足以下条件，我们称 $\boldsymbol{X}_A$ 在上下文 $\boldsymbol{x}_C \in \mathbb{X}_C$ 下与 $\boldsymbol{X}_B$ 上下文特定独立：

$$P(\boldsymbol{x}_A|\boldsymbol{x}_B, \boldsymbol{x}_C) = P(\boldsymbol{x}_A|\boldsymbol{x}_C) \tag{3}$$

对所有 $(\boldsymbol{x}_A,\ \boldsymbol{x}_B) \in \mathbb{X}_{A\cup B}$ 均成立，并记为 $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B|\boldsymbol{x}_C$。若该条件对所有 $\boldsymbol{x}_C \in \mathbb{X}_C$ 均成立，则式 (3) 中的条件可简化为式 (1) 中的标准条件独立性。

文献 [32] 引入了一种更一般的非对称条件独立定义，称为部分条件独立。如果 $\boldsymbol{X}_A$ 在取值域 $\mathcal{D}_B \subseteq \mathbb{X}_B$ 中给定上下文 $\boldsymbol{X}_C = \boldsymbol{x}_C$ 时部分条件独立于 $\boldsymbol{X}_B$，则称

$$P(\boldsymbol{x}_A|\boldsymbol{x}_B, \boldsymbol{x}_C) = P(\boldsymbol{x}_A|\tilde{\boldsymbol{x}}_B, \boldsymbol{x}_C) \tag{4}$$

对所有 $(\boldsymbol{x}_A, \boldsymbol{x}_B), (\boldsymbol{x}_A, \tilde{\boldsymbol{x}}_B) \in \mathbb{X}_A \times \mathcal{D}_B$ 均成立，并记作 $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B|\mathcal{D}_B, \boldsymbol{x}_C$。显然，当 $\mathcal{D}_B = \mathbb{X}_B$ 时，(3) 和 (4) 等价。此外，要使非平凡的部分条件独立性成立，样本空间 $\mathbb{X}_B$ 必须包含两个以上的元素。

最后一个条件是所谓的局部条件独立性，最早由 [7] 讨论。对于 $i \in [p]$ 和满足 $A \cap \{i\} = \emptyset$ 的 $A \subset [p]$，局部条件独立性表示如下形式的概率等式

$$P(x_i|\boldsymbol{x}_A) = P(x_i|\tilde{\boldsymbol{x}}_A) \tag{5}$$

对于所有 $x_i \in \mathbb{X}_i$ 和两个 $\boldsymbol{x}_A$，$\tilde{\boldsymbol{x}}_A \in \mathbb{X}_A$。请注意，就一般性而言，$3 \preceq 4 \preceq 5$。条件5仅说明某些条件概率分布是相同的，且无法检测到类似方程（3）和（4）中的明显模式。

2.3 依赖关系的类别

假设现有一个DAG模型，但我们认为变量之间还存在额外结构，且该结构不一定是对称的。这种结构可能包含与非对称独立性相关的等式，例如方程3–5中的等式。以下定义形式化了有向无环图 G中由一条边相连的两个随机变量之间可能存在的依赖关系类型。

定义1。设 P 为$\mathbf{X}$ 的联合质量函数，且 P 对于有向无环图 $G =([p], F)$满足马尔可夫性质。对于每个$(j, i) \in F$，我们称 X_i对 X_j的依赖关系属于以下类别

- 上下文，如果在某个上下文 $\mathbf{x}_C$ 下，X_i 和 X_j 是上下文特定独立的，其中 $C = \Pi_i \setminus \{j\}$。

- 部分，如果 X_i 在取值域 $\mathcal{D}_j \subset \mathbb{X}_j$ 中，给定上下文 $\mathbf{x}_C$ 时与 X_j 条件独立；并且 X_i 和 X_j 在相同上下文 $\mathbf{x}_C$下并非上下文特定独立。

- 局部，如果以上条件均不成立，且形式为 $P(x_i|\mathbf{x}^{\Pi_i}) = P(x_i|\tilde{\mathbf{x}}^{\Pi_i})$ 的局部独立性成立，其中 $x_j \neq \tilde{x}_j$。

- 完全，如果以上条件均不成立。

在标准DAG模型中，所有边都带有完全标签，但也可以考虑使用其他标签来表示更灵活的依赖类型。请注意，如果 X_i 与 X_j 之间依赖关系的类别是情境特定或部分依赖，那么也可能存在涉及这两个变量的局部独立性陈述，如方程5所示。同样，关于两个不同的上下文， X_i 与X_j 之间的依赖关系可以同时属于情境特定和部分依赖。另一方面，如果它们的依赖关系类别是局部依赖，那么根据定义，就不存在情境特定或部分相等关系。缺少边表示条件独立性，正如方程1所形式化的那样，也与标准DAG中的情形相同。

2.4 ALDAG

根据定义1中引入的依赖类别，我们现在可以为DAG补充有关非对称依赖结构的额外信息。由此得到的带标签图称为ALDAG，其中 [45], 的边根据标签着以不同颜色，因而也体现了变量之间关系的类型。

形式上，令 G 为一个DAG， F 为其边集，并且

$$\mathcal{L}^A = \{\text{‘context’}, \text{‘partial’}, \text{‘context/partial’}, \text{‘local’}, \text{‘total’}\}$$

表示依赖关系类型的边标签集合。

定义2。 *ALDAG* 是一个二元组 (G, ψ) ，其中 $G = ([p], F)$ 是一个有向无环图，且 ψ 是从 G 的边集到 $\mathcal{L}^A$的函数，即 $\psi: F \to \mathcal{L}^A$。如果联合质量函数 P 相容于 *ALDAG* (G, ψ) ，则称之为相容；这意味着 P 相对于 G 满足马尔可夫性质，并且 P 还遵循 ψ 给出的全部边标签；也就是说，对于每个 $(j, i) \in F$， X_i 与 X_j之间存在 $\psi(i, j)$ 依赖关系。

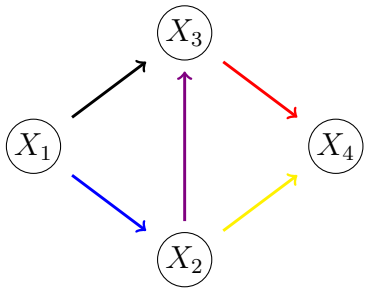

图2：一个包含四个随机变量的 ALDAG 示例。边的着色如下：红色表示上下文；蓝色表示部分；紫罗兰色表示上下文/部分；绿色表示局部；黑色表示完全。

下文中，我们用对 ALDAG 的边进行着色来表示其标签。标准贝叶斯网络具有一种 ALDAG 表示，其中所有边的标签都是“完全”。请注意，贝叶斯网络的标准特性在 ALDAG 上同样成立，例如前文提到的 D-分离准则以及快速概率传播算法。此外，近年来，人们还利用边标签 [23]提供的额外信息，将它们用于因果发现。

图2展示了一个包含四个随机变量的 ALDAG 示例。缺少从 X_1 指向 X_4 的边，对应于对称条件独立性 $X_4 \perp\!\!\!\perp X_1 | X_2,\ X_3$。带有 total 标签的黑色边表示 X_1 和X_3 之间的关系是对称的。其他所有由边连接、且没有 total 标签的变量对，则表明模型的条件概率中存在一些等式，这些等式与非对称独立性有关。此外，对于 X_2 和 X_4 之间带有 local 标签的边，这种非对称依赖非常无结构，既无法形式化为情境特定，也无法形式化为部分。

ALDAG 与 [31] 的带标签有向无环图有一些共同特征，但两者在两个关键方面有所不同：首先，带标签有向无环图只能嵌入上下文特定独立性，而 ALDAG 则能表示

任何类型的非对称独立性；其次，带标签有向无环图会明确指出独立性成立的上下文，而 ALDAG 不会。原因有二：一方面，ALDAG 中的具体独立性可从相关的分阶段树中读取（见下文）；另一方面，对于变量数量较多的应用，所需的上下文往往过于复杂，无法在有向无环图中呈现。

3 从数据中学习ALDAG

给定一组分类变量向量的观测值，我们希望使用结构学习算法来学习能够更好地描述数据依赖结构的ALDAG。此类算法将使用分阶段树图模型 [8, 38]，下文将对此进行介绍。

3.1 分阶段树

与贝叶斯网络不同，贝叶斯网络的图形表示是有向无环图，而分阶段树使用带颜色的树来直观呈现条件独立性。设 (V, E) 是一棵有向有限根树，其顶点集为 V，根节点为 v_0，边集为 E。对于每个 $v \in V$，令 $E(v) = \{(v, w) \in E\}$ 为从 v 发出的边的集合，并令 $\mathcal{C}$ 为标签集。

与 $\mathbf{X}$ 兼容的分阶段树是一个三元组 $T = (V, E, \theta)$，其中 (V, E) 是一棵有根有向树，并且：

1. $V = v_0 \cup \bigcup_{i \in [p]} \mathbb{X}_{[i]}$；

2. 对所有 $v, w \in V$，$(v, w) \in E$ 当且仅当 $w = \boldsymbol{x}_{[i]} \in \mathbb{X}_{[i]}$ 且 $v = \boldsymbol{x}_{[i-1]}$，或者 $v = v_0$ 且 $w = x_1$ 对某个 $x_1 \in \mathbb{X}_1$成立；

3. θ： $E \to \mathcal{L} = \mathcal{C} \times \cup_{i\in[p]}\mathbb{X}_i$ 是边的一种标记，使得对于某个函数 κ： $V \to \mathcal{C}$， $\theta(v,\boldsymbol{x}_{[i]}) = (\kappa(v)\ x_i)$。函数 k 称为分阶段树 T的着色。如果 $\theta(E(v)) = \theta(E(w))$，则称 v 和 w 属于同一阶段。由 $\theta(E(v))$诱导的等价类构成该树内部顶点的一个阶段划分。

以上第1点和第2点构造出一棵有根树，其中每条根到叶路径（或等价地，每个叶节点）都与样本空间 $\mathbb{X}$中的一个元素相关联。随后定义此类树的边标记，其中标签是一个有序对：一个元素来自集合 $\mathcal{C}$，另一个元素来自树中相应变量 X_i 的样本空间 $\mathbb{X}_i$。根据构造，**X**-兼容的分阶段树满足：两个顶点处于同一阶段，当且仅当它们对应于相同的样本空间。

图3给出了一个 (X_1, X_2, X_3, X_4)-兼容分阶段树模型的示例，该模型包含两个三元变量（X_1 和 X_2，水平为低/中/高）和两个二元变量（X_3和 X_4，水平为低/高）。函数 κ 给出的着色显示在顶点上，每条边（、（x_1、.....、 x_i））均标有 x_i。结合文本标签和发出边的顶点颜色，可以从图中读出边标记 θ。例如，$\theta(v_1, v_6) \neq \theta(v_2, v_7) = \theta(v_2, v_8) \neq \theta(v_2, v_9)$。这种将标记 θ 表示在顶点上的方式与表示在边上的方式等价，但更易于解释，因此下文将采用这种方式。共有31个内部顶点，分阶段方式为 $\{v_0\}$, $\{v_1, v_2\}$, $\{v_3\}$, $\{v_4, v_5, v_6\}$, $\{v_7, v_8\}$,$\{v_9\}$, $\{v_{10}\}$, $\{v_{11}\}$, $\{v_{12}\}$, $\{v_{13}, v_{14}, v_{19}, v_{20}, v_{25}, v_{26}\}$, $\{v_{15}, v_{18}, v_{21}, v_{24}, v_{27}, v_{30}\}$, $\{v_{16}, v_{22}, v_{28}\}$和 $\{v_{17}, v_{23}, v_{29}\}$。

条件独立性通过标记 θ 在分阶段树中进行形式化建模和表示。作为示例，请考虑图3中的分阶段树。v_4， v_5 和 v_6 处于同一阶段这一事实，与上下文特定独立性 $X_3 \perp\!\!\!\perp X_2 | X_1 = \text{low}$ 相关。绿色阶段 $\{v_7, v_8\}$ 表示X_2 和 X_3 之间的部分独立性，其取值域为 $X_2 \in \{\text{low,medium}\}$，上下文为 $X_1 =$medium。这一事实表明,

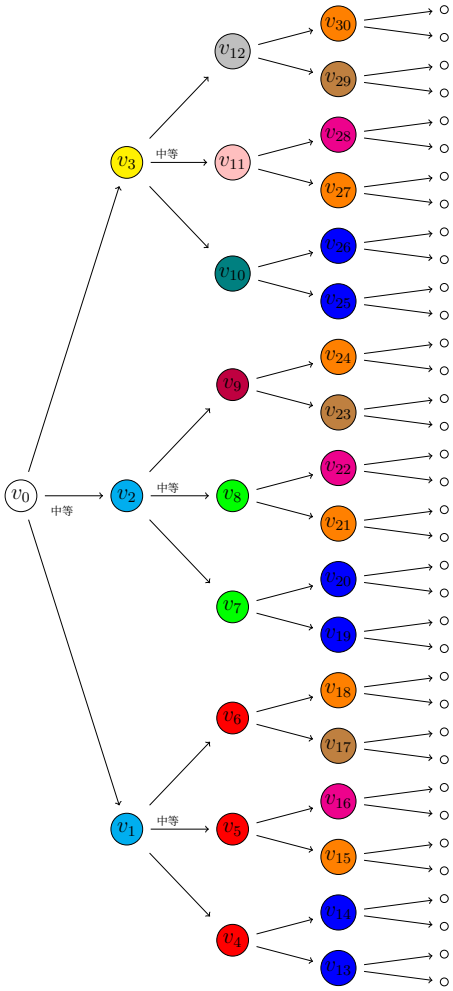

图3：由两个三元变量和两个二元变量构成的分阶段树，其 ALDAG 见图2。

例如， v_{27} 和 v_{30} 处于同一阶段，表示一种一般局部独立性，即给定 $X_1 =$ high $X_2 =$ high $X_3 =$ high时， X_4 的条件分布与给定 $X_1 =$ high $X_2 =$ medium $X_3 =$ low时 X_4 的条件分布相同。最后， $v_{13} - v_{18}$、 $v_{19} - v_{24}$ 和 $v_{25} - v_{30}$ 中的重复分阶段模式表示对称条件独立性$X_4 \perp\!\!\!\perp X_1 | X_2, X_3$。

分阶段树通过对内部顶点进行分阶段来表示多种非对称独立性。然而，随着可能的原子事件数量增加，通过目视检查来评估潜在的独立性会变得更具挑战性。因此， [45] 提出了ALDAG，将分阶段树中存储的信息压缩为一种更直观易读的表示，并提出了一种将分阶段树转换为其ALDAG表示的算法。该转换算法需要两个步骤：(i) 根据分阶段树构建有向无环图 G_T T ，使得若 $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_C$ 在 T 中，则 $\boldsymbol{X}_A$ 和 $\boldsymbol{X}_B$ 在 G_T中被 $\boldsymbol{X}_C$ d- 分离；(ii) 根据相关变量之间存在的依赖关系类型，为 G_T 中的每条边添加标签。例如，图3中分阶段树对应的ALDAG就是图2中的图。边(X_1, X_4)的缺失源于 $v_{13} - v_{18}$、 $v_{19} - v_{24}$ 和 $v_{25} - v_{30}$中的重复分阶段模式。这些带标签的边概括了分阶段树极其复杂的分阶段结构，而这种结构通常无法用标准贝叶斯网络简单表示。

3.2 按固定顺序学习ALDAG

鉴于 [45], 中介绍的从分阶段树到 ALDAG 的转换算法，从数据中学习 ALDAG 归结为学习相应的分阶段树。目前已有大量用于学习不同类型分阶段树的算法 [3, 9, 15, 25, 24, 37], ，这些算法也已在可免费获取的 `stagedtrees`R 软件包 [5] 中实现。由于分阶段树的模型搜索空间极其庞大，远大于有向无环图的模型搜索空间，因此所有这些算法

都基于贪心启发式搜索，并在每次迭代中优化模型评分（通常是贝叶斯信息准则，详见 [18]）。

研究者尝试通过仅考虑特定子类别来简化分阶段树的学习任务。在 [6] 中，定义了朴素分阶段树，其参数数量与基于相同变量的朴素贝叶斯网络相同。在 [25] 中，研究了简单分阶段树，其顶点划分类型受到限制。在 [12] 中，定义了上下文特定树，它们仅嵌入对称和情境特定类型的独立性。在 [45] 中，研究了学习其 ALDAG 不含局部边的树的算法。最后，在 [24]中，定义了 k 父节点阶段树，其 ALDAG 的父节点数有限。由于这些树将用于我们提出的稀疏 ALDAG 新算法，下面回顾其形式化定义。

定义 3. 若分阶段树 T 的最大入度小于或等于 k，则称其属于 k父节点阶段树类别。

使贝叶斯网络的结构学习具备可扩展性的早期方法之一，是限制每个变量可拥有的父节点数量 [17, 44]。这样做不仅是为了限制可能的有向无环图的模型空间，从应用角度看也合乎情理，因为通常预计只有有限数量的变量会直接影响另一个变量。标准 `bnlearn`软件 [35]也提供设置父节点最大数量的选项。

3.3 学习固定顺序下的稀疏 ALDAG

因此，当 k 设为较小的整数时，与具有 a k-个父节点的分阶段树相关联的 ALDAG 是稀疏的。请注意，使用非局部先验学习的 [9] 分阶段树，其 ALDAG 会是稀疏的；非局部先验在学习稀疏有向无环图 [2]方面也已证明有效。

在此，我们探讨比 [24], 中所讨论的算法更灵活的替代方案；那些算法要求利用与从数据中学习到的贝叶斯网络相兼容的特定顺序。我们还先假设，事先选定感兴趣变量的一个可能排序 X （下文假设该排序为正整数的顺序）。为确定哪些变量可以作为所学习 ALDAG 中某个顶点的父节点，我们使用条件互信息。回顾一下，给定（向量）$\boldsymbol{X}_C$、I（X_A； $X_B|\boldsymbol{X}_C$）时，两个变量 X_A 和 X_B 之间的条件互信息定义为

$$I(X_A; X_B|\boldsymbol{X}_C) = \sum_{\boldsymbol{x}_c \in \mathbb{X}_C} P(\boldsymbol{x}_C) \sum_{x_A \in \mathbb{X}_A} \sum_{x_B \in \mathbb{X}_C} P(x_A, x_B|\boldsymbol{x}_C) \ln \left(\frac{P(x_A, x_B|\boldsymbol{x}_C)}{P(x_A|\boldsymbol{x}_C)P(x_B|\boldsymbol{x}_C)} \right) \quad (6)$$

具体而言，通过迭代选择变量 $X_{[i-1]}$ 中与 X_i 的条件互信息最大的变量X_j ，来确定变量 X_i 的父节点。显然，如果 $i \leq k+1$，那么此前的所有变量都是X_i的父节点。请注意，方程 6 中的概率分布需要根据数据进行经验估计：为此，我们使用 `infotheo` R 软件包 [28]。

前述程序构建了一个有向无环图，但不会估计任何相关概率。随后，ALDAG 的结构学习算法将此有向无环图作为输入，并执行以下步骤：(i) 将有向无环图转换为等价的分阶段树表示（嵌入所有有向无环图条件独立性）；(ii) 运行 [5], 的后向爬山算法，该算法在每次迭代中只能合并底层分阶段树中的阶段；(iii) 将所得分阶段树转换为 G_T 并添加边标签。根据构造，所得 ALDAG 中任一顶点最多有 k 个父节点。

3.4 学习无固定顺序的稀疏 ALDAG

上一节所述的方法会输出一个与 $\boldsymbol{X}_\pi$ 兼容的分阶段树，并给出变量的一种可能排序 π 。对于变量数量较少的情况，可以直接

枚举全部 $p!$ 种可能的排序，并根据选定的标准（例如贝叶斯信息准则）选出最优排序。下面我们将探讨这种方法的可行性。

3.5 使用部分顺序学习稀疏 ALDAG

为避免考虑所有可能顺序所带来的阶乘级复杂度增长，一种折中方案是只考虑与从数据中学习得到的贝叶斯网络相容的顺序。对于每个允许的顺序，使用第3.3节中的算法从数据中学习一个 ALDAG，再根据选定的准则（例如贝叶斯信息准则）选出得分最高的一个。我们在此采用三种不同的方法来限制所考虑的顺序数量：

1. 使用禁忌算法从数据中学习贝叶斯网络，并选出与相应有向无环图相容的所有顺序；

2. 使用禁忌算法从数据中学习贝叶斯网络，并构造表示其等价类的混合图；随后，我们仅考虑该混合图中有向边所构成的有向无环图，并据此选出所有相容顺序；

3. 使用 PC 稳定算法 [10] 学习一个混合图，该图同样表示一个等价类；随后，我们仅考虑该混合图中有向边所构成的有向无环图，并据此选出所有相容顺序。

最后两种方法考虑了可能的排序，其中仅固定变量 [30] 之间的因果关系。根据构造，选项 2 考虑的可能排序数量不会少于选项 1。

请注意，上述算法需要从数据中学习贝叶斯网络或贝叶斯网络等价类，但这只是为了限制需要考虑的可能顺序。

所学习的 ALDAG 的父节点不依赖于初始贝叶斯网络的父节点，而是像第3.3节中那样，依据条件互信息来选取。

4 实验

4.1 计算研究

我们首先考虑概率图模型文献中的九个数据集，以评估稀疏 ALDAG 和不同估计程序的性能。必要时，使用等频方法对变量进行离散化。对于每个数据集，我们考虑六种不同的估计程序；对于每种程序，$k = 1$，2，3 种父节点数。首先，使用 bnlearn R 软件包中实现的禁忌算法学习贝叶斯网络模型 [35] （标记为 DAG）。随后，利用学习得到的贝叶斯网络的拓扑序，采用 [24] 中的方法学习 ALDAG（标记为 LV）。接着使用本文提出的方法：采用固定顺序（标记为 CMI），以及使用第3.5节中的三种兼容顺序方法（分别标记为 ORD1、ORD2 和 ORD3）。最后，在计算可行的情况下，即变量数少于七个时，我们会针对每一种可能的变量排序实施穷举模型搜索（标记为 ALL）。

学习得到的模型的贝叶斯信息准则值见表1。我们可以从表中得出以下结论：

- 对于每个数据集和每种算法，贝叶斯信息准则值都会随着父节点数 k的增加而降低。

- 对于 $k = 1$这种情况，DAG 方法优于部分 ALDAG 算法；但对于 $k = 2$，3 个 ALDAG 得出的贝叶斯信息准则值更低。

表1：在九个数据集上学习得到的模型的贝叶斯信息准则：贝叶斯网络（有向无环图）、采用 [24] （潜变量）方法的 ALDAG、采用固定顺序（条件互信息）的 ALDAG、使用相容顺序（ORD1、ORD2 和 ORD3）的 ALDAG、考虑全部顺序（全部）的 ALDAG。

data	k	DAG	LV	CMI	ORD1	ORD2	ORD3	ALL
asia	1	**22694.01**	**22694.01**	22919.69	22919.69	22919.69	22908.34	
	2	22214.59	22190.48	**22184.01**	**22184.01**	**22184.01**	**22184.01**	
	3	22214.59	22190.48	22181.24	**22178.24**	**22178.24**	**22178.24**	
恶病质	1	963.98	958.85	973.35	958.85	**957.14**	973.12	
	2	963.98	958.85	925.16	920.22	**904.47**	914.28	
	3	963.98	958.85	877.76	871.25	**851.55**	873.27	
chds	1	2831.56	2825.67	2825.67	2825.67	**2825.61**	2849.18	**2825.61**
	2	2831.56	2825.67	2826.09	2825.67	**2820.87**	**2820.87**	**2820.87**
	3	2831.56	2825.67	2824.87	2823.81	**2817.55**	**2817.55**	**2817.55**
冠状动脉	1	**13507.86**	**13507.86**	13508.53	13556.92	13555.14	13560.47	**13507.86**
	2	13449.09	13425.35	13424.10	13490.45	13487.30	13500.24	**13419.67**
	3	13442.02	13403.86	13385.54	13414.38	13411.23	13475.69	**13370.25**
fall	1	138118.14	**138060.31**	142495.58	**138060.31**	**138060.31**	**138060.31**	**138060.31**
	2	137637.86	**137418.54**	137561.62	**137418.54**	**137418.54**	**137418.54**	**137418.54**
	3	137637.86	**137418.54**	137503.24	**137418.54**	**137418.54**	**137418.54**	**137418.54**
ksl	1	**11927.73**	**11927.73**	12197.26	12115.97	12115.97	12186.57	
	2	11801.14	11773.84	11960.43	**11768.09**	**11768.09**	11948.90	
	3	11801.14	11773.84	11940.19	**11750.37**	**11750.37**	11916.28	
mathmarks	1	956.03	942.13	942.13	942.13	**936.78**	952.71	**936.78**
	2	956.03	942.13	930.64	930.64	**913.23**	918.01	**913.23**
	3	956.03	942.13	890.83	890.83	**876.21**	878.30	**876.21**
phd	1	8372.67	8371.22	**8369.02**	8397.87	8397.87	8402.31	**8369.02**
	2	8371.68	8356.90	8359.82	**8345.74**	**8345.74**	8348.98	**8345.74**
	3	8371.68	8356.90	8339.95	8338.37	8338.37	8338.44	**8330.42**
泰坦尼克号	1	10651.36	**10641.16**	**10641.16**	**10641.16**	**10641.16**	**10641.16**	**10641.16**
	2	10502.28	10451.72	10499.21	10451.72	**10450.72**	**10450.72**	**10450.72**
	3	10502.28	10451.72	10443.48	10432.84	**10431.85**	**10431.85**	**10431.85**

- 在所有情况下，至少有一种算法的得分不低于使用禁忌搜索的标准有向无环图方法。

- 除冠状动脉数据和博士数据（使用 $k = 3$）外，都存在一种得分与考虑所有可能顺序的算法同样优异的 ALDAG 算法。

- 从禁忌搜索有向无环图算法的等价类开始学习的算法，总体上是在不考虑每种可能顺序的算法中得分最高的；特别是，它优于从 PC 算法输出结果开始的算法。

表2报告了所有已学习模型的计算时间。正如预期，随着 k（即父节点数）的增加，计算时间也会增加，因为所学习的模型更加复杂。有向无环图算法当然是最快的，因为它考虑的模型类最小。采用固定顺序（LV 和 CMI）的 ALDAG 算法速度相近，而不采用固定顺序的算法最慢，但对于所有数据集而言仍然可行。总之，ORD1算法（该算法考虑通过禁忌搜索过程学习的贝叶斯网络的所有相容顺序）似乎在拟合效果与计算速度之间实现了合理的折中。

最后，我们在表3中报告具有完全或其他类型标签的边数。默认情况下，所有贝叶斯网络模型的边标签均为完全。每个顶点只有一个父节点的 ALDAG（$k = 1$）通常也只有完全边标签。拟合效果更好的模型（对应于在没有固定顺序的情况下学习得到的、含有 $k = 2$，3 个顶点的 ALDAG）具有更多非完全边标签，这凸显了学习能够嵌入非对称独立性的更灵活模型的必要性。

4.2 模拟研究

接下来，我们进行模拟研究，以评估所提出方法的可行性，并证明在以下假设下，该方法优于标准贝叶斯网络算法

表2：学习表1中所报告模型所需的时间（秒）。计算使用第11代英特尔®酷睿™ i7-1165G7 2.80GHz完成。

data	k	有向无环图	潜变量	条件互信息	ORD1	ORD2	ORD3	ALL
asia	1	0.02	0.07	0.16	34.72	169.97	1929.00	
	2	0.02	0.02	0.25	64.54	320.61	4530.72	
	3	0.02	0.03	0.36	102.28	526.64	7739.50	
恶病质	1	0.00	0.01	0.03	0.44	151.39	19.50	
	2	0.00	0.03	0.08	0.95	390.86	50.76	
	3	0.00	0.02	1.00	7.03	3237.02	376.25	
chds	1	0.01	0.00	0.00	0.02	0.14	0.06	0.12
	2	0.00	0.00	0.02	0.03	0.24	0.07	0.22
	3	0.00	0.01	0.02	0.06	0.50	0.18	0.47
冠状动脉	1	0.00	0.02	0.01	0.03	0.50	0.58	14.58
	2	0.02	0.02	0.05	0.08	1.19	1.39	34.48
	3	0.00	0.01	0.07	0.08	1.72	2.00	52.14
fall	1	0.16	0.22	0.14	0.32	4.53	4.39	4.32
	2	0.14	0.20	0.17	0.35	5.82	5.65	5.70
	3	0.14	0.22	0.06	0.58	6.03	5.94	6.10
ksl	1	0.00	0.03	0.03	10.33	10.11	226.57	
	2	0.01	0.02	0.06	11.33	10.88	442.58	
	3	0.00	0.02	0.12	12.80	12.15	751.77	
mathmarks	1	0.00	0.01	0.00	0.02	1.32	0.42	1.25
	2	0.00	0.00	0.03	0.05	4.42	1.50	4.31
	3	0.00	0.02	0.49	0.44	54.95	18.20	54.46
phd	1	0.00	0.02	0.02	0.23	0.92	3.53	10.58
	2	0.01	0.02	0.04	0.47	1.78	7.00	23.11
	3	0.00	0.01	0.11	1.30	4.59	17.28	68.21
泰坦尼克号	1	0.00	0.00	0.02	0.01	0.19	0.17	0.19
	2	0.00	0.02	0.02	0.03	0.45	0.41	0.43
	3	0.00	0.01	0.05	0.06	0.95	0.97	0.93

表1中的学习得到的模型中，完全标签（左侧数字）和非 total 标签（右侧数字）的边数。

data	k 个有向无环图		LV	CMI	ORD1	ORD2	ORD3	全部	
asia	1	(6, 0)	(6, 0)	(5, 0)	(5, 0)	(5, 0)	(6, 0)		
	2	(7, 0)	(3, 4)	(2, 6)	(2, 6)	(2, 6)	(1, 8)		
	3	(7, 0)	(3, 4)	(0, 12)	(0, 12)	(0, 12)	(0, 12)		
恶病质	1	(6, 0)	(4, 2)	(3, 3)	(4, 2)	(2, 4)	(4, 2)		
	2	(6, 0)	(4, 2)	(1, 10)	(1, 10)	(0, 11)	(0, 11)		
	3	(6, 0)	(4, 2)	(1, 14)	(1, 14)	(0, 15)	(0, 15)		
chds	1	(3, 0)	(2, 1)	(2, 1)	(2, 1)	(1, 2)	(1, 1)	(1, 2)	
	2	(3, 0)	(2, 1)	(1, 3)	(2, 1)	(1, 2)	(1, 2)	(1, 2)	
	3	(3, 0)	(2, 1)	(1, 5)	(2, 3)	(0, 4)	(0, 4)	(0, 4)	
冠状动脉	1	(5, 0)	(5, 0)	(5, 0)	(5, 0)	(5, 0)	(4, 0)	(5, 0)	
	2	(8, 0)	(3, 5)	(2, 7)	(2, 7)	(2, 7)	(1, 7)	(1, 8)	
	3	(8, 0)	(4, 4)	(2, 10)	(2, 9)	(2, 9)	(1, 10)	(2, 9)	
fall	1	(3, 0)	(2, 1)	(2, 1)	(2, 1)	(2, 1)	(2, 1)	(2, 1)	
	2	(5, 0)	(0, 5)	(1, 4)	(0, 5)	(0, 5)	(0, 5)	(0, 5)	
	3	(5, 0)	(0, 5)	(1, 5)	(0, 6)	(0, 6)	(0, 6)	(0, 6)	
ksl	1	(8, 0)	(8, 0)	(8, 0)	(7, 0)	(7, 0)	(7, 0)		
	2	(12, 0)	(3, 9)	(2, 12)	(2, 11)	(2, 11)	(1, 13)		
	3	(12, 0)	(3, 9)	(2, 15)	(0, 18)	(0, 18)	(0, 18)		
mathmarks	1	(4, 0)	(0, 4)	(0, 4)	(0, 4)	(0, 4)	(1, 3)	(0, 4)	
	2	(4, 0)	(0, 4)	(0, 7)	(0, 7)	(0, 7)	(0, 7)	(0, 7)	
	3	(4, 0)	(0, 4)	(0, 9)	(0, 9)	(0, 9)	(0, 9)	(0, 9)	
phd	1	(5, 0)	(4, 1)	(3, 2)	(3, 2)	(3, 2)	(4, 1)	(3, 2)	
	2	(6, 0)	(3, 3)	(3, 4)	(3, 6)	(3, 6)	(2, 6)	(3, 6)	
	3	(6, 0)	(3, 3)	(2, 9)	(2, 9)	(2, 9)	(1, 10)	(2, 7)	
泰坦尼克号	1	(3, 0)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	
	2	(5, 0)	(0, 5)	(0, 5)	(0, 5)	(0, 5)	(0, 5)	(0, 5)	
	3	(5, 0)	(0, 5)	(0, 6)	(0, 6)	(0, 6)	(0, 6)	(0, 6)	

真实模型是一个 ALDAG。我们从随机生成的、复杂程度各异的 ALDAG 中模拟数据：变量数量（$p \in \{4, 5, 6\}$）；每个变量的父节点数（$k \in \{1, 2, 3\}$）；相关分阶段树中每个变量的阶段数（$t \in \{2, 3, 4\}$）；样本量（$N \in \{250, 500, 1000, 3000, 5000, 10000\}$）。对于每组参数组合，我们重复进行 20 次实验，每次都随机打乱变量的顺序，以消除搜索启发式可能产生的任何偏差。估计程序与上一节中的相同。

我们首先考察这些算法能否找回用于模拟 ALDAG 的正确变量排序。如果假设变量之间的关系具有因果性，这一点通常很有意义 [30]。图 4 报告了每种变量数量（ p ）和父节点数（ p）组合下，肯德尔τ距离随样本量 N 的变化情况。肯德尔τ距离是在用于模拟 ALDAG 的真实排序与使用 `PerMallows` R 软件包 [21]实现得到的估计排序之间计算的。对于 $k = 1$ ，我们可以看到，各种方法在找回正确变量排序方面几乎没有差异。对于 $k = 3$，使用所有可能排序的算法表现优于其他算法，这与预期一致；但 ORD2 和 ORD3 的结果优于标准 DAG 方法。不出所料，CMI 算法采用任意的变量排序，其表现是所有方法中最差的。

图 5 报告了模型估计所需的平均时间随时间的变化情况。正如预期，有向无环图方法速度最快，但其耗时与固定顺序方法相近。遍历所有可能顺序的算法最慢：不过，对于六个变量的情况，其平均耗时仅为十秒。请注意，父节点数 k 似乎不会对计算时间产生重大影响。

每种变量数 (p)、父节点数

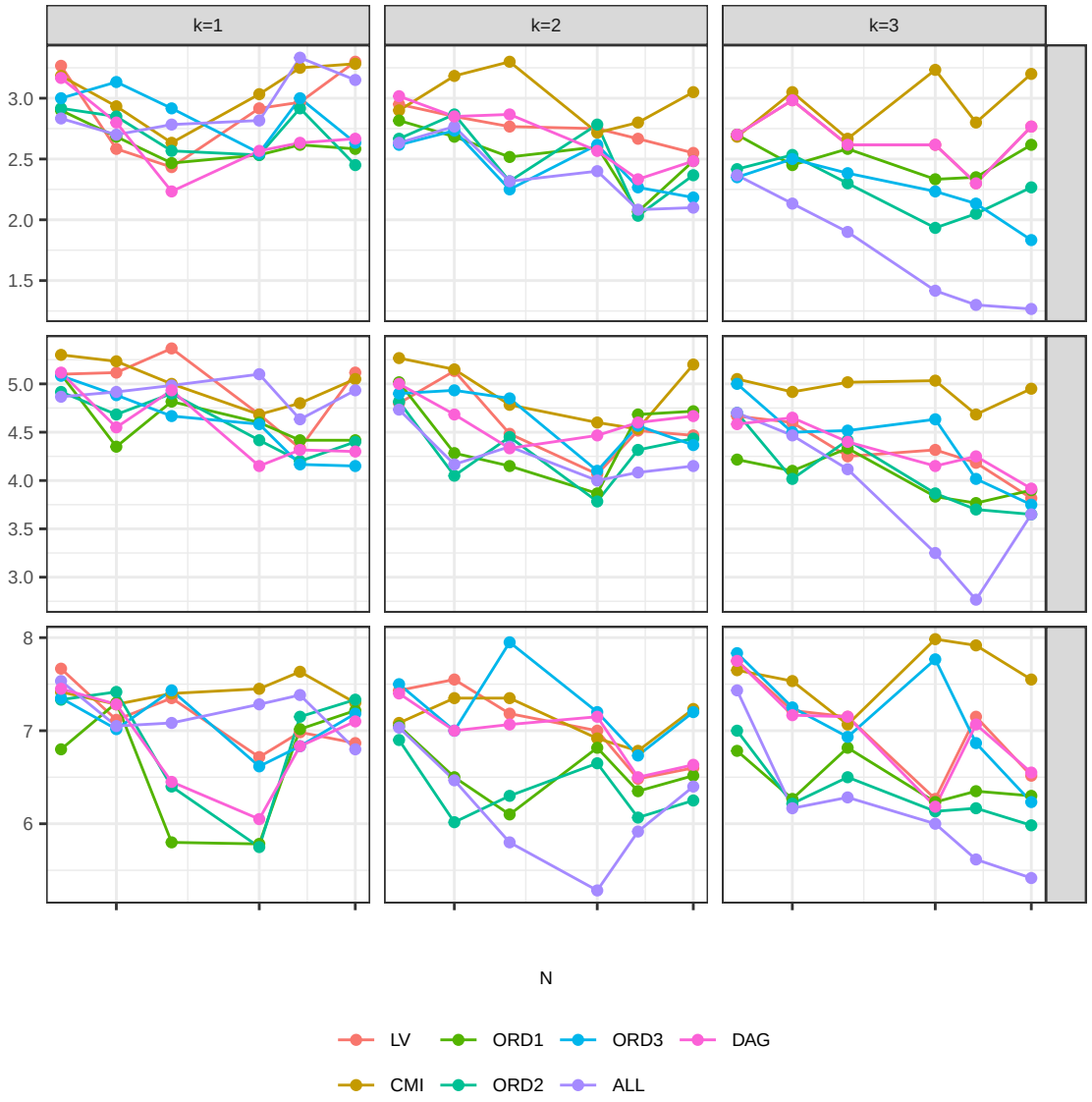

图 4：d 的估计模型与真实模型之间的肯德尔τ距离。

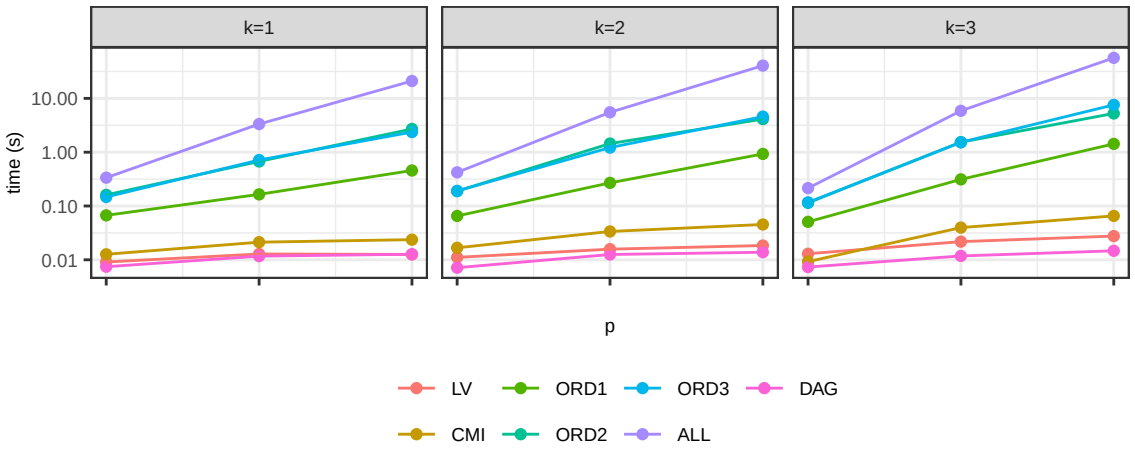

图 5：模拟实验中，计算时间（秒）随变量数量 p 的变化。

（k）和样本量（N）也针对每种估计程序进行了计算（篇幅所限，此处未报告）。结果表明，随着样本量增加，不采用固定顺序的 ALDAG 程序比标准有向无环图算法拟合效果更好。正如上文报告的计算研究所示，拟合效果的差异也会随 p 和 k增加而增大。

5 数据应用：COVID-19恐惧量表

新冠疫情震撼了世界，也因教育活动和非必要商业活动遭到数月封锁，改变了大多数人的生活习惯。对于全球几乎所有公民而言，如此规模且影响重大的事件，在其一生中前所未有。新冠疫情的诸多方面，例如对患者预后的不确定性、身边是否有熟人感染，以及生活习惯被迫改变，都让全球许多人产生了普遍的恐惧感 [20]。在以往的流行病中，恐惧

已被证明与焦虑和抑郁等其他心理障碍有关 [14], ，而这些障碍也被证明与社会隔离密切相关[34], 。在新冠疫情期间，政府的社交距离政策强制实施了社会隔离。

为了量化和测量个体对新冠疫情的恐惧，Fear of COVID-19量表于 [1],开发，共包含以下七个条目：

- 我最害怕COVID-19（恐惧）。
- 想到COVID-19会让我感到不适（思考）。
- 想到COVID-19时，我的手会变得湿冷（手）。
- 我害怕因COVID-19而失去生命（生活）。
- 在社交媒体上观看有关COVID-19的新闻和故事时，我会变得紧张或焦虑（媒体）。
- 我担心自己会感染 COVID-19，因此无法入睡（睡眠）。
- M每当我想到自己可能感染 COVID-19 时，心脏就会怦怦直跳或悸动（心 t).

参与者使用五级李克特量表表明自己对这些陈述的认同程度。选项包括“强烈不同意”“不同意”“既不同意也不反对”“同意”和“强烈同意”。每道题的最低分为 1 分，最高分为 5 分。将各条目的得分相加，计算总分（范围为 7 至 35 分）。分数越高，对 COVID-19 的恐惧程度越高。由于新冠疫情带来了普遍的恐惧感，该量表已在包括中国 [46], 意大利 [40] 和西班牙 [13]在内的许多其他国家使用并得到验证。

可从 [39]获取 COVID-19恐惧量表的意大利验证数据，其中包含 149 名受试者对量表七个条目的回答，以及他们的年龄和性别信息。年龄采用等频方法进行二分；量表条目的回答则转换为三元变量：不同意（包括强烈不同意和不同意）、中立（既不同意也不反对）和同意（包括同意和强烈同意）。

本分析旨在了解这两个人口统计因素对恐惧量表作答的影响，以及量表各条目之间的关系。为此，我们采用多种方法从数据中学习DAG和ALDAG模型：使用禁忌算法学习DAG；使用 [24]程序学习ALDAG；按照第3.3节所述的固定顺序学习ALDAG；以及考虑与初始贝叶斯网络相容的全部顺序来学习ALDAG。此处不考虑其他方法，因为对于总共九个变量，需要考虑约362880种顺序，这在计算上不可行。不过请注意，上一节的实验表明，使用与初始贝叶斯网络相容的全部顺序来学习ALDAG的过程，在拟合效果方面可与更精细的方法相竞争。对于所考虑的每种算法，均使用 $k = 1$，2，3个父节点进行评估。此外，对于所有算法，仅考虑年龄和性别均没有条目作为父节点的结果，因为我们关注的是人口统计变量对个体作答方式的影响。

表 4 报告了所有学习得到的模型的贝叶斯信息准则。可以注意到，所有有向无环图都满足任一顶点至多有一个父节点，因为所有贝叶斯信息准则值都相同。因此，使用 [24] 方法学习得到的所有 ALDAG 也都满足任一顶点至多有一个父节点。得分最高的模型具有三个父节点，其顺序是从与使用禁忌算法学习得到的初始贝叶斯网络相容的顺序中选取的。

图 6 展示了学习得到的贝叶斯网络。可以看出，确实每个变量最多只有

表 4：表1中学习得到的模型里，标记为完全（左侧数字）和非完全（右侧数字）的边数。

k	DAG	LV	CMI	ORD1
1	3275.158	3264.242	3337.566	3264.242
2	3275.158	3264.242	3251.970	3191.428
3	3275.158	3264.242	3225.318	**3166.348**

一个父节点。重要的是，这表明两个人口统计变量不会影响个体对量表条目的回答。

图 7 所示得分最高的 ALDAG 表明，人口统计变量与量表条目之间的严格独立性假设并不成立。年龄直接影响新闻条目，而性别直接影响除思考条目以外的所有条目。此外，该模型显示量表条目之间存在广泛得多的依赖关系，因为它包含了图 6 中有向无环图的所有边（尽管标签不同），以及其他边。有趣的是，没有边被标记为完全标签；有七条边标记为上下文，两条标记为部分，八条标记为上下文/部分。

对称条件独立性可以像在标准贝叶斯网络中一样从 ALDAG 中读出。例如，该网络表明，在给定性别、生活和思考的情况下，年龄和其余条目不会影响个体对恐惧条目的回答。换言之，恐惧条目概述了个体的恐惧程度，它只直接受性别以及生活和思考条目回答情况的影响。

边的标记提供了有关变量与其父节点之间关系的更多信息。在 [45], 中，引入了所谓的依赖子树，以便从 ALDAG 中提取变量与其父节点之间的实际依赖结构。

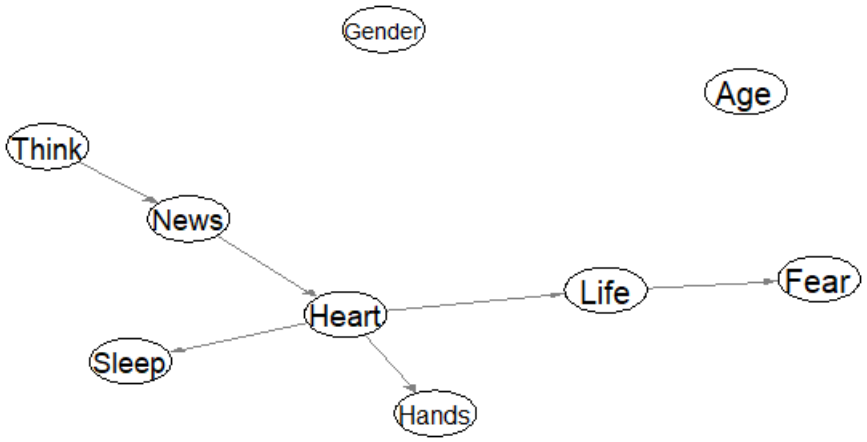

图 6：使用 COVID-19恐惧量表数据和禁忌算法学习得到的贝叶斯网络。

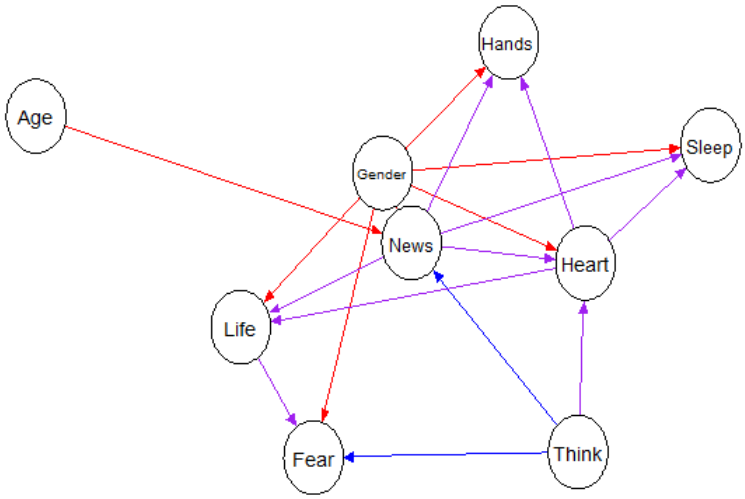

图 7：COVID-19恐惧量表数据的最佳评分 ALDAG。边着色如下：红色——上下文；蓝色——部分；紫罗兰色——上下文/部分；绿色——局部；黑色——完全。

变量“新闻”的结果如图 8 所示。该变量受到人口统计变量以及“思考”条目的直接影响。与分阶段树中的情况一样，与“新闻”变量（$v_9 - v_{20}$）相关的顶点着色表示非对称条件独立性。着色表明，对于在“思考”条目上回答“同意”或“不同意”的个体，性别和年龄不会影响他们对“新闻”的回答。对于在“思考”条目上回答 = “都不”的成年群体，性别也独立于“新闻”。

接下来，我们分析“生活”条目，因为它或许是七个条目中最关键的一个。根据有向无环图中的标准条件独立性，我们可以得出：给定性别、“新闻”和“心脏”后，“生活”独立于年龄、“思考”、“手”和“睡眠”。图 9 中的依赖子树揭示了“生活”与其父节点之间的依赖结构。与图 8 中的结构相比，我们可以看到一个稍欠规则的结构，其中包含各种非对称独立性。例如，对于在“新闻”上回答 = “都不”的男性，“壁炉”与“生活”相互独立。类似地，给定“壁炉”和“新闻” = “不同意”后，“生活”独立于性别。

如此复杂的依赖模式无法使用贝叶斯网络进行建模，但它们提供了有关数据依赖结构的宝贵信息。ALDAG 的优良性可由模型的贝叶斯信息准则加以证明，其值远低于贝叶斯网络模型的贝叶斯信息准则。

6 讨论

ALDAG 是标准有向无环图的一种富有表现力的扩展，携带了有关模型条件分布中额外等式的信息。在本文中，我们介绍了用于 ALDAG 的高效学习算法，这些算法能够返回具有可解释性的稀疏模型，并可通过小维度的分阶段树加以丰富。计算

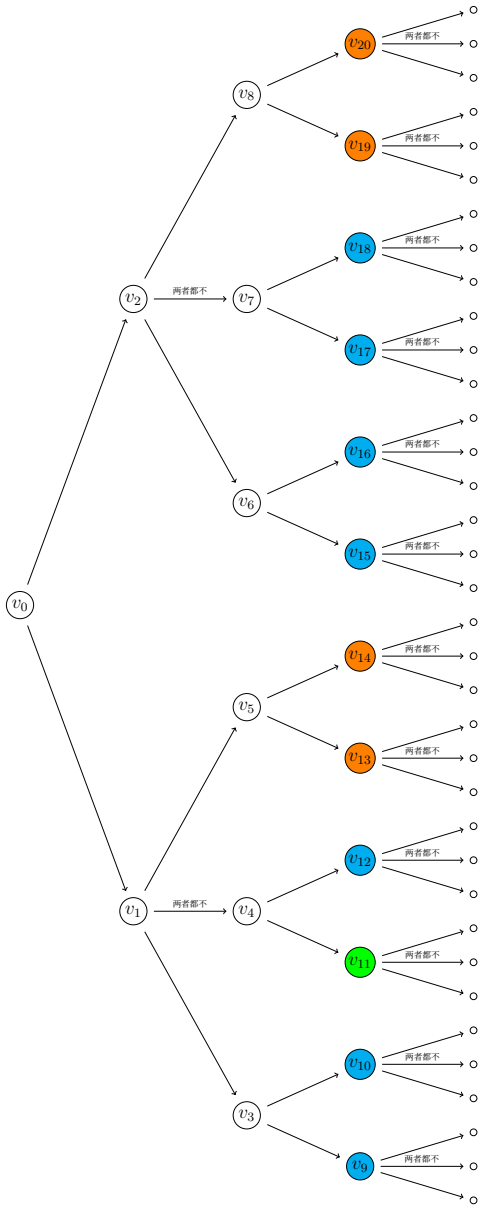

图 8：变量新闻的依赖子树，其父节点为（年龄、思考、性别）。

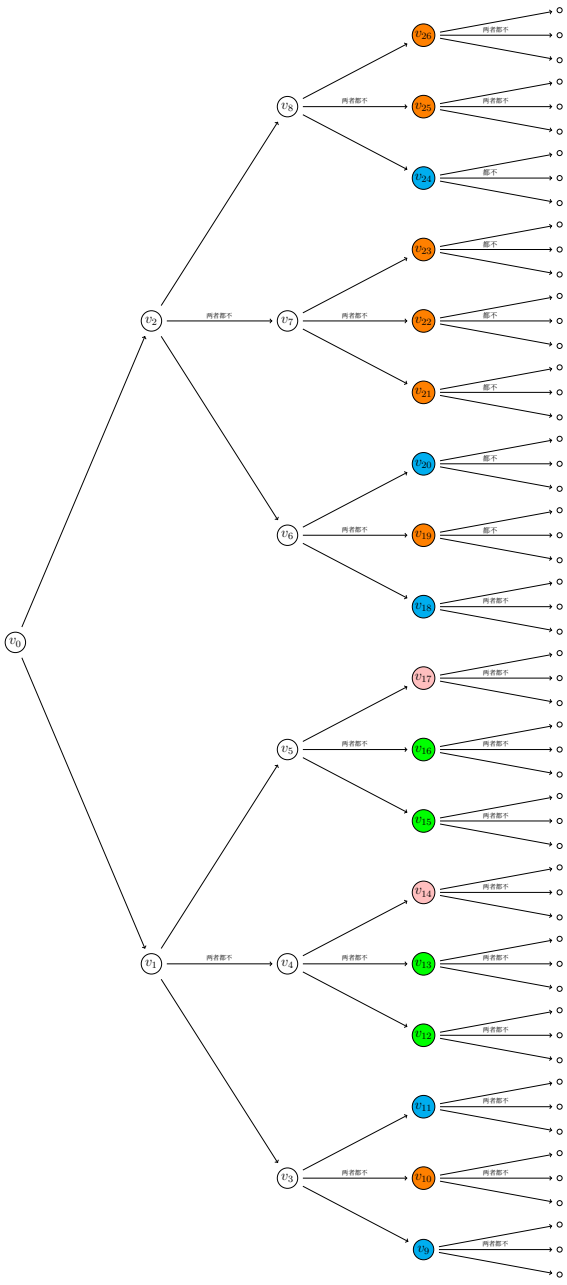

图 9：变量生活的依赖子树，其父节点为（性别、新闻、壁炉）。

研究和真实世界数据应用展示了 ALDAG 及其学习算法在理清复杂的非对称依赖结构方面的强大能力。

一个有前景的研究方向是扩展 ALDAG，使其能够传递有关变量间关系的因果信息。利用 ALDAG 进行因果发现，能够比使用有向无环图的情况提供更详细的因果效应信息。我们已开始在 [23], 中研究这一主题，但对于因果 ALDAG 的全面刻画仍然缺失。

在学习得到的模型中施加稀疏性的另一种途径是采用贝叶斯学习方法。这样一来，对所有可能的 ALDAG 空间设定先验分布，就可以惩罚更复杂的有向无环图结构，从而倾向于更简单的结构。我们目前正在研究此类方法的实现，以及不同先验分布的使用。

参考文献

[1] D.K. Ahorsu、C.Y. Lin、V. Imani、M. Saffari、M.D. Griffiths、A.H. Pakpour，《COVID-19恐惧量表：开发与初步验证》，《国际心理健康与成瘾杂志》，**20** （2022年），1537–1545。

[2] D. Altomare、G. Consonni、L. La Rocca，使用非局部先验对有序变量的高斯有向无环图模型进行客观贝叶斯搜索，《生物统计学》，**69** （2013），478–487。

[3] L. M. Barclay、J. L. Hutton、J. Q. Smith，使用链式事件图改进贝叶斯网络，《国际近似推理杂志》，**54** （2013），1300–1309。

[4] C. Boutilier、N. Friedman、M. Goldszmidt、D. Koller，《贝叶斯网络中的上下文特定独立性》。载于：第十二届人工智能不确定性会议论文集；1996年。第115–123页。

[5] F. Carli、M. Leonelli、E. Riccomagno、G. Varando，用于分层阶段树结构学习的 R 包 stagedtrees，《统计软件杂志》，**102**（2022年），1–30。

[6] F. Carli, M. Leonelli, G. Varando，基于阶段树模型的一类新型生成式分类器，*arXiv:2012.13798*

[7] D. M. Chickering、D. Heckerman、C. Meek，利用局部结构学习贝叶斯网络的贝叶斯方法。载于：第13届人工智能不确定性会议论文集；1997年，第80–89页。

[8] R.A. Collazo、C. G¨orgen、J.Q. Smith，链式事件图，1st 版，Chapmann & Hall，2018年。

[9] R.A. Collazo、J. Q. Smith，面向链事件图模型选择的一类新型非局部先验，《贝叶斯分析》，**11** (2016年)，1165–1201。

[10] D. Colombo、M.H. Maathuis，顺序无关的基于约束的因果结构学习，《机器学习研究杂志》，**15**，3741–3782。

[11] J. Corander, A. Hyttinen, J. Kontinen, J. Pensar, J. V¨a¨an¨anen，《上下文特定独立性的一种逻辑方法》，《纯粹与应用逻辑年刊》，**170** (2019), 975–992。[12] E. Duarte, L. Solus，《利用观测数据和干预数据表示上下文特定因果模型》，*arXiv:2101.09271*[13] B. Espejo, I. Checa，《西班牙版COVID-19恐惧量表（FCV-19S）：适配及构念效度与同时效度的验证证据》，《数学》，**9** (2021), 2512。

[14] B.N. Ford, R.H. Yolken, F.B. Dickerson, T.K. Teague, M.R. Irwin, M.P. Paulus, J. Savitz, 《重度抑郁障碍成人的麻疹免疫力下降》, 心理医学, **49** (2019), 243–249。[15] G. Freeman, J. Q. Smith, 《链事件图的贝叶斯最大后验模型选择》, 多元分析杂志, **102** (2011), 1152–1165。[16] N. Friedman, M. Goldszmidt, 《利用局部结构学习贝叶斯网络》。载于: 第十二届人工智能不确定性会议论文集; 1996年。第252–262页。[17] N. Friedman, I. Nachman, D. Pe'er, 《从海量数据集学习贝叶斯网络结构: "稀疏候选"算法》。载于: 第15届人工智能不确定性会议论文集; 1999年。第206-215页。[18] C. G¨orgen, M. Leonelli, O. Marigliano, 《分阶段树的曲指数族》, 电子统计学杂志, **16** (2022年), 2607–2620。[19] H. Hagras, 《迈向易于人类理解且可解释的人工智能》, 《计算机》, **51** (2018年), 28–36。[20] Y. Huang,N. Zhao, 《COVID-19疫情期间中国的广泛性焦虑障碍、抑郁症状与睡眠质量: 一项基于网络的横断面调查》, 精神病学研究, **288** (2020年), 112954。[21] E. Irurozki, B. Calvo, J.A. Lozano, 《PerMallows: 用于Mallows模型和广义Mallows模型的R软件包》, 《统计软件杂志》, **71** (2016年), 1–30。[22] M. Jaeger, J.D. Nielsen, T. Silander, 《学习概率决策图》, 国际近似推理杂志, **42** (2006), 84–100。

[23] M. Leonelli、G. Varando，使用分阶段树对分类数据进行上下文特定因果发现，*arXiv:2106.04416*[24] M. Leonelli、G. Varando，稀疏分阶段树的高效结构学习。载于：国际概率图模型会议；2022年。第 193–204 页。[25] M. Leonelli、G. Varando，简单分阶段树的结构学习，*arXiv:2203.04390*。[26] P. Linardatos、V. Papastefanopoulos、S. Kotsiantis，可解释人工智能：机器学习可解释性方法综述，*Entropy*，**23**（2020年），18。

[27] S.M. Lundberg、G. Erion、H. Chen、A. DeGrave、J.M. Prutkin、B. Nair、R. Katz、J. Himmelfarb、N. Bansal、S. Lee，从局部解释到借助面向树模型的可解释人工智能实现全局理解，*Nature Machine Intelligence*，**2**（2020年），56–67。

[28] P.E. Meyer，*infotheo:* 信息论度量，R 软件包版本 1.2.0.1，`https://CRAN.R-project.org/package=infotheo`。[29] J. Pearl，《智能系统中的概率推理：合理推断网络》，1st 版，Morgan Kaufmann，1988年。[30] J. Pearl，因果性，剑桥大学出版社，2009年。

[31] J. Pensar、H. Nyman、T. Koski、J. Corander，带标签有向无环图：有向图模型中上下文特定独立性的一种推广，*DataMining and Knowledge Discovery*，**29**（2015年），503–533。

[32] J. Pensar、H. Nyman、J. Lintusaari、J. Corander，《局部部分独立性在贝叶斯网络学习中的作用》，《国际近似推理杂志》，**69**（2016年），91–105。

[33] D. Poole、N.L. Zhang，《利用上下文独立性进行概率推理》，《人工智能研究杂志》，**18** （2003），263–313。

[34] Z.I. Santini, P.E. Jose, E.Y. Cornwell, A. Koyanagi, L. Nielsen, C. Hinrichsen, C. Meilstrup, K.R. Madsen, V.Koushede，社会脱节、感知孤立与美国老年人的抑郁和焦虑症状（NSHAP）：一项纵向中介分析，《柳叶刀·公共卫生》，**5** （2020年），e62-e70。

[35] M. Scutari，使用 bnlearn R 软件包学习贝叶斯网络，《统计软件杂志》，**35** （2010年），1–22。

[36] Y. Shen, A. Choi, A. Darwiche，学习上下文特定独立性的新视角。载于：第10届国际概率图模型会议；2020年。第425-436页。

[37] T. Silander, T.Y. Leong，用于学习链式事件图的动态规划算法。载于：国际发现科学会议；2013年。第201–216页。

[38] J.Q. Smith，P.E. Anderson，《条件独立性与链式事件图》，《人工智能》，**172** （2008），42–68。

[39] P. Soraci，《COVID-19恐惧量表》意大利语版本的验证与心理测量学评估。*UniData*—比科卡数据档案馆，米兰。研究编号*SN225*。数据文件版本*1.0* （2022年）。

[40] P. Soraci, A. Ferrari, F.A. Abbiati, E. Del Fante, R. De Pace, A. Urso, M.D. Griffiths，《COVID-19恐惧量表意大利语版本的验证与心理测量学评估》，《国际心理健康与成瘾杂志》，**20** (2022年)，1913—1922。

[41] L. E. Sucar，《概率图模型：原理与应用》，2nd 版，施普林格，伦敦，2021年。[42] T. Talvitie, R. Eggeling, M. Koivisto，利用局部结构、混合变量和精确算法学习贝叶斯网络，《国际近似推理杂志》，**115** (2019)，69—95。[43] S. Tikka, A. Hyttinen, J. Karvanen，通过上下文特定独立关系识别因果效应。载于：神经信息处理系统进展；2019年。第32页。[44] I. Tsamardinos, L. E. Brown, C. F. Aliferis，最大-最小爬山贝叶斯网络结构学习算法，《机器学习》，**65** (2006)，31–78。[45] G. Varando, F. Carli, M. Leonelli，分阶段树与不对称性标记有向无环图，*arXiv:2108.01994*。[46] W. Yang, P. Li, Y. Huang, X. Yang, W. Mu, W. Jing, X. Ma, X. Zhang，针对中国大学生的COVID-19恐惧量表跨文化适配与验证：一项横断面研究，《国际环境研究与公共卫生杂志》，**19** (2022年)，8624。