arXiv:2302.00293v3 [cs.LG] 1 Jan 2024

因果关系中的方法、挑战与展望综述

GAËL GENDRON，新西兰奥克兰大学MICHAEL WITBROCK，新西兰奥克兰大学GILLIAN DOBBIE，新西兰奥克兰大学

深度学习模型通过从高维数据中提取相关性模式，在各种任务中都取得了成功，但在初始数据分布之外进行泛化时仍然面临困难。由于因果引擎旨在学习不依赖于数据分布的机制，将深度学习与因果性结合起来，可能会对这两个领域产生重大影响。本文进一步阐释这一观点。我们从不同角度全面综述因果性的理论与方法，重点关注深度学习以及这两个领域面临的挑战。我们介绍了将这两个领域结合起来的早期尝试，并探讨了未来可能的发展方向。最后，我们列举了因果性技术的各种应用。

1 引言

研究因果性的一项动因，源自“相关性不意味着因果关系”这句话，也就是说，仅凭两个在统计上相关的变量 X 和 Y，并不足以推断其中一个导致了另一个。 共同原因原则 [138] 对这一说法作出限定，指出可能是 X 导致 Y、Y 导致 X，或者第三个变量 Z 是二者的共同原因。然而，这仍不足以完整刻画因果关系，因此人们为此发展出了不同的理论。许多科学领域都致力于寻找因果关系，例如社会科学 [36, 37], 经济学 [67], 神经科学 [47]。这也说明了人们对因果性这一主题的浓厚兴趣，以及应对这一问题的多种研究方法。

深度学习（DL）通过从高维数据中提取相关性模式，在各种任务中取得了巨大成功。然而，当前方法在不同于训练数据的数据分布上进行泛化时仍然面临困难。这可以用数据中存在混杂效应来解释，而混杂效应会导致伪相关。因果性领域研究此类混杂因素的影响，旨在寻找能够系统地学习不受数据分布影响的一般因果效应的方法。因此，可以假设，将因果关系理论引入深度学习能够提升其性能和泛化能力。然而，深度学习与因果性两个领域一直是分别发展的，直到最近，人们才逐渐对融合这两个领域产生了兴趣[158]。

这些观察结果促成了本篇关于因果关系理论与技术的综述。我们从机器学习的角度进行探讨，但也试图涵盖广泛的方法，这些方法应用于通常彼此没有交集的各种领域。我们介绍针对静态数据和时间序列的因果性观点，并呈现其应用所在的领域。我们还描述因果性与深度学习所面临的挑战，以及结合这两个领域的技术如何被应用于解决这些挑战。仍有许多开放性问题有待解决；我们将介绍这些问题，并概述当前研究文献中针对这些主题的观点，以探讨如何回答这些问题。我们在此总结本文的贡献：

- 我们对结构因果模型领域中的因果性进行了全面综述，并将通常被分别考察的潜在结果框架和格兰杰因果关系纳入其中。

作者地址：Gaël Gendron，奥克兰大学，新西兰奥克兰，ggen187@aucklanduni.ac.nz；MichaelWitbrock，奥克兰大学，新西兰奥克兰，m.witbrock@auckland.ac.nz；GillianDobbie，奥克兰大学，新西兰奥克兰，g.dobbie@auckland.ac.nz。

类别和子类别				论文
因果结构发现	平稳	图形	约束（4.1）	[18, 26, 27, 121, 137, 143, 167, 168, 168, 178, 183]
			评分（4.2）	[23, 28, 69, 114, 121, 136, 216]
			非高斯（4.3）	[164, 190, 214, 215]
			变分（4.4）	[4]
			大语言模型（4.5）	[81, 210]
	时间序列	图形	约束 (4.1)	[39, 148]
因果推断	平稳	do 演算	(5.1)	[77, 127]
		学习	重新加权 (5.2)	[80, 140, 144]
			匹配 (5.3)	[6, 79, 86, 95, 144, 170, 202]
			决策树 (5.4)	[24, 102]
			神经因果模型 (5.5)	[208, 209]
			变分（5.6）	[49, 88, 104, 152, 196, 209]
	时间序列	学习	Transformer（5.7）	[115]
			VAR（5.8）	[120]
因果表征学习	平稳	解耦	特征（6.1）	[1, 21, 40, 63, 72, 73, 84, 85, 92, 93, 99, 100, 110, 191]
		解	领域（6.1）	[12, 48, 68, 129, 213]
		增强	(6.2)	[8, 50, 109, 184, 212]

表 1. 所列论文的因果关系分类 rs.

- 我们阐述了因果性在深度学习中的应用价值，以及该领域未来可能的发展方向。
- 我们发现并描述了因果性与深度学习之间的若干联系，包括图神经网络、稀疏架构、数据增强和解耦。
- 我们详细列出了因果性的相关研究及其应用领域。

据我们所知，这些贡献首次汇集了因果性的各种方法，并整合了采用这些方法的应用。我们还建立了因果性与深度学习之间的新颖联系。相比之下，Zanga 等人 [207] 的综述聚焦于因果结构发现，而 Yao 等人 [202] 仅讨论因果推断方法。范围更广的综述 [53, 65, 120, 187] 涵盖了这两类任务，但未涉及最新的深度学习技术。此外，Moraffah 等人 [118] 讨论了时间序列方法，Kaddour 等人 [82] 则对因果性机器学习方法进行了另一项全面综述。

本文其余部分的结构如下：第 2节介绍主要因果关系理论的概念和术语。第 3节阐述并介绍神经因果模型，这是一种基于结构因果模型、利用深度学习技术进行因果推理的新颖框架。在第 4节和 5节中，我们从计算机科学的视角，对因果关系框架以及因果结构发现和因果推断方法进行全面综述。第 6 节介绍了一种随着深度学习兴起而出现的另一种因果任务——因果表征学习，并介绍该领域近期开发的技术。第7节介绍因果性在科学各领域中的应用。我们将在第8节总结本文。

2 因果关系理论

因果性是一门研究领域，旨在寻找系统性方法来揭示因果关系。它试图回答以下问题：推断因果联系需要哪些证据，以及我们能从中推断出什么？ [127]。因果性已在科学和哲学的多个领域得到研究 [193], , 但本文将仅讨论计算机科学领域的方法。

因果关系的恢复可分为两项任务：因果结构恢复和因果推断。因果结构恢复旨在从非结构化来源（例如表格数据）中提取并表示信息的因果结构。目前，这项任务完全由图模型（如结构因果模型（SCM））来处理 [127]。相比之下，因果推断旨在回答因果查询，即估计给定情境下因果变量的取值。因果推断方法要么尝试解决特定情形（例如回答单个查询），要么生成能够解决所有查询的函数（例如，估计结构因果模型的映射函数，下一节将对此进行说明）。因果推断这一术语常被用来涵盖这两项任务 [127]，但为了更清楚地说明本文所介绍的各类方法旨在解决哪些任务，我们将在全文中对二者加以区分。因果性领域的大多数文献都探讨因果推断。

本综述将重点讨论三种用于表示因果关系的范式：Pearl的结构因果模型 [127],*Rubin* 因果模型 [146], 和格兰杰因果关系 [60]。它们分别在第2.1、2.2和2.3节中讨论。

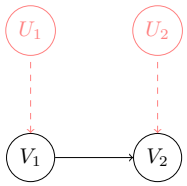
2.1 结构因果模型

结构因果模型 [127, 128, 197] 是一种用于表示因果关系的图模型。图 $\mathbf{G} = (\mathbf{V}, \mathbf{E})$由节点（或顶点）$\mathbf{V}$ 和连接这些节点的边$\mathbf{E} \subset \mathbf{V} \times \mathbf{V}$ 组成，其中 $N = |\mathbf{V}|$ 表示节点集的大小，$M = |\mathbf{E}|$ 表示边集的大小。边可以是有向边，也可以是无向边。在结构因果模型中，边为表示因果关系而设为有向边，并用箭头表示 $\rightarrow$。图中的**有向路径**是由有向边连接起来的一系列节点（沿箭头方向延伸）。图中的**环**是首尾节点相同的有向（非空）路径。只包含有向边且没有环的图称为有向无环图（DAG）。父节点通过边与子节点相连：$parent \rightarrow children$。节点的祖先节点 V 是位于通往 V 的有向路径上的节点，而 V 的后代节点则位于从 V 出发的有向路径上。节点 V 对应的父节点集、子节点集、祖先节点集和后代节点集分别记为**pa**、**ch**、**anc** 和**desc**。结构因果模型是基于有向无环图扩展而成的模型，其中包含一组函数和一个概率分布，节点表示因果变量，边表示变量之间的因果依赖。每个节点都关联一个函数，该函数以其父节点（原因）为输入，并计算出该节点的（结果）值。

2.1.1 形式化定义。结构因果模型（SCM）定义为一个元组 $\mathbf{M} = \mathbf{U}, \mathbf{V}, \mathbf{F}\ P(\mathbf{U})$，其中：

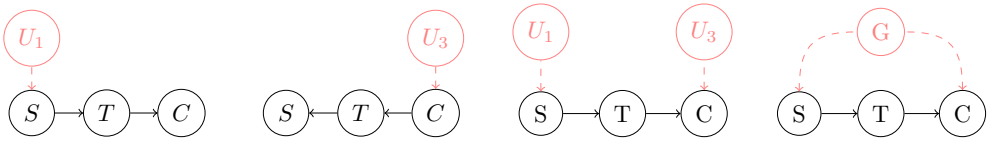
- **U**是未观测或外生 变量的集合
- **V** 是已观测或内生 变量的集合
- **F** 是一组映射函数 $\{f_{V_i} \in \mathbf{F} : V_i \leftarrow f_{V_i}(\mathbf{pa}_{V_i}, \mathbf{U}_{V_i})\}_{i \in 1 \ldots |V|}$，$f_{V_i}$用于计算内生变量的实例值 V_i，其条件是已知该变量的原因：父内生变量的实例值 $\mathbf{pa}_{V_i} \subset \mathbf{V} \setminus (V_i \cup \mathbf{desc}(V_i))$ 和外生变量 $\mathbf{U}_{V_i} \subset \mathbf{U}$
- $P(\mathbf{U})$ 是定义在外生变量 $U_i \in \mathbf{U}$上的概率函数

结构因果模型（SCM）的结构（即每个观测变量之间的依赖关系）可以用有向无环图表示。图1 给出了一个结构因果模型的示例。结构因果模型旨在对观测变量**V**背后的因果机制进行建模。未观测变量**U**的机制无法确定。这意味着，在结构因果模型中，**U** 中的变量没有父节点，并通过概率分布而非确定性函数建模，以表示不确定性。因果图的无环性假设有助于区分原因与结果，详见第4节。完整的结构因果模型可以充分表示一个因果系统。然而，为一组变量构建结构因果模型是一项艰巨的任务，需要识别并估计变量之间的因果联系（结构发现与推断）。第4节和5节将介绍生成结构因果模型的方法。



$$\mathbf{U} = \{U_1, U_2\}$$
$$\mathbf{V} = \{V_1, V_2\}$$
$$\mathcal{F} = \begin{cases} V_1 \leftarrow f_{V_1}(U_1) \\ V_2 \leftarrow f_{V_2}(V_1, U_2) \end{cases}$$
$$P(U_i) \sim \mathcal{N}(0,1)\ \forall U_i \in \mathbf{U}$$

图 1. 包含两个内生变量 V_1 和 V_2 以及两个外生变量 U_1 和 U_2的结构因果模型（SCM）示例。F 是将 V_1、 V_2 与 U_1、 U_2 链接在一起的映射函数集合。 $P(\mathbf{U})$ 是定义在 U_1 和 U_2 上的概率函数。



(a) 马尔可夫相对有向无环图。(b) 马尔可夫等价图，包含另一个原因。(c) 含外生变量的马尔可夫相对有向无环图。(d) 含外生变量的半马尔可夫有向无环图。

图 2. 马尔可夫等价和半马尔可夫（无环）有向无环图因果结构示例。（2a）一个马尔可夫相对有向无环图，用于描述吸烟（S）对癌症风险（C）的因果效应，其中中间变量表示肺部焦油（T）的存在。（2b）一个马尔可夫等价模型，可通过对相同观测数据作不同解释得到。（2c）另一个马尔可夫相对有向无环图，其中外生变量排除了先前的解释。（2d）同一模型，其中未观测变量被合并为一个潜在混杂因素：遗传因素 (G) 同时导致吸烟倾向和较高的癌症风险。最后一个图是半马尔可夫的。

2.1.2 隐藏混杂因素与马尔可夫等价性。恢复变量系统的因果结构颇具挑战，而且并非总是可行。首先，因果任务面临的主要挑战是处理混杂效应。所谓隐藏混杂因素，是指一个共同导致多个内生变量的外生变量。在马尔可夫图中，某个观测变量的取值不依赖于其父节点以外的其他观测变量（即 $V \perp\!\!\!\perp \mathbf{anc}_V \setminus \mathbf{pa}_V | \mathbf{pa}_V$）。包含隐藏混杂因素的图是非马尔可夫的（若图无环，也称为半马尔可夫），因为外生因子$\mathbf{U}$ 彼此并不独立，并且可能影响图中相距较远的多个变量$\mathbf{V}$，从而产生非马尔可夫依赖 [197]。其次，即使不存在隐藏混杂因素，真实的因果结构也可能只能恢复到某个马尔可夫等价类。马尔可夫等价的结构因果模型具有不同的因果结构，却会产生相同的观察数据。

图2 展示了这两个概念：以吸烟和癌症为例。吸烟（S）与肺部焦油（T）的存在相关，而焦油与较高的癌症风险（C）相关。图2a 和2b 展示了两个可从相同观察数据中得到的马尔可夫等价图。在这些例子中，各变量始终相关，因为它们都是同一外生过程的确定性结果。如果还涉及另一个随机性来源，如图2c 所示（由此产生不同的数据分布），因果方向可能可以恢复（例如，T 总是在 S 之后出现，而 C 只在某些情况下被观察到）。另一种可能的因果结构可能包含一个未观察到的共同原因（例如遗传学），如图2d （$C \leftarrow G \rightarrow S \rightarrow T$是另一种可能的图）。如果我们能找到一个与数据相符的因果结构，那么该图就是**可识别**的。如果有两个因果模型都与观察数据相符，则可以根据它们回答干预查询和反事实查询的能力加以区分，详见第2.1.3节。

表 2. Pearl 因果层级。每一层都对应一种因果查询，所需数据的数量逐层增加。每一层都关联着相应的直观活动和概率公式。这些公式展示了各层的最简单情形，并假设所需变量数为最少。

	层级	活动	公式
$\mathcal{L}_1$	关联	观察	$P(y\|x)$
$\mathcal{L}_2$	干预	行动	$P(y\|do(x))$
$\mathcal{L}_3$	反事实	想象	$P(y\|do(x)\ x',\ y')$

2.1.3 Pearl 因果层级。结构因果模型（SCM）的目的是提供一个用于回答因果查询的框架。这些查询可分为三种类型，并按照难度递增的顺序排列成一个层级。这个层级称为因果阶梯或*Pearl* 因果层级（PCH）[10]，并展示在表2中。

层级结构中的第一层 $\mathcal{L}_1$ 对应于 观察性 查询，即基于被动收集的数据提出的查询。其形式为：“**X** 能告诉我们关于 **Y**的什么信息？”，可用条件概率 $P(y|x)$建模，并可通过贝叶斯统计求解。基于监督和无监督方法的机器学习只能回答这类查询 [10]。

层级结构中的第二层 L_2 对应于干预查询，即关于环境变化／干预效果的查询。其形式为：“如果我们强制设定 X，会对 Y 产生什么影响？”。一种新的算子用于对此类查询进行数学建模，即 do算子 [126]。do(X = x) 表示将变量 X 强制赋值为 x（即使从未观察到这种情况）。$\mathrm{Y_x}$ 也常用于表示 Y|do(X = x)。使用第 5 节介绍的 do 演算规则，可以将此算子化简为条件概率。强化学习可以通过与其环境交互来回答这一层级的查询 [10]。

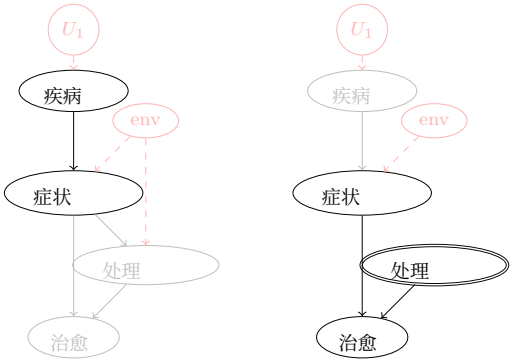
第三层也是最后一层$\mathcal{L}_3$ 表示反事实查询，即关于假想情境的问题。其形式如下：“如果发生的不是**X'**，而是**X** ，那么**Y**（导致了**Y'**）会发生什么？”在因果层级中，这类查询最难解决，因为它们要求对未曾发生的事情进行推理（“如果.....会怎样？”问题）。然而，对外生变量进行边缘化后，查询可以化简为以下表达式（对于离散情形，假设最简单的情况，即只有一个外生变量 U） [10, 127]:

$$P(y|do(x), x', y') = \sum_{u \in U} P(y|do(x), u)P(u|x', y') \tag{1}$$

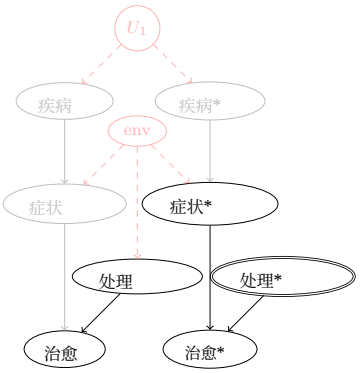
核心思想是利用外生变量作为共享证据，将事实世界与反事实世界联系起来。计算过程分为三个步骤：(1) 溯因计算 $P(u|x',\ y')$；(2) 行动执行干预 $do(x)$；(3) 预测使用运算数计算最终结果。

图3 通过一个示例说明了每种查询类型。一种疾病在某个人群中传播，其机制未包含在我们的因果模型中，并由一个外生变量表示 U_1。患者感染疾病后会出现症状。症状可能轻微，也可能严重，具体取决于未观测到的环境因素（例如，医疗服务可及性可能会减轻症状的严重程度）。患者随后可以接受处理，从而治愈。能否获得处理可能取决于与症状相关的未观测因素（例如，通过同一医疗保健系统可以获得处理）。这里的环境因素充当隐藏混杂因素。在这种情形下，问题在于评估处理对治愈概率的影响。在我们的模型中，答案取决于处理和症状的严重程度（也可能不取决于它们），但这两个变量并不独立，因为它们有一个共同的隐藏原因。

结构因果模型能够准确表示因果机制。借助完全指定的结构因果模型，可以计算 PCH 中的所有查询层。在结构因果模型框架下，因果推断旨在尽可能准确地估计



(a) L_1 查询：这些$_2$ 查询：这些症状能告诉我们患者是否会治愈治愈，如果他接受疾病的处理?



(c) L_3 查询：如果患者没有接受处理，他会治愈吗?

图 3. 计算 L_i 层查询所涉及的因果图。对于 L_2 查询，受干预的变量（用双边表示）被强制赋值并纳入图中，不再依赖其父变量。对于 L_3 查询，事实世界与反事实世界（带 * 的变量）由相同的外生因子联系起来。利用事实世界的观察数据估计 U_1 和*env* 变量（溯因），然后对反事实图进行干预（行动），并计算最终估计值（预测）。

真实的因果模型。然而，恢复真实的结构因果模型很难实现，**因果层级定理**(CHT) [10] 还指出，解决某一层$\mathcal{L}_i$ 查询所需的信息量通常不足以解决$\mathcal{L}_{i+1}$层的查询。具体而言，CHT 指出，对于能够解决$\mathcal{L}_1$ 层查询的因果模型，总能找到另一个模型，它对$\mathcal{L}_1$ 层查询给出相同的响应，却对$\mathcal{L}_2$层查询给出不同的响应。对于某些特殊的结构因果模型，$\mathcal{L}_2$ 层与$\mathcal{L}_3$ 层之间可能出现坍缩；但通常总能找到两个因果模型，它们对$\mathcal{L}_2$ 层查询给出相同的响应，却对$\mathcal{L}_3$层查询给出不同的响应。获取信息这一问题是因果性面临的主要挑战，因为一旦解决了这一问题，问题便可归结为概率推理 [119]。

2.2 鲁宾因果模型

Rubin 模型，又称 Neyman-Rubin因果模型或潜在结果框架（POF） [146,169, 202], , 是一种基于观测结果与反事实结果这一理念的因果模型。与 SCM 不同，该模型并不旨在恢复数据的因果结构，而是估计应用处理 变量后结果的分布。这种情况在医学应用中十分常见，研究人员希望估计某种处理对疾病的影响。实现这一目标的标准方法是进行随机对照试验，将处理随机分配给一组人，以减轻患者年龄、族裔和健康状况等因素造成的混杂效应。然而，这种方法成本高昂，并且出于伦理、可行性或实际条件等原因，在某些情况下可能无法实施。例如，将死亡率很高的药物注射给处理组是不道德的。再如，在衡量一项经济决策对某个国家的影响时，设置处理组和对照组并不可行。研究独特议题时甚至不可能做到这一点：在决定是否采取行动应对气候变化后，只能观察到两种结果中的一种。另一种结果称为反事实结果，永远无法观测到。最后这个例子说明了

因果推断的**基本问题**：我们希望估计一个未被观测到的情境所产生的结果。

2.2.1 形式化定义。 在 Rubin 模型中，会对一个感兴趣的对象施加处理 ，该对象称为个体。处理变量记为 $X \in \{0、1、.....、N_X\}$，其中 N_X 表示可能的处理数量，0 表示未接受处理。为与上一节的记号保持一致，我们将与潜在结果对应的随机变量记为 $Y(X = x)$、Y_x 或 $Y|do(x)$。 Y_0 表示对照组的结果， Y_1 表示接受处理 1 的组（若只研究一种处理，则为处理组）的结果。与患者属性对应的背景随机变量，即协变量，记为 Z。我们注意到，文献中也会采用其他记号；为与 SCM 使用的理论记号保持一致，我们选择了这一种。在潜在结果框架中，因果推断的目标是在不同粒度层面上确定处理效应：总体层面、亚组层面或个体层面。

平均处理效应（*ATE*）。ATE 在公式 2 中定义，指接受处理的个体与未接受处理的个体之间结果效应的期望平均差异。

$$\text{ATE} = \mathbb{E}[Y_1 - Y_0] \tag{2}$$

处理组平均处理效应（*ATT*）。ATT 在公式 3 中定义，指处理组的预期结果与同一处理组未接受处理时的反事实结果之间的差异。

$$\text{ATT} = \mathbb{E}[Y_1|X = 1] - \mathbb{E}[Y_0|X = 1] \tag{3}$$

条件平均处理效应（*CATE*）。CATE 在公式 4 中定义，指在给定背景变量 Z 条件下，某个子群体中接受处理与未接受处理的个体之间结果效应的差异。

$$\text{CATE} = \mathbb{E}[Y_1|Z = z] - \mathbb{E}[Y_0|Z = z] \tag{4}$$

个体处理效应（*ITE*）。ITE 在公式 5 中定义，对于特定个体或个体 i，指实施处理与未实施处理两种情形下结果效应的差异。在后一种情形中，似乎无法观察到其中一种情形，因为同一个体不可能同时接受和不接受处理。这个反事实结果（可能是 Y_1 或 Y_0）必须通过其他方法确定。

$$\text{ITE}^{(i)} = Y_1^{(i)} - Y_0^{(i)} \tag{5}$$

使用 Rubin 模型进行因果推断的目标是计算上述公式，最终目的是计算任意给定个体的 ITE。为实现这一目标，尽管缺乏观察数据，仍有必要计算反事实结果。为此，人们已经开发出相关技术，详见第5节。

2.3 格兰杰因果关系

读者或许已经注意到，迄今为止我们提到的因果性概念并不包含任何时间概念。这或许令人惊讶，因为时间是我们直观理解因果关系时的一个基本概念，也是大脑进行因果推断时坚实的归纳偏置 [16]。然而，正如我们在前几节所见，在大多数流行的因果框架中，时间并非核心，甚至并非

必需。现在，我们关注光谱的另一端：依赖时间敏感信息，并将因果学习应用于时间序列的因果模型。

时间序列是一组其取值随时间演变的变量 $\mathbf{X}(t) = (x_1(t)\ x_2(t) \ldots, x_N(t))$，其中 t 对应当前时间步，$\{X_i \in \mathbf{X}\}_{i \in 1 \ldots N}$ 是时间敏感变量的集合，而 $\mathbf{X}(t)$ 是 $\mathbf{X}$ 中每个变量在时间步 t的取值快照。格兰杰因果关系 [60] 是一种针对时间序列发展的因果性理论。在该模型中，如果利用所有可用信息（记为 $\mathbf{U}$）对 $\mathbf{Y}(t)$ 的预测优于剔除 $\mathbf{X}$ 过去取值后的信息 $\mathbf{U} \setminus \{\mathbf{X}(0), \mathbf{X}(1) \ldots, \mathbf{X}(t-1)\}$，那么序列 $\mathbf{X}$ 会导致（或“格兰杰导致”[118]）$\mathbf{Y}$。更形式化地说：如果$\sigma_t^2(\mathbf{Y}|\mathbf{U}) < \sigma_t^2(\mathbf{Y}|\mathbf{U} \setminus \{\mathbf{X}(0), \mathbf{X}(1) \ldots, \mathbf{X}(t-1)\})$，其中 $\sigma_t^2(\mathbf{A}|\mathbf{B})$ 是预测误差 $\epsilon_t(\mathbf{A}|\mathbf{B}) = \mathbf{A}(t) - P(\mathbf{A}(t)|\mathbf{B})$ 的方差。需要强调的是，格兰杰因果关系无法还原真正的原因 [112]。在上述定义中，如果 $\mathbf{X}$ 和 $\mathbf{Y}$ 相关，但二者由第三个隐藏序列 $\mathbf{Z}$ 导致（$\mathbf{X} \leftarrow \mathbf{Z} \rightarrow \mathbf{Y}$），格兰杰因果关系检验将无法区分这种情况与因果情形 $\mathbf{X} \rightarrow \mathbf{Y}$。

另一类从时间序列推断因果性的方法基于信息论，并使用熵度量来推断因果关系。变量的香农熵 [163] X可定义如下，如公式6所示。同样，我们可以定义两个变量之间的联合熵 X和 Y，如公式 7所示；还可以定义在给定 条件熵 X Z，如公式 8所示。

$$H(X) = -\sum_x P(x) \log P(x) \tag{6}$$

$$H(X, Y) = -\sum_x \sum_y P(x, y) \log P(x, y) \tag{7}$$

$$H(X|Z) = -\sum_x \sum_z P(x, z) \log P(x|z) \tag{8}$$

熵度量变量的不确定性大小。熵越低，随机变量的取值就越容易预测。根据这些定义，我们可以定义给定 Z 时，两个变量 X和 Y之间的**互信息**：

$$I(X; Y|Z) = H(X|Z) + H(Y[Z) - H(X, Y|Z) \tag{9}$$

两个变量之间的互信息表示二者共享的信息量。基于互信息的方法试图检测因果系统中信息随时间的流动 [123]。上述公式具有对称性，因此无法反映信息流动的方向。为解决这一问题，Schreiber 引入了**传递熵** 这一概念 [159] ，用于从数据中检测变量之间的不对称性。传递熵的定义见公式10 （为简洁起见，时间步以下标表示）。

$$T_{X \to Y} = H(Y_t|Y_{t-1}, Y_{t-2}, \ldots, Y_{t-K}) - H(Y_t|Y_{t-1}, \ldots, Y_{t-K}, X_{t-1}, \ldots, X_{t-K}) \tag{10}$$

传递熵是一种非对称度量，用于衡量信息从 X流向 的情况。它比较过程 Y在时间步 t的熵（已知其截至步骤 $t-K$的过去取值），以及Y 的熵（已知其过去取值和第二个变量 X的过去取值）。它可用于恢复时间序列之间的因果结构。格兰杰因果关系和传递熵是两个相互关联的概念，尤其是在高斯变量的情况下，这两种度量是等价的 [11]。传递熵适用于两个过程，但当多个变量都可能影响系统时则难以奏效 [118, 173]。后续研究试图通过提出新的时间序列因果性度量来解决这一问题 [173, 180, 185]。

3 神经因果模型

因果性提供了大量用于推理原因与结果的工具，但尽管理论上正确，这些传统方法在处理高维变量时仍面临困难。相比之下，深度学习已被证明非常擅长从大型数据集中提取模式并将其应用于各种任务。然而，这些方法在初始分布之外的泛化能力较差，因果推理能力不足，也很容易受到混杂效应的误导 [55, 57, 90, 158]。将这两个领域的优势结合起来似乎顺理成章，但有趣的是，直到最近，这两个领域的发展路径一直各不相同。人们已尝试构建能够进行因果推理的神经模型，但这些方法仍处于起步阶段，如何大规模开发因果推理引擎仍是一个尚未解决的问题。本节重点探讨因果推理引擎的潜在优势及其相关理论。第4节和第 5 节将介绍基于该框架的因果结构发现与推断方法，以及基于第2节所述框架的方法。第6节将从因果视角探讨表征学习问题，即如何从低层次数据中学习因果变量。

3.1 动机

如第2.1.3节所述，基于监督学习的传统机器学习方法仅依赖观察数据。CHT [10] 指出，此类模型通常无法回答因果查询（表2所示层级结构中的第二层和第三层查询）。将因果关系理论融入机器学习技术，尤其是深度学习（DL），有助于解决这一问题，并推动该领域取得进一步进展。我们总结了因果模型可能为深度学习带来的影响。

3.1.1 泛化。 首先，现代监督学习技术依赖于数据满足独立同分布（i.i.d.）这一假设。当该假设不成立时，模型在 [55, 57, 90, 158] 相应环境中的表现可能会很脆弱。i.i.d. 假设确保学到的模式能够应用于一般情形。然而，数据集可能存在偏差，或不足以代表真实世界分布。一个模型在某个图像数据集上表现良好，在另一个分布略有不同的数据集上却可能表现不佳（例如，在照片上训练的模型可能无法很好地处理素描）。模型不仅容易受到偏差影响，还可能被伪相关样本、对抗攻击和分布偏移所误导 [57, 157, 158]。为抵消这一影响，人们尝试了许多技术，其中最成功的做法是尽可能增大模型和数据的规模，由此形成了深度学习的**以数据为中心**的方法，该方法由基础模型 [14]主导。另一方面，因果模型可被视为**以模型为中心的**方法，因为它们不依赖于数据中不存在伪效应这一假设。因果模型旨在表示数据生成机制的因果结构，且这种表示不受数据分布方式的影响。此外，如第2节所述，它们能够识别伪相关和混杂效应并将其纳入考虑，而且无需大量数据。近期研究证明，在小数据场景下，因果模型较不容易过拟合，并能取得更好、更快且更稳健的结果 [155]。将因果引擎融入深度学习任务，有助于构建更稳健、较少受偏差影响且泛化能力更强的模型。

3.1.2 模块化。其次，深度学习引擎通常是由密集且纠缠的计算构成的黑箱。这使其难以解释，且计算成本很高。许多研究认为，模块化架构有助于解决这两个缺点，并学习可迁移机制，以便在不同任务之间重复使用 [14, 58, 134]。具体而言，**独立因果机制 (ICM)** 原则 [130, 157]提出，因果

机制是自主的，彼此之间互不交互。**稀疏机制转变**(SMS) 假设 [157] 进一步指出，分布的微小变化只会导致因果模型发生局部变化，而不会影响所有因子。密集计算并非必要，只会带来额外的伪相关。近期研究表明，模块化神经网络的梯度幅度要低得多，并且在具有稀疏因果结构的任务上表现更好[155]。在深度学习领域，人们已尝试构建此类稀疏且模块化的架构 [34, 43, 58, 134]。这些尝试大多依赖于专家混合模型，并使用路由机制来选择合适的专家。这些架构由多个专家层叠组成，并不具备图结构。它们的模块化程度仍然相当有限。另一方面，图形因果模型本质上基于图结构，通常是稀疏的，非常适合这一目的。神经科学理论也认为，大脑以模块化方式运作。每个模块都是某一领域的专家，并与其他专家进行稀疏通信，只传递少量信息 [7, 41]。由有向无环图构成、节点间稀疏交换信息的因果模型，将更接近大脑的实际行为。

3.1.3 干预与反事实。 第三，如本节开头所述，目前的 DL 模型还无法回答以下类型的查询："如果我们执行 X，会发生什么？如果当时发生的是 Y，又会怎样？"这些问题对应于 CHT 在第 $\mathcal{L}_2$ 和 $\mathcal{L}_3$, 节 2.1.3中规定的因果查询层级。在线 强化学习 （RL）技术能够与环境交互，并以此直接执行干预。因此，它们可以回答部分层级为 $\mathcal{L}_2$ 的查询，但因果推理可能有助于提升 RL 模型在想象空间中进行推理以及回答反事实查询的能力 [29, 157]。进一步而言，我们可以认为，反事实推理是迈向更类人化人工智能智能体的关键。实现这一目标的重要一步，被认为是生成 **世界模型**，即能够理解其周围环境内部机制，并在环境中进行推理和规划的模型 [66, 111, 156]。

3.2 神经因果模型

神经因果模型（NCM） [197] 是一个理论框架，建立了结构因果模型与神经网络之间的联系。与第2.1.1节中定义的结构因果模型类似，NCM 定义为三元组$\mathcal{M}(\theta) = \langle \mathbf{U}, \mathbf{V}, \mathcal{F} \; P(\mathbf{U}) \rangle$，其中：

- 与结构因果模型一样，U是外生变量的集合
- 与结构因果模型一样，V 是内生变量的集合
- 与结构因果模型一样，F 是映射函数的集合，用于根据变量的原因（其父内生变量 $\mathbf{pa}_{V_i} \subset \mathbf{V} \setminus (V_i \cup \mathbf{desc}(V_i))$ 的实例值以及外生变量 $\mathbf{U}_{V_i} \subset \mathbf{U}$）计算该变量的效应值。NCM 将函数类别限制为前馈神经网络：$\{f_i \in \mathbf{F} : v_i \leftarrow f_{V_i}(\theta_{V_i}; \mathbf{pa}_{V_i}, \mathbf{U}_{V_i})\}_{i \in 1...|V|}, f_{V_i}$
- $\boldsymbol{\theta} = \{\theta_{V_i} : V_i \in \mathbf{V}\}$ 是映射函数的参数集合
- 至于结构因果模型，$P(\mathbf{U})$ 是定义在 $U_i \in \mathbf{U}$ 中噪声变量上的概率函数；神经因果模型还进一步将概率律约束为 $\mathbf{u_i} \sim \text{Unif}(0,1) \;\; \forall U_i \in \mathbf{U}$

简而言之，神经因果模型是结构因果模型的一个子类，其中表示变量间因果关系的映射函数仅限于参数化神经网络这一类。初始论文 [197] 进一步将外生变量的分布限制为均匀分布，并将映射函数限制为多层感知机（MLP）。由于多层感知机是通用函数逼近器，且所有分布都可以根据均匀分布进行估计 [3], , 这一限制保证了神经因果模型类与更广泛的结构因果模型类具有相同的表达能力。神经因果模型能够表示 $\mathbf{L}_1, \mathbf{L}_2$ 层的查询，以及 $\mathbf{L}_3$ 层的查询（观察数据、干预和反事实）[197]。然而，在实践中，使用多层感知机进行因果推断是难以处理的 [208]。后续研究放宽了

这两项假设。这种选择可能会导致表达能力下降，但能得到可学习的模型。我们还强调，将神经模型与因果引擎相结合的研究不一定遵循 NCM 原则。

4 因果结构发现

因果结构发现的任务是学习一组变量之间的因果关系，包括两个变量之间是否存在因果联系，以及在可能的情况下确定因果关系的方向。结构发现方法可分为多个类别：基于约束的方法、基于评分的方法和非高斯方法是因果结构发现的传统方法。有些非高斯方法也利用神经模型。相比之下，变分方法和 Transformer 方法依赖于近期的深度学习技术。

4.1 基于约束的方法

这些基于约束的因果结构发现方法依靠独立性检验，在变量之间建立因果连接 [120] 并构建结构因果模型。它们基于归纳因果（IC）算法 [183]。在原始论文中，该算法可用于存在隐藏混杂因素的情形。然而，遵循 Pearl [127] 所作的区分，为了便于理解，我们先介绍无混杂情形下的 IC 版本，即假设模型中不存在隐藏混杂因素。

IC 算法分为三个步骤。第一步构建无向因果依赖关系。根据感兴趣的（观测到的）变量集合**V**构建无向图。如果无法找到集合 $S_{AB} \subset V \times V \setminus A, B$ 使得 $A \perp\!\!\!\perp B | S_{AB}$，就会在两个变量 A 和 B 之间添加一条边。换言之，A 和 B 共享一种不依赖于其他变量的依赖关系。此条件被认为足以说明 A 和 B 之间存在因果联系。第二步和第三步用于确定因果方向。第二步构建 *v-structures*，即形状为 $A \rightarrow C \leftarrow B$ 的结构（也称为碰撞点）。对于每一对不相邻的变量（A, B），如果它们有一个共同邻居 C（$A - C - B$），且满足 $C \notin S_{AB}$，则可以推断出 C 对 A 和 B 没有因果效应，也不能是其中任何一个的原因。因此，这三个变量只能构成 $A \rightarrow C \leftarrow B$ 结构。第三步为剩余可能的边定向。如果将一条边朝某个方向定向会形成一个环，则必须朝另一个方向定向。如果此过程会形成 *v-structure*，也要应用该规则，因为所有此类结构都应在第二步中发现。

IC 算法的步骤较为抽象，无法直接实现。许多算法以 IC 的思想为基础，为该方法提供了实用的实现方案。最早的一种应用是 SGS 算法，其最坏情况下的复杂度为指数级 [167]。最受欢迎的算法之一是 Peters 和 Clark（PC）算法 [168], ，其运行时间为多项式级。改进算法试图通过减少条件独立性检验的次数，并放宽对节点顺序的假设，来降低 PC 的复杂度并提高其稳定性，例如 PC-stable [26], 、conservative-PC [137], 、PC-select（原论文中称为 PC-simple）[18], ，以及 MMHC [178], ——一种由极小极大启发式引导的基于 PC 的方法。上述所有方法都假设模型中不存在隐藏混杂因素。这一假设有助于简化问题，但在实践中很少成立。

现在，我们介绍在存在隐藏混杂因素时恢复图结构的方法。如果所执行的检验具有统计显著性，那么无混杂情形下的算法可以将因果有向无环图结构恢复到马尔可夫等价类 [44, 113, 183]。但在混杂情形下，情况不再如此，因为可能不存在能够解释观测数据的有向无环图结构。针对隐藏混杂因素的 IC 算法（在 Pearl [127] 的文献中称为 IC*）

其第三步经过修改，以考虑混杂效应：当无法区分因果关系 $A - B$ 究竟属于 $A \to B$、$B \to A$ 还是 $A \leftarrow Z \to B$，且Z 存在隐藏混杂因素时，边保持无向。

快速因果推断（FCI）算法 [168] 提供了一种在存在隐藏混杂因素时实现 IC 的方法，而真实快速因果推断（RFCI）算法 [27] 缩短了运行时间。最近，迭代因果发现（ICD）方法 [143] 通过迭代运行，减少了所需的条件检验次数。该算法从完整图开始，逐步移除错误的边。独立性检验以局部邻域为条件，而该邻域的规模会在每一步增大；但随着每一步移除边，图变得越来越稀疏，因此即使邻域中包含大量变量，因果推断仍然很快。

最后，这些方法也已应用于时间序列。时间序列快速因果推断（TsFCI）算法 [39] 是将 FCI 算法应用于时间序列的一种改编方法。该方法使用滑动窗口将时间序列划分为多个数据框，每个数据框都有自己的一组变量，且变量不包含时间分量。同一时间序列中的各个时点在一个数据框中是彼此独立的变量，因此需要添加额外规则来纳入这些背景知识。最近，PCMCI [148] 将 PC 算法与瞬时条件独立性检验相结合。该框架分为两个阶段：第一阶段为每个变量移除父节点候选项，直至得到一个最小集合；第二阶段则根据这组数量较少的变量建立因果联系。

4.2 基于评分的方法

基于评分的方法会根据一组变量，为每个潜在的因果图计算似然评分。由于可能的图的搜索空间会随着节点数量的增加而呈指数增长，因此这类算法面临的主要挑战是剪枝搜索空间 [120]。变量 $\mathbf{V}$ 的联合概率分布，其因果结构由图 $\mathcal{G}$给出，可以通过贝叶斯网络 $\mathcal{B}$表示，如公式 11所示。本节所述方法试图恢复与该分布匹配的真实有向无环图 $\mathcal{G}$ 。

$$P_{\mathcal{B}}(V_1 = v_1, V_2 = v_2, \ldots, V_N = v_N) = \prod_{i=1}^{N} P(V_i = v_i | \mathbf{pa}_{V_i}^{\mathcal{G}}) \tag{11}$$

早期的一种搜索方法是 K2过程 [28]。该方法基于对父节点进行贪心搜索。如果向节点添加父节点能提高结构的概率，就会将其添加进来。不过，K2 的评分并不严格等同于条件独立性，因为它假设数据分布存在先验 [23]。广受欢迎的贪心等价搜索（GES）方法 [23, 114]分两个阶段，从空有向无环图开始，迭代构建因果图：第一阶段中，如果添加一条边能提高有向无环图的评分，就在每一步添加一条边。当评分不再提高时，便进入第二阶段。如果移除一条边能提高有向无环图的评分，就在每一步移除一条边。与基于约束的方法类似，该方法是精确的，并能返回最优有向无环图（在不存在混杂因素的假设下，精确到一个等价类）。贪心干预等价搜索（GIES）[69] 通过执行干预改进了这一方法，将等价解类限制在一个更小的子集中，而该子集在执行干预时仍然有效。快速贪心等价搜索（fGES） [136] 修改了 GES 算法，使其能够并行化，并扩展到变量数量较多的情况。这些方法能够高效地计算因果模型，但都假设不存在隐藏混杂因素。贪心快速因果推断（GCFI）通过结合 GES 与 FCI [121]消除了这一限制。与以往在离散域中进行的研究不同，基于迹指数与增广项的非组合优化

结构学习的拉格朗日方法（NOTEARS）模型 [216]将有向无环图学习视为一个连续优化问题。节点值用特征矩阵表示 $\mathbf{X}$，图的边则用连续的加权邻接矩阵表示 $\mathbf{W}$。该方法还假设因果变量之间的关系是线性的。基于这一假设，要最小化的损失函数可简单定义为 $\mathcal{L}(\mathbf{W}) = \|\mathbf{X} - \mathbf{X}\mathbf{W}\|_2$。正则化项可增强图的稀疏性，约束则确保图无环且非平凡。通过将权重 $\mathbf{W}$ 中对应权重非零的情况视为存在边，即可得到离散邻接矩阵。这种方法缓解了有向无环图组合搜索空间带来的问题。然而，Kaiser 和 Sipos [83] 指出，当因果变量的方差不同时，NOTEARS 会产生有偏图，并认为该方法不适用于因果发现。

4.3 非高斯方法

基于约束的方法和基于评分的方法是因果发现的基本途径。除了 NOTEARS 之外，所有方法都依赖统计检验。不过，近年来，研究者开始探索无需统计检验的方法，以解决这一问题，因为统计检验计算量大，而且需要大量数据才能达到统计显著性 [53]。所提出的方法受到独立成分分析（ICA）领域的启发 [78]，并基于数据服从非高斯分布这一假设。该假设能够提取数据中原因与结果分布的不对称性，并确定因果结构。此外，这些方法还假设模型中不存在混杂因素。

一项早期研究——线性非高斯无环模型（LiNGAM） [164] ——将变量之间的因果联系表示为如下形式的线性函数：

$$v_i = \sum_{v_j \in \mathbf{pa}_{v_i}} b_{ij} v_j + \epsilon_i + c_i \tag{12}$$

其中，模型中的任意变量都是其父节点的线性函数。ϵ_i 是扰动项，c_i 是一个附加常数。扰动项用于表示噪声（或结构因果模型中的外生变量），并且假设所有 ϵ_i 相互独立（即不存在混杂因素）。LiNGAM 的第二个也是最关键的假设是，ϵ_i 服从非高斯分布。这一限制源自独立成分分析（ICA）领域，在该领域中，这一假设用于分离相互独立变量的线性组合。如果混合结果服从高斯分布，就无法恢复构成它们的变量（这是中心极限定理的结果）。如果混合结果及其所有底层变量都服从非高斯分布，则可以利用它们的不对称性将其分离。回到我们的问题，上述方程可用矩阵形式表示为：

$$\mathbf{v} = \mathbf{B}\mathbf{v} + \epsilon \tag{13}$$

$$\mathbf{v} = (1 - \mathbf{B})^{-1}\epsilon \tag{14}$$

$$\mathbf{v} = \mathbf{A}\epsilon \tag{15}$$

因此，待解决的问题可归结为计算矩阵$\mathbf{A}$的值。在 LiNGAM 中，这通过 ICA 实现：所有混合项 $v \in \mathbf{v}$ 都被分解为来自$\mathbf{v} \setminus v$的相互独立的（原因）变量。LiNGAM 的直接扩展是后非线性模型（PNL） [214, 215]，它将线性映射替换为非线性函数的复合：$v_i = f_2(f_1(v_j) + \epsilon_i)$。

当维度较低且提供了足够的数据以进行统计检验时，传统的因果结构发现和因果推断方法效果良好。然而，

尽管因果结构发现已有所改进，但大规模恢复因果结构仍然困难 [157]。为缓解这一问题，我们研究了用于结构发现的神经网络方法。尽管名称如此，神经因果推理网络（NCINet）[190] 是一种基于神经网络、适用于双变量场景的因果结构发现框架。由此得到的图可以表示因果关系（$X \rightarrow Y$）、反因果关系（$X \leftarrow Y$）或无关联（$X \; Y$）。其核心思想是，因果方向上的回归均方误差应低于反因果方向上的回归均方误差；也就是说，如果 X 导致 Y，那么 $y = f_1(x)$ 应比 $x = f_2(y)$ 具有更强的预测能力。神经编码器会为变量 X 和 Y 生成隐藏表示 H_X 和 H_Y，并将它们输入两个回归变量，分别尝试学习函数 f_1 和 f_2。随后，分类器根据结果输出两个变量之间的关系类型（因果、反因果或无关联）。

4.4 变分模型

该**变分因果网络(VCN)** [4] 是一种从数据中迭代学习因果图邻接矩阵的方法。该图通过将其反复输入由 θ 参数化的长短期记忆网络 (LSTM) q_θ 来自回归地构建，该网络生成图 G。模型最大化以下证据下界(ELBO)损失：

$$\mathcal{L} = \mathbb{E}_{q_\theta(\mathcal{G})}[\log P(\mathcal{D}|\mathcal{G})] - \mathrm{KL}(q_\theta(\mathcal{G})||P(\mathcal{G})) \tag{16}$$

其中，$\mathcal{D}$ 是数据分布，而 $P(\mathcal{D}|\mathcal{G})$ 表示给定因果图时获得数据的概率。损失函数的第一项验证了 $\mathcal{G}$ 能够生成数据分布 $\mathcal{D}$。$P(\mathcal{G})$ 是生成的因果图的先验分布，即关于因果图的先验知识（此处为无环图）。第二项使用 Kullback-Leibler散度来确保生成的图符合先验分布。VCN 使用了变分模型所用证据下界损失的一个修改版本 [87]。后者也用于因果推断（第5.6节）和因果表征学习（第6.1节）。

4.5 大型语言模型

近期有一系列研究考察了大语言模型（LLMs）执行因果结构发现的能力。这些研究聚焦于上下文学习 [198],，即通过提示模型，根据对问题的简短描述和几个示例来回答问题。研究并未针对给定任务进行显式训练。Zecevic 等人 [210] 对多个语言模型进行了实验，其中包括广受欢迎的 GPT-3 [17]。当被提示判断两个概念之间是否存在因果联系时，LLM 的回答大多准确。然而，回答会因提示措辞不同而发生显著变化，模型所掌握的知识也不一致。尽管这些实验表明，当前的 LLM 不具备稳健的因果推理能力，但也凸显出评估模型究竟理解了因果联系还是仅仅复述它们的困难。特别是，作者认为，LLM 可以通过学习训练数据中真实因果图的结构，在 CHT 的 L_2 和 L_3 查询上表现良好。为评估 LLM 能否发现新的因果结构（即训练数据中未泄露的结构），Jin 等人 [81] 提出了一个抽象的因果发现评测基准。作者发现，包括 GPT-3 在内的 LLM 表现不佳。若在训练集上进行微调，其表现会大幅提升；但如果测试集的分布与训练集不同，这种提升便不复存在，这表明 LLM 主要依赖非因果线索。

4.6 联结问题

虽然这不是本节的重点，我们还是简要探讨一下通过合并多个数据来源来恢复因果结构的问题 [25, 117, 174, 177]。这被称为**结构因果边缘问题** [62]（即合并若干共享某些未知节点的结构因果模型）。一些研究还尝试生成带环的因果图 [2, 15, 46, 145]，或识别非平稳数据中的不变机制 [76, 158]。

5 因果推断

因果结构发现方法可以恢复变量之间的联系，但无法估计因果量，也无法解释感兴趣的变量如何在因果上相互影响。我们将因果推断描述为计算因果效应（针对观察数据、干预事件或反事实事件）或构建能够计算这些效应的函数这一问题。本节将全面概述并解释因果推断方法。

5.1 do演算

do 演算*do* 演算 [127] 为将干预查询化简为仅凭观察数据即可求解的概率形式建立了推断规则。do 演算本身并未直接提供因果推断的计算模型，但可与此类模型结合使用，以估计因果效应。只有在确定了当前问题的因果结构后，才能应用 do 演算。do 演算的三条规则见公式 17、18和 19。

- **规则 1** 删除观测

$$P(y|do(x),z,w) = P(y|do(x),w) \text{ if } (Y \perp\!\!\!\perp Z|X,W)_{\mathcal{G}_{\overline{X}}} \tag{17}$$

- **规则 2** 将干预简化为观测

$$P(y|do(x),do(z),w) = P(y|do(x),z,w) \text{ if } (Y \perp\!\!\!\perp Z|X,W)_{\mathcal{G}_{\overline{X}\underline{Z}}} \tag{18}$$

- **规则 3** 删除干预

$$P(y|do(x),do(z),w) = P(y|do(x),w) \text{ if } (Y \perp\!\!\!\perp Z|X,W)_{\mathcal{G}_{\overline{X},\overline{Z(W)}}} \tag{19}$$

其中，G 是描述问题结构的因果有向无环图。简化规则基于 G 的结构和**d-分离** [125]。节点 Z（或代表因果变量的一组节点）*d*-分离 （或阻断）两个节点 X和 Y，如果后两者在给定 Z的条件下相互独立（对于图 G 所表示的所有分布均如此）。

规则 1（17）指出，如果在给定某些背景干预（对 X）和观察数据（W）的条件下，Z 与 Y 相互独立，那么可以忽略对 Z 的观察。G_X 表示对 X 进行干预后的图 G，即移除了 X 的入边。值得注意的是，X 和 W 也可以是（可能为空的）变量集。规则 2（18）指出，如果 Z 与 Z 没有共同原因（或其共同原因已被 d 分离），那么对 Z 进行干预等同于观察 Z。如果移除从 Z 到 Y 的有向路径（G_Z 移除 Z 的出边）后，它们变得相互独立，就意味着从 Z 到 Y 的唯一路径是有向路径。由于 **do(*Z*)** 操作会强制设定 Z 的值，并使其与其原因断开，因此该干预没有影响。规则 3（19）指出，如果从 Z 到 Y不存在有向路径，那么可以忽略对 Z 的干预。如果 Z 不是 W 的祖先，$G_{Z(W)}$就会移除 Z的入边。这是为了避免捕捉到 $Y \to Z$ 路径。这些规则使我们能够仅凭观察数据估计干预设定下某些变量的值。查询能否得到解答，取决于图的局部形状。当且仅当查询可以使用 do 演算的三条规则 [77] 化简为观察数据时，该查询才是可识别的。许多关于可识别性和



（a）前门路径　　（b）后门路径

图 4. 前门路径和后门路径示例。隐藏混杂因素 C 虽未知，但在这些图结构下，仍可使用 do 演算回答查询 $P(y|do(x))$。



（a）不可识别模型　　（b）可识别（前门路径）

图 5. 因果图示例。吸烟（S）是癌症（C）的父因，对应遗传因素的隐藏变量 (G)起混杂效应。如果能够通过中间变量（T）——肺部中的焦油——建立因果联系，该模型便可识别。

不可识别性模式已在 Pearl [127]中讨论过。其中最著名的称为**前门**和**后门**路径，如图4所示。

根据　　前门准则，可以将 do 演算规则推导为以下公式：

$$P(y|do(x)) = \sum_{z \in Z} P(z|x) \sum_{x' \in X} P(y|x', z)P(x') \tag{20}$$

根据后门准则，可以将 do 演算规则推导为以下公式：

$$P(y|do(x)) = \sum_{z \in Z} P(y|x, z)P(z) \tag{21}$$

我们继续讨论图 2 中的示例，现在的目标是估计感兴趣变量的值。已知存在一张将吸烟（S）与癌症（C）链接起来的潜在因果图，我们能否估计吸烟对癌症的影响？或者它们是否存在一个未观察到的共同遗传原因 (G)？如果没有其他信息（图5a），仅凭观察数据无法得出结论，因此有必要进行干预研究（通过随机对照试验）。然而，如果我们掌握了其他信息，例如图5b所示的肺部中存在焦油（T），那么我们就处于前门路径的情形。我们可以使用公式 20 来求解查询：$P(c|do(s)) = \sum_t P(t|s) \sum_{s'} P(c|s', t)P(s')$，对应的问题是“……是什么

患者开始吸烟后患癌症的概率是多少？”。值得注意的是，ATE 或 ITE（来自 Rubin 模型）恰好对应我们要寻找的答案，因为它比较的是 $P(c|do(s))$ 与 $P(c|do(\neg s))$。do 演算可以完整回答因果查询，但它需要有关因果图结构和数据分布的先验知识，这限制了它在实践中的应用。

5.2 重新加权

现在，我们讨论在没有图模型的情况下进行因果推断的技术。我们旨在训练一个模型（例如，使用机器学习）来计算来自数据的干预查询。

$P(y|do(x))$来自数据。重新加权方法会为数据中的样本分配权重，以抵消选择偏差，并使分布更接近独立同分布（i.i.d.）。罕见的情况会被赋予较高权重，而经常发生的情况则会被赋予较低权重。重新加权方法背后的隐含假设是存在一条后门路径（图4b），其中Z变量对应于分布偏差。这类技术近似于方程21 中的后门公式，其中 $P(z)$ 是与当前样本（x， y）关联的权重。可以从 Rubin 模型的角度理解重新加权方法。在这种情况下，目标是使背景变量在不同处理组中的分布相同。例如，在医学应用中，如果数据主要由男性患者构成，女性患者较少，那么重新加权就会为女性患者分配较高权重。如果其他背景变量也不平衡，也应加以校正。这通常通过平衡得分 [202]来实现。广泛使用的一种评分指标是**倾向得分** [144]，它表示在给定协变量 $X = 1$的条件下，接受处理 z的概率，定义如下：

$$e(z) = P(X = 1|Z = z) \tag{22}$$

一种常见的为每个样本分配权重的方法是使用**逆倾向加权**（IPW）估计量 [144]。给定样本的权重 w取决于患者是否接受处理：若患者接受处理（$X = 1$），则除以倾向得分；若患者未接受处理（$X = 0$），则除以倾向得分的逆概率。其定义如下：

$$w = \frac{X}{e(z)} + \frac{1 - X}{1 - e(z)} \tag{23}$$

其他估计量则致力于在倾向得分难以轻易估计时提高模型的稳健性，例如**双重稳健**（DR）估计量 [140], ；或者更好地估计倾向得分，例如**协变量平衡倾向评分**（CBPS） [80]。然而，在存在隐藏混杂因素的情况下，这一问题仍然难以解决。

5.3 匹配

匹配方法试图通过将样本划分为协变量完全相同的子组（或个体），来再现随机试验的条件。随后，在单个子组内进行推断。处理组中的每个个体都会与对照组中具有相同背景变量的另一个个体进行匹配。匹配背后的直观思想是，由于同一子组内的样本具有相同的协变量，因此只有处理的实施情况能够区分处理组个体和对照组个体 [6, 202]。将样本分配到子组的一种常用技术是**倾向得分匹配** [144]。倾向得分被用作组分配的距离度量： $D(z_i, z_j) = |e(z_i) - e(z_j)|$。这种分配可以通过最近邻搜索、子分类或加权调整来完成 [170]。然而，近期研究表明，PSM 可能造成不平衡、模型依赖和偏差 [86]。其他匹配方法不使用倾向得分，可以缓解这一问题 [79, 95, 202]。

5.4 决策树

决策树 是另一种广泛用于因果推断的结构。我们遵循潜在结果框架的设定，旨在根据背景变量$Z = z$预测处理的结果 Y。与匹配方法类似，其原理依赖于学习在单一背景协变量组 z 内运作的函数。这类模型的一个著名例子是**分类与回归树** 模型（CART） [102],，其描述为 $Y = g(z; \mathbf{T}, \mathbf{M}) + \epsilon$，其中$\mathbf{M} = \{\theta_1, \ldots\ldots, \theta_B\}$是一组参数，每个参数对应决策树 T 的一个 B 终端节点。在 Z和 Y之间进行回归，由线性函数 $g(\cdot)$表示。该函数的参数化方式为

由 θ_i 与所到达的终端节点相对应的值决定（具体取决于 z 值）。 $\epsilon \sim N(0, \sigma^2)$ 是一个噪声项。树和参数通过马尔可夫链蒙特卡洛（MCMC）算法进行学习。**贝叶斯加性回归树**（BART） [24] 模型是基于随机森林对 CART 的扩展。BART 并非依赖单棵树，而是由 T 棵树组成，并将每棵树的输出相加：

$$Y = \sum_{j=1}^{T} g(x; \mathcal{T}_j, \mathbf{M}_j) + \epsilon \tag{24}$$

依靠多个模型的组合可以缩小树的规模。在 CART 方法中，每棵树的复杂度都远低于原始树，并充当弱分类器。由于使用了多棵树，BART 可以纳入许多协变量效应，因此比单树模型更稳健，并能获得很高的准确率 [202]。

5.5 神经因果模型

第3节所述神经因果模型（NCM）的一种改进版本称为可处理NCM（TNCM） [208],，它用**和-积网络**（SPN） [132, 151]取代多层感知机。SPN 是一种有向无环图，具有单一根节点，叶节点表示因果变量$\mathbf{X}$上的概率分布。在 SPN 中，根节点表示输出节点。在因果推断的语境下，它包含查询 y 的值 $P(y|x)$。中间节点则分别表示求和或求积参数化运算。

GNN-SCM [209] 是另一种神经因果模型，它用图神经网络取代因果推断中的参数化函数（即多层感知机）。图神经网络（GNN） [89, 153]是一种在图结构上运行的特定类型的神经网络。GNN 的一层以图作为输入，图由带有特征向量的节点和边列表组成；该层根据参数化函数和图结构聚合节点特征，并返回聚合结果。GNN 方程的一般形式见方程25：

$$\mathbf{h}_V^{(l+1)} = \phi_{\theta_2}\left(\mathbf{h}_V^{(l)}, \bigoplus_{U \in \mathcal{N}_V} \gamma_{\theta_1}(\mathbf{h}_V^{(l)}, \mathbf{h}_U^{(l)})\right) \tag{25}$$

其中，向量 $\mathbf{h}_V^{(l)}$ 表示第 (l) 层节点 V的节点特征，N_V是 V的邻居节点集合（即存在指向该节点的出边 $U \to V$的节点 U），而 $\bigoplus$ 是通用聚合算子。例如，它可以表示求和 $\sum$ 或求积 $\prod$。GNN层由函数γ和 ϕ组成，其参数为 $\theta = \{\theta_1, \theta_2\}$。

GNN的图结构可自然地用于结构因果模型：因果图对应于GNN的输入图，每个节点都是一个因果变量。节点V的邻居对应于有向无环图中的父节点 $\mathbf{pa}_V$，而GNN的输出为 $P(V|\mathbf{pa}_V)$。按照上述GNN公式，可以自然地定义对GNN中节点V的干预：从N_V中移除 V的父节点集合。新的邻居集合定义为 $\mathrm{M}_V = \{j | j \in \mathrm{N}_V \setminus \mathbf{pa}_V\}$。与上述模型不同，GNN-SCM对所有节点推断都使用相同的GNN映射函数。它只修改输入图，使其与目标节点 V的父节点 $\mathbf{pa}_V$ 相匹配。然而，使用单个全局函数来表示许多局部机制会使优化变得困难。因此，作者在其研究中不考虑反事实 L_3 查询。

5.6 变分图自编码器

NCM 为构建基于神经网络的因果模型提供了一个理论框架，但这并不是使用深度神经网络表征因果关系的唯一方式。我们介绍另一条

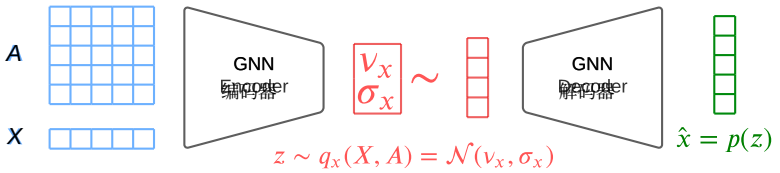


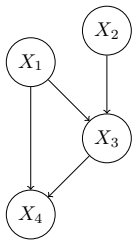
图 6. 变分图自编码器流程。将特征矩阵 X 和邻接矩阵 A 作为输入馈送给图神经网络编码器，以生成多元正态分布的参数，并从中采样得到编码向量 z 。图神经网络解码器以 z 和 A为输入，生成近似值 $\hat{X}$ ，以重构原始特征 X。

基于变分自编码器（VAEs） [87] 和图神经网络的研究路线。VAEs 也广泛用于表征学习，正如第6节所述。

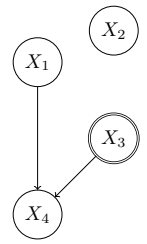
变分自编码器（VAE）是一种通过优化证据下界（ELBO；第 4.4 节简要介绍过一种类似的损失函数）进行训练的编码器—解码器架构。VAE 的核心思想是将数据编码为潜在分布。解码器从该分布中采样，并将数据解码还原。潜在分布是一个多元正态分布，其参数由编码器输出。编码器和解码器都是神经网络。它们的参数通过优化两个项进行学习：（i）数据重构项确保解码器的输出接近输入数据；（ii）正则化项迫使潜在分布服从正态分布（使用 Kullback-Leibler散度）。VAE 通常应用于图像等网格数据或表格数据，但 Kipf 和 Welling [88] 将 VAE 模型应用于图结构。这种模型称为变分图自编码器（VGAE），如图 6 所示。它用于为较稀疏的输入特征矩阵 X 生成嵌入稠密向量 z。图神经网络编码器q学习生成 z $\sim$N (ν, σ) 的分布，而图神经网络解码器 p学习重构 X。编码器分为两项 ν = GNN$_\nu$(X, A) 和 σ = GNN$_\sigma$(X, A)。VGAE 的工作方式与 VAE 类似，但编码器和解码器还会接收额外的邻接矩阵 A。因此，编码后的 z 与 X 的变量数量相同 N，但维度更少。

干预式VGAE（iVGAE） [209] 是一种能够执行 L_2 因果推断的因果模型，其基础是VGAE架构。假设存在一个因果图 G（如图 7a 所示），以及一个表示 G 中因果变量取值的矩阵 X，重构项 $\hat{X} = p(q(X, A_{\mathcal{G}}), A_{\mathcal{G}})$ 会基于每个变量 $X \in X$ 的父节点 pa$_X$，计算 L_1 查询。这得益于编码器-解码器的图神经网络特性，它在推断时遵循图结构。假设因果图正确、完整且不存在隐藏混杂因素，iVGAE 会计算查询 $P(y|x)$ $\forall y \in X$ $x \in \text{pa}_Y \subset X$。现在，如果我们从图 G 中删去特定的边，就能够对 G 执行 do干预。例如，图 7b 中的图 G′ 对应于对 X_3 的干预。在图 G′ 下，$\hat{X} = p(q(X, A_{\mathcal{G}'}), A_{\mathcal{G}'}) \approx p(X|\text{do}(X_3 = x_3))$。iVGAE 是一种用作因果模型的VGAE架构。通过删去输入图中的边并强制设定干预节点的取值，即可执行 L_2 因果推断。然而，iVGAE 无法对外生变量进行边缘化，也无法计算 L_3 查询。

变分因果图编码器（VACA） [152] 是另一种基于变分图自编码器架构、用于因果推断的模型。VACA 的定义与 iVGAE 类似，但增加了额外约束，使其能够进行反事实推断，并假设模型中不存在隐藏混杂因素：第一项约束针对解码器，解码器应至少包含 l-1 个隐藏层，其中 l 是因果图 G 中最长的有向路径。这确保图神经网络中的消息传播能够将因果效应传递到每个节点。第二项约束针对编码器，编码器不应包含隐藏层。这确保每个因果变量的编码分布仅取决于其父节点。VACA 通过修改因果图来执行干预，与 iVGAE 的做法类似，



(a) 原始图 G。



(b) 截断图 G$_{\prime}$。

图 7. 对图 G 的节点 X_3 进行干预。

将因果图 G 转换为图 G′。反事实推断通过结合两个输入来完成。如第 2.1.3 节所示，反事实推断需要溯因、行动和预测。首先，使用 X 的取值和图 G 对一组潜变量 Z^{fact} 进行编码（溯因）；其次，使用 X$^{X_i=x_i}$对另一组 Z$_{inter}$进行编码，即在对感兴趣变量 X_i 实施 do($X_i = x_i$)干预后的 X，以及修改后的图 G′。将与被干预变量 X_i的编码值相对应的向量 z$_i$合并到集合 Z^{fact}中，形成 Z。随后将这个新集合输入解码器，解码器会返回每个因果变量的期望值（预测）。如果遵守上述约束和无混杂假设，那么 VACA 就是一个能够表示干预分布和反事实分布的神经因果模型。即使存在混杂因素，只要图神经网络具有足够的表达能力，仍可近似回答某些查询 [152]。因果效应变分自编码器（CEVAE） [104]和可识别治疗条件变分自编码器（Intact-VAE） [196] 是基于变分自编码器的模型，无需使用图结构即可进行因果效应推断。这些方法旨在识别个体处理效应（ITE）。因果图具有简单且已知的结构 $X \rightarrow Y \leftarrow T$（结果 Y 取决于协变量 X 和处理 T），并且可能存在隐藏混杂因素。深度端到端因果推断（DECI）[49] 模型是一个同时进行因果结构发现和因果推断的流程。它可以避免学习到的因果结构与估计任务之间的不对齐问题。该模型基于因果图结构的先验 p（G）以及用于编码后验分布的通用神经网络 q_θ（A$_{\mathcal{G}}$）来优化证据下界损失。因果推断通过图神经网络架构计算。DECI 假设图是可识别的，且不包含隐藏混杂因素。DECI 融合了第 4.4 节所述 VCN 和变分图自编码器的思想。

5.7 Transformer

The **因果Transformer**(CT) [115] 是一种用于时间序列因果推断的基于Transformer的架构 [182]。它尝试根据三个独立输入恢复反事实结果或个体处理效应 ˆ(ITE) Y_t：过去的协变量 Z$_t$、过去的结果 Y$_{t-1}$，以及过去的干预或处理 X$_{t-1}$，并包含下一步要执行的干预或处理 X_t。每个输入都会馈入一个独立的自注意力子网络。随后，所有输出都会合并到一个带有交叉注意力的联合注意力网络中，并依次馈入两个以对抗方式训练的分类器：第一个分类器经过训练以预测结果 Y_t，第二个分类器则经过训练以预测当前干预X_t（X_t 仅馈入第一个分类器）。这种对抗训练旨在学习一种能够预测结果、但不能预测干预的表征，因为背景变量不应包含任何与干预有关的信息。来自协变量或过去处理的当前处理信息必然源于伪相关。此处采用对抗训练来减轻混杂效应。

5.8 向量自回归模型

人们采用了不同方法利用时间信息进行因果推断，其中最常用的是**向量自回归模型**（VAR）。格兰杰因果关系通常与VAR模型结合使用 [118, 120], ，在该模型中，各变量之间遵循线性关系。一种广泛使用的模型是自回归积分滑动平均（ARIMA）模型。ARIMA由三个部分组成：自回归（AR）部分根据序列 $X(t)$的过去取值和误差项预测下一个取值；滑动平均（MA）部分根据AR的过去误差项进行预测，以考虑噪声项之间的相关性；积分（I）部分则在序列趋势不稳定时使其平稳。这些部分分别由方程26、27和28表示。

$$X_{AR}(t) = c_1 + \epsilon_t + \sum_{i=1}^{p} \alpha_i X(t-i) \tag{26}$$

$$X_{MA}(t) = c_2 + \epsilon_t + \sum_{j=1}^{q} \beta_j \epsilon_{t-j} \tag{27}$$

$$X_I(t) = X(t) - X(t-d) \tag{28}$$

其中， α_i、 β_j、 ϵ_t、 c_1 和 c_2 是可训练参数， p、 q 和 d 是超参数。这三个方程可以组合起来构建 ARIMA 模型：序列 $X(t)$ 经过 I 分量移位后，其结果被输入合并后的 ARMA 方程。通过调整（或将其设为零）AR、I 和 MA 分量的系数，这种分解可以表示多种架构。其他用于因果推断的时间序列建模方法包括动态贝叶斯网络（离散时间贝叶斯网络的扩展）、隐马尔可夫模型或高斯过程，Moraffah 等人 [118]对此进行了综述。

6 因果表征学习

第 3 节介绍了神经因果模型，即利用深度学习技术的因果推理引擎，以及这类模型的优势。然而，深度学习应用于高维低层级数据，而因果推理则针对高层级变量，这些变量通常具有语义含义且维度较低 [57]。从前者转向后者是一项具有挑战性的任务，需要运用表征学习，而生成可用于因果结构发现和推断（即因果变量）的变量，则称为**因果表征学习** [157]。本节将探讨当前针对因果变量的表征学习技术。

6.1 解耦

如前文所述，因果变量代表与其他变量稀疏连接的抽象概念。具体而言，ICM原则指出，因果机制彼此独立，修改其中一个不应影响其他机制 [157]。解缠特征 符合这些性质。它们还只需更少的样本即可达到相近的性能，对噪声不那么敏感，也更加稳健 [181]。因此，对数据中的语义特征进行解耦，或许能够生成适用于因果发现和推断的高层因果变量。本节将回顾解缠特征的基本性质、当前的解耦方法，以及这些方法如何助力因果表征学习。

尽管解耦是深度学习研究中一个备受关注的主题，但目前尚无明确的定义。人们提出了许多定义，但没有一个得到广泛认可。例如，Higgins 等人 [72] 将向量表征定义为：如果它可以分解为

正交子空间，且对一个子空间进行对称修改不会影响其他子空间，那么这种向量表征就是解耦的。尽管目前还没有严格的定义，大多数关于解耦的研究都遵循一些共同原则。研究界普遍认同的解耦核心思想是，解耦表示应将影响数据的**变化因素**彼此分离 [85, 99]。这些因素代表了描述数据的、相互独立且可解释的语义信息 [21, 110]。解耦研究文献中的局部修改概念至关重要，并且可以与因果性中的干预概念相协调 [158]。修改（即对某个解耦变量进行干预）不应影响其他不受该变量因果影响的变量。例如，改变图像中场景的光照，不应改变物体的姿态、方向或标签。文献中常见这一示例，因为解耦研究和应用的核心领域是受控图像生成 [21, 73]。由于对解耦潜在空间进行干预所产生的效果很容易从输出图像中观察到，因此这一领域非常适合用于开发解耦模型。

该领域的大多数文章都依赖变分自编码器，并提出了限制编码空间各维度、使其实现解耦的技术。深度卷积逆图形网络（DC-IGN） [191] 是这一方向上的早期研究。为了实现潜在空间的解耦，训练样本被划分为若干小批量，每个小批量中的图像几乎完全相同，仅有一处差异，对应于某个变化因素的改变（例如方向、颜色、光照）。对于每个小样本，只更新一个维度的参数，以迫使该维度仅表示所需的效果。这种方法需要大量仅测量特征发生变化的样本。在处理真实世界数据集时，这一要求在实践中几乎无法实现。Achille 和 Soatto [1] 使用信息瓶颈拉格朗日函数 [176] 来减少神经网络权重中包含的数据的互信息，并提高潜在表示的不变性和解耦程度。IB 拉格朗日函数定义于方程29中，使用熵 H 和互共享信息 I 作为度量，定义见前文方程8 和9：

$$\mathcal{L} = H(\mathbf{Y}|\mathbf{Z}) + \alpha I(\mathbf{Z};\mathbf{X}) \tag{29}$$

其中，$\mathbf{X}$ 和$\mathbf{Y}$ 分别表示输入图像和输出图像的随机变量集合。$\mathbf{Z}$ 是我们试图解耦的潜在空间。可以注意到，$\mathbf{Z}$ 与隐藏混杂因素发挥着相同的作用，因为它表示生成数据$\mathbf{X}$的底层机制。第一项控制**充分性**，即如果 H 是最小的，那么$\mathbf{Y}$ 就可以完全由$\mathbf{Z}$预测，并且$\mathbf{Z}$ 足以表征$\mathbf{Y}$。第二项控制**最小性**，即 I越小，从$\mathbf{X}$ 共享到$\mathbf{Z}$的信息量就越少。该论文的主要观点是，最小且充分的表征$\mathbf{Z}$ 还应当具有不变性并实现解耦。然而，信息度量并不能衡量表征的质量，而且该表征所包含的信息可能并不比输入$\mathbf{X}$更多。

较新的方法建立在 β-VAE 模型 [73]之上。该模型对变分自编码器（VAE）所学习并最大化的损失函数进行了修改，即证据下界（ELBO） [87], ，如公式30所示：

$$\mathcal{L}_\beta = \mathbb{E}_{q_\phi}[\log p_\theta(\mathbf{X}|\mathbf{Z})] - \beta\,\mathrm{KL}(q_\phi(\mathbf{Z}|\mathbf{X})||P(\mathbf{Z})) \tag{30}$$

其中 p_θ 表示从解码器输出中观察到输入数据分布$\mathbf{X}$ 的概率，该输出以$\mathbf{Z}$的分布为条件；q_ϕ 表示通过编码器从$\mathbf{X}$获得的$\mathbf{Z}$ 的分布，KL表示Kullback-Leibler散度。 ϕ和 θ分别是编码器和解码器学习得到的参数。 $\beta > 0$ 是一个超参数。 β-VAE使用值 $\beta > 1$ ， 而情况$\beta = 1$ 对应于经典的ELBO损失。该论文没有对为何该方法应产生解耦表示作出完整解释，

但后续研究通过分解Kullback–Leibler散度项对此作出了解释。例如， [21, 40, 110]-TCVAE [21]将KL项分解为三个分量之和：

$$\mathrm{KL}(q_\phi(\mathbf{Z}|\mathbf{X})||p(\mathbf{Z})) = I(\mathbf{Z};\mathbf{X}) + \mathrm{KL}(q_\phi(\mathbf{Z})||\prod_j q_\phi(Z_j)) + \sum_j \mathrm{KL}(q_\phi(Z_j)||P(Z_j)) \tag{31}$$

其中，第一项称为互信息（I）分量，表示**X** 与**Z**之间共享的信息。第二项是总相关（TC）分量，是衡量**Z**各维度独立程度的常用指标。第三项是逐维KL散度，用于衡量**Z** 的每个维度是否与其先验不同。 β-TCVAE背后的直觉是，最小化总相关 $\mathrm{TC}(z) = \mathrm{KL}(q(z)||\prod_j q(z_j))$ 能够提高**Z**的解耦程度。解耦表示**Z** 的分布应与其各组成部分的分布乘积相同 Z_j。作者通过实验表明，提高TC的权重可以增强解耦效果。FactorVAE [85] 采用类似的分解方法，以提高**Z**各维度之间的独立性。Disentangled Inferred prior VAE（DIP-VAE） [92] 没有最小化TC项，而是引入了一个基于**X** 和**Z**这两个分布协方差的正则项。

另一类基于生成对抗网络的研究旨在分离与标签相关的变量（有时称为感兴趣因子或属性），以及背景 变量或残余 属性 [12, 48, 68, 129]。这些研究并不试图逐一识别每个变量，而是尝试将它们分离到两个不同的潜在空间中：不变变量$\mathbf{Z}_I$（例如所表示的对象）和特定于域的变量$\mathbf{Z}_D$（例如环境、风格、方向）。

近期，一系列实验和理论分析 [99]表明，在没有关于任务性质的额外假设或归纳偏置的情况下，无法以无监督方式学习解耦表示。作者证明，存在一个无限函数族，它们在最小化 TC 的同时会产生纠缠的**Z** 。

这一发现促使研究从无监督解耦转向半监督方法。自适应组VAE（ada-GVAE） [100]是一种用于解耦的弱监督方法。作者表明，只需很少的假设即可实现弱监督解耦。如果两个样本之间发生变化的因子数量较少，就有可能恢复每个因子的边缘分布，并学习解耦表示。然而，无监督解耦是否可行仍存在争议 [75, 101]。尽管如此，解耦仍在多个领域受到积极研究：视觉（用于可控图像生成、图像到图像翻译或目标检测） [48, 52, 96, 195] , 独立成分分析（ICA） [84, 93], 医学成像 [98], 或图表示 [106]。此外，文本到图像扩散模型 [135, 149] 近期在生成图像方面取得的成功，或将推动解耦领域的发展；该领域主要基于VAE架构或对抗网络。近期关于解耦的研究从ICA和因果性等其他领域汲取灵感。可识别VAE（iVAE，不要与前文介绍的干预式VGAE混淆） [84]和正则化VAE [93] 提出了恢复解耦潜变量及其稀疏因果图链接关系的方法。它们与CEVAE [104]和Intact-VAE [196], 相关，后者在第5.6节中有所介绍，并运用类似思路进行因果推断。DDG网络（解耦约束域泛化） [213] 通过在语义空间和变化空间中施加约束，将领域属性与关注属性解耦。作者利用该方法学习真正解耦的因果变量，并在语义空间或变化空间中进行干预，从而实现数据增强并提升泛化能力。近期提出的一种方法——独立机制分析（IMA）——将ICA与ICM原则相结合 [63]。基于IMA的正则化

旨在恢复并区分导致信号混合物生成的真实来源，并假设每个来源独立地影响数据。研究表明，即使上述假设不成立，该方法仍然有效，并能在存在混杂因素时恢复真实原因。

Schölkopf 等人 [157] 提出，自动解耦是从大规模数据中学习因果机制的必要步骤。这一观点存在争议 [57]。然而，即使这一观点被证明成立，也并不足够，因为提取出高层次的解缠特征后，因果表征学习仍需要恢复特征之间的因果联系。如何找到粒度恰当的特征（在精确性与抽象性之间取得平衡），仍然是一个难题。

6.2 数据增强

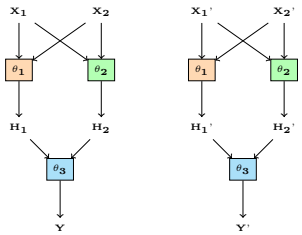
生成对现实的良好表征，很大程度上依赖于模型可用的观察数据。深度学习中的一个基本假设是，训练数据独立同分布（i.i.d.）。如果这一假设成立，模型就只能学习所需的区分性特征，以提升性能。然而，如果这一假设不成立，模型可能会学习到由混杂效应导致的不良行为。迄今讨论的技术均由模型驱动，主要依靠改进所用的架构，从而放宽 i.i.d. 假设。另一方面，数据驱动方法旨在为模型提供更好的数据。具体而言，数据增强可以确保数据保持独立同分布，并且已被证明能够提升深度学习模型的性能 [8]。本节将简要介绍一些方法，旨在为模型训练提供更多或质量更高的数据，使模型能够近似因果行为。

交换干预训练（IIT） [50]是一种通过向神经网络输入干预数据，使其匹配因果模型的训练方法。该过程包含两个步骤：将神经网络的隐藏表示与真实因果模型对齐，并训练网络，使其在反事实设定下匹配因果模型的结果。IIT 将两种因果情境混合在一起，创造出两个并行运行的反事实场景，如图8所示。网络接收两个输入，并在进行推断时交换两个隐藏表示，这分别对应于在每个模型中执行干预。随后，将输出与真实因果模型的反事实结果进行比较。IIT 方法可视为一种数据增强方法，其中增强后的数据会直接插入网络的子部分。然而，这一训练过程要求已知任务的真实结构因果模型（SCM），并将潜在空间分解为多个可映射到因果变量的组成部分。在实践中，给定任务的真实结构因果模型通常未知，这大大限制了该方法的适用性。

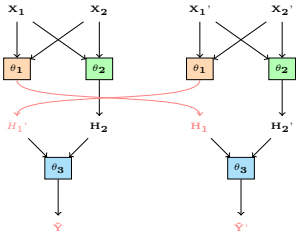
我们将这项工作与一种名为 流形 *mixup* [109, 184, 212] 的数据增强技术联系起来。后者通过对神经网络某一层的输入和输出标签进行插值来实现，如下所示：

$$\tilde{\mathbf{H}}_i = \alpha \cdot \mathbf{H}_i + (1-\alpha) \cdot \mathbf{H}_i' \qquad (32) \qquad\qquad \tilde{\mathbf{Y}} = \alpha \cdot \mathbf{Y} + (1-\alpha) \cdot \mathbf{Y}' \qquad (33)$$

IIT 可以看作流形混合技术的一种特殊情况，其中 $\alpha = 0$ 在方程32中，标签以监督方式利用因果模型确定。上述技术的目标是让神经模型在比训练集所能提供的样本更多样的样本上进行训练，从而提高模型鲁棒性和泛化能力。研究表明，流形混合也能提升少样本任务的表现 [109]。我们可以认为，样本所生成的表征对应于神经网络中的干预。在网络隐藏层中交换或混合输入，会产生



(a) 标准前向流程。



(b) IIT 前向流程。

图 8. 在简单神经网络中进行交换干预训练（IIT）的示意图。两个输入 $\mathbf{X}=\{\mathbf{X}_1, \mathbf{X}_2\}$ 和 $\mathbf{X}'=\{\mathbf{X}'_1, \mathbf{X}'_2\}$ 被输入到参数为 $\theta=\{\theta_1, \theta_2, \theta_3\}$ 的同一个网络中。进行 IIT 时，在将部分隐藏表示送入最后一层之前，会将其进行交换。这里，H_1 和 H'_1。将干预后的输出 Y 和 Y' 与一个和网络对齐的真实结构因果模型在输入相同观测数据和干预数据时得到的值进行比较。

仅通过向网络输入数据无法获得的表征 [50, 184]。这可以增强现有数据增强技术的表达能力，或减少所需样本数量 [8]。

7 应用

因果性领域的发展与科学的发展密切相关。因果性最早出现的记载之一可追溯到亚里士多德的《物理学》第二卷 [42]。因此，因果性在科学的众多广泛领域中都有应用，从生物学到经济学皆是如此。本节将简要概述使用因果方法的领域。附录中提供了更全面的综述。

如前文各节所述，因果性可用于提升深度学习模型的性能。一些文章已将因果关系理论应用于视觉任务 [133], 复杂推理任务 [172], 自动驾驶 [30] 以及因果常识 [56]，以提升这些任务的表现。因果性也应用于生物学，例如对生态系统等动力系统进行建模 [171, 194] ，或将个体的基因表达归因于其表型属性[32, 154, 204]。早期针对时间序列的因果关系理论是为经济应用[59, 166], ，尤其是计量经济学而提出的。这些研究试图利用因果性提取关系，并构建理论经济模型。因果性的另一项重要应用是金融领域，以及对股票市场趋势和危机的预测 [13, 124]。由于格兰杰框架最初是为经济领域开发的，Rubin因果模型则是为医学应用[146], 而建立的，尤其用于研究随机试验的影响。在此类研究中，因果推理可以减少偏差和混杂效应；研究人员认为，应更广泛地采用因果方法，以减少导致错误结论的伪相关 [71]。流行病学中也存在因果模型，用于识别传播疾病的协变量和混杂因素，并对人群中的流行病进行建模 [61, 70]。因果性在气候学中的主要用途，是将气候事件建模为网络。例如，Malik等人 [108]利用网络表征研究印度的降雨。气候学任务颇具挑战，因为其中涉及许多因子，且存在复杂的非线性交互。一些研究指出了基于相关性的方法在这些任务中的局限性 [38]。因果性的应用还见于动物行为学 [20, 51, 103, 175]、神经科学 [45, 91, 142], 和社会科学 [16, 189]。

8 结论

因果性在科学的许多领域都有分支。近年来，深度学习领域对因果性日益关注，因为它有助于应对深度学习面临的关键挑战，包括强泛化能力和复杂因果推理。本综述深入介绍因果性，以及将因果性与深度学习相结合的尝试，旨在为希望将因果性用于深度学习的研究人员提供指导。我们介绍静态数据和时间序列数据的因果关系理论，并综述因果任务所采用的各种方法。我们进一步探讨将深度学习与因果性相结合的最新方法。我们认为，弥合这两个领域之间的差距，有助于解决深度学习模型面临的泛化、模块化、干预推理和反事实推理挑战。本文介绍的所有方法都为构建更出色的因果人工智能模型铺平了道路，助其应对这些挑战，不过因果性的研究视角并不局限于本文所述的范畴。例如，因果性可以帮助洞察并更好地理解现有系统的因果推理能力，并有助于构建更具可解释性的模型。

参考文献

[1] Alessandro Achille 和 Stefano Soatto。2018年。《深度表征中不变性与解耦的涌现》。《机器学习研究杂志》19（2018年），50:1–50:34。http://jmlr.org/papers/v19/17-646.html [2] Ragib Ahsan、David Arbour 和 Elena Zheleva。2022年。《含环的关系因果模型：表征与推理》。载于第1届因果学习与推理会议（CLeaR 2022），红杉会议中心，美国加利福尼亚州尤里卡，2022年4月11日至13日（《机器学习研究论文集》，第177卷），Bernhard Schölkopf、Caroline Uhler 和 Kun Zhang（编）。PMLR，1–18。https://proceedings.mlr.press/v177/ahsan22a.html [3] John E Angus。1994年。概率积分变换及相关结果。SIAM评论36，4（1994年），652–654。[4] Yashas Annadani、Jonas Rothfuss、Alexandre Lacoste、Nino Scherrer、Anirudh Goyal、Yoshua Bengio 和 Stefan Bauer。2021年。变分因果网络：对因果结构进行近似贝叶斯推断。CoRR abs/2106.07635（ 2021 2106 07635 2106 07635 ）。ar Xiv: 。https://arxiv.org/abs/ 。Mike Austin。 。物种分布模型与生态学理论：批判性评估及若干可能的新方法。《生态建模》200，1-2（2007），1–19。[6] Peter C. Austin。2011年。倾向评分方法简介：在观察性研究中降低混杂影响。《多变量行为研究》46，3（2011），399–424。https://doi.org/10.1080/00273171.2011. 568786 arXiv:https://doi.org/10.1080/00273171.2011.568786 PMID: 21818162。[7] Bernard J. Baars。1997年。《意识剧场》。全球工作空间理论：一种严谨的意识科学理论。[8] Randall Balestriero、Ishan Misra 和 Yann LeCun。2022年。数据增强胜过千份样本：基于解析增强样本矩的精确量化。CoRR abs/2202.08325（2022年）。arXiv:2202.08325 https://arxiv.org/abs/ 2202 08325 。Fabien Baradel、Natalia Neverova、Julien Mille、Greg Mori 和 Christian Wolf。2020年。CoPhy：物理动力学的反事实学习。载于第8届国际学习表征会议（ICLR 2020），埃塞俄比亚亚的斯亚贝巴，2020年4月26日至30日。OpenReview.net。https://openreview.net/forum?id=SkeyppEFvS [10] Elias Bareinboim、Juan David Correa、Duligur Ibeling 和 Thomas F. Icard。2022年。论Pearl层级与因果推断的基础。《概率与因果推断》（2022年）。[11] Lionel Barnett、Adam B. Barrett 和 Anil K. Seth。2009年。对于高斯变量，格兰杰因果关系与传递熵等价。《物理评论快报》 103 2009 238701 23 10 1103 103 238701 （12月），。第。期。https://doi.org/。/PhysRevLett.。David Bau、Jun-Yan Zhu、Hendrik Strobelt、Bolei Zhou、Joshua B. Tenenbaum、William T. Freeman 和 Antonio Torralba。2019年。GAN剖析：生成对抗网络的可视化与理解。载于第7届国际学习表征会议（ICLR 2019），美国路易斯安那州新奥尔良，2019年5月6日至9日。OpenReview.net。https: // openreview.net/forum?id=Hyg_X2C5FX [13] Monica Billio、Mila Getmansky、Andrew W Lo 和 Loriana Pelizzon。2012年。金融业和保险业关联性与系统性风险的计量经济学度量。《金融经济学杂志》104，3（2012年），535–559。[14] Rishi Bommasani、Drew A. Hudson、Ehsan Adeli、Russ Altman、Simran Arora、Sydney von Arx、Michael S. Bernstein、Jeannette Bohg、Antoine Bosselut、Emma Brunskill、Erik Brynjolfsson、Shyamal Buch、Dallas Card、Rodrigo Castellon、Niladri S. Chatterji、Annie S. Chen、Kathleen Creel、Jared Quincy Davis、Dorottya Demszky、Chris Donahue、Moussa Doumbouya、Esin Durmus、Stefano Ermon、John Etchemendy、Kawin Ethayarajh、李飞飞、Chelsea Finn、Trevor Gale、Lauren Gillespie、Karan Goel、Noah D. Goodman、Shelby Grossman、Neel Guha、Tatsunori Hashimoto、Peter

Henderson、John Hewitt、Daniel E. Ho、Jenny Hong、Kyle Hsu、Jing Huang、Thomas Icard、Saahil Jain、Dan Jurafsky、Pratyusha Kalluri、Siddharth Karamcheti、Geoff Keeling、Fereshte Khani、Omar Khattab、Pang Wei Koh、Mark S. Krass、Ranjay Krishna、Rohith Kuditipudi 等。2021年。基础模型的机遇与风险。CoRR abs/ 2108 07258 2021 2108 07258 2108 07258。（）。arXiv: 。https://arxiv.org/abs/ 。Stephan [15] Bongers、Patrick Forré、Jonas Peters 和 Joris M Mooij。2021。含环和潜变量的结构因果模型基础。《统计年刊》49，5（2021年），2885–2915。[16] Neil R. Bramley、Tobias Gerstenberg、Ralf Mayrhofer 和 David A. Lagnado。2018年 44，12 2018 1880 1910 10 1037 0000548 。因果结构学习中的时间。（），–。https://doi.org/。/xlm 地点：美国 出版方：美国心理学会。[17] Tom B. Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared Kaplan、Prafulla Dhariwal、Arvind Neelakantan、Pranav Shyam、Girish Sastry、Amanda Askell、Sandhini Agarwal、Ariel Herbert-Voss、Gretchen Krueger、Tom Henighan、Rewon Child、Aditya Ramesh、Daniel M. Ziegler、Jeffrey Wu、Clemens Winter、Christopher Hesse、Mark Chen、Eric Sigler、Mateusz Litwin、Scott Gray、Benjamin Chess、Jack Clark、Christopher Berner、Sam McCandlish、Alec Radford、Ilya Sutskever 和 Dario Amodei。2020年。语言模型是少样本学习者。载于《神经信息处理系统进展》第33卷：神经信息处理系统年度会议2020（NeurIPS 2020），2020年12月6日至12日，线上，Hugo Larochelle、Marc'Aurelio Ranzato、Raia Hadsell、Maria-Florina Balcan 和 Hsuan-Tien Lin（编）。https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html [18] Peter Bühlmann、Markus Kalisch 和 Marloes H Maathuis。2010年。高维线性模型中的变量选择：部分忠实分布与PC-simple算法。《Biometrika》97，2（2010年），261–278。[19] Tommaso Caselli 和 Piek Vossen。2017年。事件故事线语料库：因果与时间关系抽取的新基准。载于ACL 2017新闻中的事件与故事研讨会论文集，加拿大温哥华，2017年8月4日，Tommaso Caselli、Ben Miller、Marieke van Erp、Piek Vossen、Martha Palmer、Eduard H. Hovy、Teruko Mitamura 和 David Caswell（编）。计算语言学协会，77–86。https://doi.org/10.18653/v1/ w17-2711[20] Duxin Chen、Mingyu Kang 和 Wenwu Yu。2020年。鸽群协同运动的概率因果推断。《EPL》（《欧洲物理快报》） 130, 2 5 2020 28004 10 1209 0295 5075 130 28004 （），。https://doi.org/。/-// [21] Tian Qi Chen、Xuechen Li、Roger B. Grosse 和 David Duvenaud。2018。分离变分自编码器中解耦的来源。载于《神经信息处理系统进展》第31卷：神经信息处理系统年度会议2018（NeurIPS 2018），2018年12月3日至8日，加拿大蒙特利尔，Samy Bengio、Hanna M. Wallach、Hugo Larochelle、Kristen Grauman、Nicolò Cesa-Bianchi 和 Roman Garnett（编）。2615–2625。https: //proceedings.neurips.cc/paper/2018/hash/1ee3dfcd8a0645a25a35977997223d22-Abstract.html[22] Ziyue Chen、Miaoqing Xu、Bingbo Gao、George Sugihara、Feixue Shen、Yanyan Cai、Anqi Li、Qi Wu、Lin Yang、Qi Yao 等。2022年。复杂大气环境中的因果推断。《环境污染》303（2022年），119057。[23] David Maxwell Chickering。2002年。贪心搜索下的最优结构识别。《机器学习研究杂志》3（2002年），507–554。http://jmlr.org/papers/v3/chickering02b.html [24] Hugh A. Chipman、Edward I. George 和 Robert E. McCulloch。2010年。BART：贝叶斯加性回归树。《应用统计年刊》 4, 1 2010 266 298 10 1214 09 285 （），–。https://doi.org/。/-AOAS Tom [25] Claassen 和 Tom Heskes。2010。通过不同实验在多个模型中发现因果关系。载于《神经信息处理系统进展》第23卷：神经信息处理系统第24届年度会议论文集，会议于2010年12月6日至9日在加拿大不列颠哥伦比亚省温哥华举行，John D. Lafferty、Christopher K. I. Williams、John Shawe-Taylor、Richard S. Zemel 和 Aron Culotta（编）。Curran Associates, Inc.，415–423。https://proceedings.neurips.cc/paper/2010/hash/5807a685d1a9ab3b599035bc566ce2b9-Abstract.html [26] Diego Colombo 和 Marloes H. Maathuis。2014 年。与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, 1 2014 3741 3782 10 5555 2627435 2750365 （），–。https://doi.org/。/。Diego Colombo、 [27] Marloes H. Maathuis、Markus Kalisch 和 Thomas S. Richardson。2012年。利用潜变量和选择变量学习高维有向无环图。 40, 1 2012 10 1214 11 940 （）。https://doi.org/。/- [28] AOS Gregory F. Cooper 和 Edward Herskovits。1992。从数据归纳概率网络的贝叶斯方法。《机器学习》 9 1992 309 347 10 1007 00994110 （），–。https://doi.org/。/BF Ishita Dasgupta、 [29] Jane X. Wang、Silvia Chiappa、Jovana Mitrovic、Pedro A. Ortega、David Raposo、Edward Hughes、Peter W. Battaglia、Matthew M. Botvinick 和 Zeb Kurth-Nelson。2019年。从元强化学习进行因果推理。CoRR abs/ 1901 08162 2019 1901 08162 1901 08162。（）。arXiv: 。http://arxiv.org/abs/ 。 [30] Pim de Haan、Dinesh Jayaraman 和 Sergey Levine。2019。模仿学习中的因果混淆。载于《神经信息处理系统进展》第32卷：神经信息处理系统年度会议2019（NeurIPS 2019），2019年12月8日至14日，加拿大不列颠哥伦比亚省温哥华，Hanna M. Wallach、Hugo Larochelle、Alina Beygelzimer、Florence d'Alché-Buc、Emily B. Fox 和 Roman Garnett（编）。11693–11704。https://proceedings.neurips.cc/paper/2019/hash/ 947018640bf36a2bb609d3557a285329-Abstract.html

[31] Ksander N De Winkel、Mikhail Katliar、Daniel Diers 和 Heinrich H Bülthoff。2018。垂直性感知中的因果推断。《科学报告》8, 1（2018），1–12。[32] Franck Delaplace、Hanna Klaudel 和 Amandine Cartier-Michaud。2010。生物网络的离散因果模型视角。载于《计算系统生物学方法国际会议》第八届会议论文集。4–13。[33] Li Du、Xiao Ding、Kai Xiong、Ting Liu 和 Bing Qin。2022。e-CARE：用于探索可解释因果推理的新数据集。载于计算语言学协会第60届年会论文集（第一卷：长篇论文），ACL 2022，爱尔兰都柏林，2022年5月22日至27日，Smaranda Muresan、Preslav Nakov 和 Aline Villavicencio（编）。计算语言学协会，
432 446 10 18653 1 2022 33 –。https://doi.org/ 。 /v / 。
[34] acl-long。Nan Du、Yanping Huang、Andrew M. Dai、Simon Tong、Dmitry Lepikhin、Yuanzhong Xu、Maxim Krikun、Yanqi Zhou、Adams Wei Yu、Orhan Firat、Barret Zoph、Liam Fedus、Maarten Bosma、Zongwei Zhou、Tao Wang、Yu Emma Wang、Kellie Webster、Marie Pellat、Kevin Robinson、Kathy Meier-Hellstern、Toju Duke、Lucas Dixon、Kun Zhang、Quoc V. Le、Yonghui Wu、Zhifeng Chen 和 Claire Cui。2021。GLaM：通过混合专家高效扩展语言模型。（ 2021 2112 06905 2112 06905 ）。arXiv: 。
[35] 2022 http://arxiv.org/abs/ 。Jiafei Duan、Arijit Dasgupta、Jason Fischer 和 Cheston Tan。。《直觉物理建模的机器学习方法综述》。载于第三十一届国际人工智能联合会议论文集，IJCAI 2022，奥地利维也纳，2022年7月23日至29日，Luc De Raedt（编）。ijcai.org，5444–5452。https://doi.org/10.24963/ijcai. 2022/763[36] Otis Dudley Duncan。2014。《结构方程模型导论》。爱思唯尔。[37] Selma Dündar-Coecke、Gideon Goldin 和 Steven A Sloman。2022。没有机制的因果推理。《PLoS One》17, 5（2022），e0268219。[38] Imme Ebert-Uphoff 和 Yi Deng。2012。利用图模型发现气候研究中的因果关系。《气候学报》25, 17（2012），5648–5665。[39] Doris Entner 和 Patrik Hoyer。2010。使用FCI从时间序列数据中发现因果关系。第五届欧洲概率图模型研讨会论文集，PGM 2010（2010年9月）。[40] Babak Esmaeili、Hao Wu、Sarthak Jain、Alican Bozkurt、N. Siddharth、Brooks Paige、Dana H. Brooks、Jennifer G. Dy 和 Jan-Willem van de Meent。2019。结构化解耦表示。载于第二十二届人工智能与统计国际会议，AISTATS 2019，日本冲绳那霸，2019年4月16日至18日（《机器学习研究论文集》第89卷），Kamalika Chaudhuri 和 Masashi Sugiyama（编）。PMLR，2525–2534。http: //proceedings.mlr.press/v89/esmaeili19a.html [41] John Essington。2022。我们如何学习：为什么大脑比任何机器都更善于学习……至少目前如此。《教育评论》74, 4（
2022 899 899 10 1080 00131911 2021 1930914 10 1080 00131911 2021 1930914 ），–。
[42] 2006 2006 https://doi.org/ 。 / 。 。arXiv:https://doi.org/ 。 / 。 。Andrea Falcon。《亚里士多德论因果性》。（）。[43] William Fedus、Barret Zoph 和 Noam Shazeer。2021。Switch Transformer：以简单高效的稀疏性扩展至万亿参数模型。（ 2021 2101 03961 2101 03961 ）。arXiv: 。
[44] 2007 http://arxiv.org/abs/ 。Ildikó Flesch 和 Peter JF Lucas。《贝叶斯网络中的马尔可夫等价》。施普林格柏林海德堡，德国柏林，
3 38 10 1007 978 3 540 68996 6_1 –。https://doi.org/ 。 / - - - - Alex Fornito、
[45] 2013 Andrew Zalesky 和 Michael Breakspear。《人类连接组的图分析：前景、进展与陷阱》。NeuroImage 80（2013），426–444。[46] Patrick Forré 和 Joris M. Mooij。2018。针对含环和潜在混杂因素的非线性结构因果模型的基于约束的因果发现。载于第三十四届人工智能不确定性会议论文集，UAI 2018，美国加利福尼亚州蒙特雷，2018年8月6日至10日，Amir Globerson 和 Ricardo Silva（编）。AUAI出版社，269–278。
http://auai.org/uai2018/proceedings/papers/117.pdf [47] Ranran L French 和 Gregory C DeAngelis。2020。多感官神经处理：从线索整合到因果推断。《生理学当前观点》16（2020），8–13。https://doi.org/10.1016/j.cophys.2020.04.004 视觉生理学。[48] Aviv Gabbay、Niv Cohen 和 Yedid Hoshen。2021。一张图胜过千言万语：迈向真实环境中的解耦。载于《神经信息处理系统进展》第34卷：2021年神经信息处理系统年会，NeurIPS 2021，2021年12月6日至14日，线上，Marc'Aurelio Ranzato、Alina Beygelzimer、Yann N. Dauphin、Percy Liang 和 Jennifer Wortman Vaughan（编）。9216–9228。https://proceedings. neurips.cc/paper/2021/hash/4cef5b5e6ff1b3445db4c013f1d452e0-Abstract.html[49] Tomas Geffner、Javier Antoran、Adam Foster、Wenbo Gong、Chao Ma、Emre Kiciman、Amit Sharma、Angus Lamb、Martin Kukla、Nick Pawlowski、Miltiadis Allamanis 和 Cheng Zhang。2022。深度端到端因果推断。CoRR abs/ 2202 02195 2022 2202 02195 2202 02195 。（）。arXiv: 。
[50] https://arxiv.org/abs/ 。Atticus Geiger、Zhengxuan Wu、Hanson Lu、Josh Rozner、Elisa Kreiss、Thomas Icard、Noah D. Goodman 和 Christopher Potts。2022。为可解释神经网络诱导因果结构。载于国际机器学习会议，ICML 2022，美国马里兰州巴尔的摩，2022年7月17日至23日（《机器学习研究论文集》第162卷），Kamalika Chaudhuri、Stefanie Jegelka、Le Song、Csaba Szepesvári、Gang Niu 和 Sivan Sabato（编）。PMLR，

7324–7338。https://proceedings.mlr.press/v162/geiger22a.html[51] Gaël Gendron、Yang Chen、Mitchell Rogers、Yiping Liu、Mihailo Azhar、Shahrokh Heidari、David Arturo Soriano Valdez、Kobe Knowles、Padriac O'Leary、Simon Eyre、Michael Witbrock、Gillian Dobbie、Jiamou Liu 和 Patrice Delmas。2023。通过因果结构发现和图神经网络对社会性动物进行行为建模。CoRR abs/ 2312 14333 2023 2312 14333 2312 14333 。 （）。arXiv: 。http://arxiv.org/abs/ 。Gaël Gendron、Michael Witbrock 和 Gillian Dobbie。《通过因果干预解缠结潜在表征》。载于第三十二届国际人工智能联合会议论文集，IJCAI 2023，2023年8月19日至25日，中国澳门特别行政区。ijcai.org，2023。https://doi.org/ 。 / IJCAI. /[53] 2019 Clark Glymour、Kun Zhang 和 Peter Spirtes。《基于图模型的因果发现方法综述》。《遗传学前沿》 10 2019 10 3389 2019 00524 （）。https://doi.org/ 。 /fgene. 。Mariel K Goddu、Tania Lombrozo 和 Alison Gopnik。《变换与迁移：学龄前儿童理解抽象关系，并在因果任务中进行类比推理》。《儿童发展》91, 6（2020），1898–1915。[55] Ian J. Goodfellow、Jonathon Shlens 和 Christian Szegedy。2015。解释并利用对抗样本。载于第三届国际学习表征会议，ICLR 2015，美国加利福尼亚州圣迭戈，2015年5月7日至9日，会议论文集，Yoshua Bengio 和 Yann LeCun（编）。http://arxiv.org/abs/ 1412 6572 。Andrew S. Gordon、Zornitsa Kozareva 和 Melissa Roemmele。《SemEval任务7：合理备选项的选择——常识因果推理评测》。载于第六届国际语义评测研讨会论文集，SemEval@NAACL-HLT 2012，加拿大蒙特利尔，2012年6月7日至8日，Eneko Agirre、Johan Bos 和 Mona T. Diab（编）。计算语言学协会，394–398。https://aclanthology.org/S12-1052/ [57] Anirudh Goyal 和 Yoshua Bengio。2020。高层认知深度学习的归纳偏置。CoRR abs/ 2011 15091 2020 2011 15091 2011 15091 。 （）。arXiv: 。https://arxiv.org/abs/ 。Anirudh Goyal、Alex Lamb、Jordan Hoffmann、Shagun Sodhani、Sergey Levine、Yoshua Bengio 和 Bernhard Schölkopf。2020 2020 1909 10893 1909 10893 。循环独立机制。 （）。arXiv: 。http://arxiv.org/abs/ 。Clive WJ Granger。《利用计量经济模型和互谱方法探究因果关系》。《计量经济学：计量经济学会杂志》（1969），424–438。[60] C. W. J. Granger。2001。《利用计量经济模型和互谱方法探究因果关系》。哈佛大学出版社，美国，31–47。[61] Sander Greenland、Judea Pearl 和 James M Robins。1999。流行病学研究中的因果图。《流行病学》（1999），37–48。[62] Luigi Gresele、Julius von Kügelgen、Jonas M. Kübler、Elke Kirschbaum、Bernhard Schölkopf 和 Dominik Janzing。2022。通过结构因果边缘问题进行因果推断。载于国际机器学习会议，ICML 2022，美国马里兰州巴尔的摩，2022年7月17日至23日（《机器学习研究论文集》第162卷），Kamalika Chaudhuri、Stefanie Jegelka、Le Song、Csaba Szepesvári、Gang Niu 和 Sivan Sabato（编）。PMLR，7793–7824。https://proceedings.mlr.press/v162/gresele22a.html [63] Luigi Gresele、Julius von Kügelgen、Vincent Stimper、Bernhard Schölkopf 和 Michel Besserve。2021。独立机制分析：一个新概念？载于《神经信息处理系统进展》第34卷：2021年神经信息处理系统年会，NeurIPS 2021，2021年12月6日至14日，线上，Marc'Aurelio Ranzato、Alina Beygelzimer、Yann N. Dauphin、Percy Liang 和 Jennifer Wortman Vaughan（编）。28233–28248。https: //proceedings.neurips.cc/paper/2021/hash/edc27f139c3b4e4bb29d1cdbc45663f9-Abstract.html [64] Yong Gu、Gregory C DeAngelis 和 Dora E Angelaki。2012。背侧内侧颞上区神经元与多感觉方向感知之间的因果联系。《神经科学杂志》32, 7（2012），2299–2313。[65] Ruocheng Guo、Lu Cheng、Jundong Li、P. Richard Hahn 和 Huan Liu。2021。《基于数据学习因果性的综述：问题与方法》。《ACM计算调查》 53, 4 2021 75 1 75 37 10 1145 3397269 （），：–：。https://doi.org/ 。 / David Ha 和 Jürgen Schmidhuber。《世界模型》。CoRR abs/ 。 （）。arXiv: .10122 http: // arxiv.org/abs/ 1803 10122 。Trygve Haavelmo。《联立方程组的统计学含义》。《计量经济学：计量经济学会杂志》（1943），1–12。[68] Naama Hadad、Lior Wolf 和 Moni Shahar。2018。两步解缠结方法。载于2018年IEEE计算机视觉与模式识别会议，CVPR 2018，美国犹他州盐湖城，2018年6月18日至22日。计算机视觉基金会 / IEEE计算机学会， 772 780 10 1109 2018 00087 –。https://doi.org/ 。 /CVPR. 。Alain Hauser 和 Peter Bühlmann。《有向无环图干预马尔可夫等价类的刻画与贪心学习（摘要）》。载于UAI 2011，第二十七届人工智能不确定性会议论文集，西班牙巴塞罗那，2011年7月14日至17日，Fábio Gagliardi Cozman 和 Avi Pfeffer（编）。AUAI出版社， 851 =1 =2 _ =2180 _ =27 。https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu &smnu &article id &proceeding id Miguel A Hernán、Sonia Hernández-Díaz、Martha M Werler 和 Allen A Mitchell。《因果知识是混杂评估的前提：其在出生缺陷流行病学中的应用》。《美国流行病学杂志》155, 2（2002），176–184。

[71] Miguel A Hernán、John Hsu 和 Brian Healy。2019年。让因果推断回归正轨的第二次机会：数据科学任务的分类。《Chance》32，1（2019），42–49。[72] Irina Higgins、David Amos、David Pfau、Sébastien Racanière、Loïc Matthey、Danilo J. Rezende 和 Alexander Lerchner。2018年。《迈向解耦表示的定义》。《CoRR》abs/1812.02230（2018）。arXiv:1812.02230 http://arxiv.org/abs/ 1812 02230 。Irina Higgins、Loïc Matthey、Arka Pal、Christopher P. Burgess、Xavier Glorot、Matthew M. Botvinick、Shakir Mohamed 和 Alexander Lerchner。2017年。β-VAE：通过受约束的变分框架学习基本视觉概念。载于第五届国际学习表征会议，ICLR 2017，法国土伦，2017年4月24日至26日，会议论文集。OpenReview.net。https://openreview.net/forum?id=Sy2fzU9gl [73] [74] Kateřina Hlaváčková-Schindler 和 Sergiy Pereverzyev。2015年。Lasso 格兰杰因果模型：其在基因表达调控网络中的若干策略及效率。载于《决策：不确定性、不完美性、审议与可扩展性》。施普林格，91–117。[75] Daniella Horan、Eitan Richardson 和 Yair Weiss。2021年。无监督解耦何时可能？载于《神经信息处理系统进展第34卷》：神经信息处理系统进展年度会议 2021，NeurIPS 2021，2021年12月6日至14日，线上，Marc'Aurelio Ranzato、Alina Beygelzimer、Yann N. Dauphin、Percy Liang 和 Jennifer Wortman Vaughan（编）。5150–5161。https://proceedings.neurips.cc/paper/2021/hash/ 29586cb449c90e249f1f09a0a4ee245a-Abstract.html[76] Biwei Huang、Kun Zhang、Jiji Zhang、Joseph D. Ramsey、Ruben Sanchez-Romero、Clark Glymour 和 Bernhard Schölkopf。2020年。异质/非平稳数据中的因果发现。《机器学习研究杂志》21（2020），89:1–89:53。http://jmlr.org/papers/v21/19-232.html [77] Yimin Huang 和 Marco Valtorta。2006年。Pearl 干预演算是完备的。载于不确定性人工智能会议 2006，《第22届人工智能中的不确定性会议论文集》，美国马萨诸塞州剑桥，2006年7月13日至16日。AUAI出版社。https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu =1 =2 _ =1301 _ =22 &smnu &article id &proceeding id [78] Aapo Hyvärinen 和 Erkki Oja。2000 独立成分分析：算法与应用。《神经网络》13, 4 5 2000 411 430 10 1016 0893 6080 00 00026 5 -（），–。https://doi.org/ . /S -（）- Stefano M. Iacus、Gary King 和 Giuseppe Porro。2012 因果推断中的无平衡性检验：粗化精确匹配。《政治分析》[79] 20, 1 2012 1 24 10 1093 013 （），–。https://doi.org/ . /pan/mpr Kosuke Imai 和 Marc Ratkovic。2014 协变量平衡倾向评分。《英国皇家统计学会杂志：B辑（统计方法学）》76，1（2014），243–263。https://doi.org/10.1111/rssb.12027 arXiv:https://rss.onlinelibrary.wiley.com/doi/pdf/[80] 10 1111 12027 。/rssb。Zhijing Jin、Jiarui Liu、Zhiheng Lyu、Spencer Poff、Mrinmaya Sachan、Rada Mihalcea、Mona T. Diab 和 Bernhard Schölkopf。2023年。大语言模型能从相关性推断因果关系吗？《CoRR》abs/2306.05836（2023）。https://doi.org/ 10 48550 2306 05836 2306 05836 。/[81] ARXIV。arXiv：。Jean Kaddour、Aengus Lynch、Qi Liu、Matt J. Kusner 和 Ricardo Silva。2022年。因果机器学习：综述与开放问题。《CoRR》abs/[82] 2206 15475 2022 10 48550 2206 15475 2206 15475 。（）。https://doi.org/ . /ARXIV。arXiv：。2022 Marcus Kaiser 和 Maksim Sipos。NOTEARS 不适用于处理量纲量时的因果图发现。《神经过程通讯》[83] 54, 3 2022 1587 1595 10 1007 11063 021 10694 5 （），–。https://doi.org/ . /S - - - Ilyes Khemakhem、Diederik P. Kingma、Ricardo Pio Monti 和 Aapo Hyvärinen。2020 变分自编码器与非线性独立成分分析：统一框架。载于第二十三届国际人工智能与统计会议，AISTATS 2020，2020年8月26日至28日，线上 *[*巴勒莫，西西里岛，意大利*]*（《机器学习研究论文集》，第108卷），Silvia Chiappa 和 Roberto Calandra（编）。PMLR，2207–2217。http://proceedings.mlr.press/v108/khemakhem20a。html[84] [85] Hyunjik Kim 和 Andriy Mnih。2018年。通过因子分解实现解耦。载于第三十五届国际机器学习会议，ICML 2018，瑞典斯德哥尔摩，2018年7月10日至15日（《机器学习研究论文集》，第80卷），Jennifer G. Dy 和 Andreas Krause（编）。PMLR，2654–2663。http://proceedings.mlr. press/v80/kim18b.html [86] Gary King 和 Richard Nielsen。2019年。27，4（2019），435–454。10 1017 2019。为什么不应使用倾向评分进行匹配。《政治分析》（），–。https://doi.org/ . /pan。[87] 2014 Diederik P. Kingma 和 Max Welling。自编码变分贝叶斯。载于第二届国际学习表征会议，ICLR 2014，加拿大艾伯塔省班夫，2014年4月14日至16日，会议论文集，Yoshua Bengio 和 Yann LeCun（编）。http://arxiv.org/abs/ 1312 6114 。Thomas N. Kipf 和 Max Welling。2016 变分图自编码器。《CoRR》abs/[88] 1611 07308 1611 07308 。http://arxiv.org/abs/。Thomas N. Kipf 和 Max Welling。半监督分类与图卷积网络。载于第五届国际学习表征会议，ICLR 2017，法国土伦，2017年4月24日至26日，会议论文集。OpenReview.net。[89] 2017 https://openreview.net/forum?id=SJU4ayYgl

[90] Pang Wei Koh、Shiori Sagawa、Henrik Marklund、Sang Michael Xie、Marvin Zhang、Akshay Balsubramani、Weihua Hu、Michihiro Yasunaga、Richard Lanas Phillips、Irena Gao、Tony Lee、Etienne David、Ian Stavness、Wei Guo、Berton Earnshaw、Imran Haque、Sara M. Beery、Jure Leskovec、Anshul Kundaje、Emma Pierson、Sergey Levine、Chelsea Finn 和 Percy Liang。2021年。WILDS：真实环境分布偏移基准。载于第三十八届国际机器学习会议，ICML 2021，2021年7月18日至24日，线上会议（《机器学习研究论文集》，第139卷），Marina Meila 和 Tong Zhang（编）。PMLR，5637–5664。http://proceedings.mlr.press/v139/koh21a.html [91] Konrad P Körding、Ulrik Beierholm、Wei Ji Ma、Steven Quartz、Joshua B Tenenbaum 和 Ladan Shams。2007年。多感官知觉中的因果推断。《PLoS One》2，9（2007），e943。[92] Abhishek Kumar、Prasanna Sattigeri 和 Avinash Balakrishnan。2018年。从无标签观测中对解耦潜在概念进行变分推断。载于第六届国际学习表征会议，ICLR 2018，加拿大不列颠哥伦比亚省温哥华，2018年4月30日至5月3日，会议论文集。OpenReview.net。https://openreview.net/forum?id=H1kG7GZAW[93] Sébastien Lachapelle、Pau Rodríguez、Yash Sharma、Katie Everett、Rémi Le Priol、Alexandre Lacoste 和 Simon Lacoste-Julien。2022年。通过机制稀疏正则化实现解耦：非线性独立成分分析的新原则。载于第一届因果学习与推理会议，CLeaR 2022，美国加利福尼亚州尤里卡红杉会议中心，2022年4月11日至13日（《机器学习研究论文集》，第177卷），Bernhard Schölkopf、Caroline Uhler 和 Kun Zhang（编）。PMLR，428–484。https://proceedings.mlr.press/v177/lachapelle22a.html [94] Zachary M Laubach、Eleanor J Murray、Kim L Hoke、Rebecca J Safran 和 Wei Perng。2021年。生物学家的模型选择与因果推断指南。《英国皇家学会学报B辑》288，1943（2021），20202815。[95] Sheng Li、Nikos Vlassis、Jaya Kawale 和 Yun Fu。2016年。通过降维匹配估计数字营销活动中的处理效应。载于第二十五届国际人工智能联合会议论文集（美国纽约州纽约市）（IJCAI'16）。AAAI出版社，3768–3774。[96] Xinyang Li、Shengchuan Zhang、Jie Hu、Liujuan Cao、Xiaopeng Hong、Xudong Mao、Feiyue Huang、Yongjian Wu 和 Rongrong Ji。2021年。通过层次化风格解耦实现图像到图像翻译。载于IEEE计算机视觉与模式识别会议，CVPR 2021，线上，2021年6月19日至25日。计算机视觉基金会 / 电气电子工程师学会，8639–8648。https://openaccess.thecvf.com/content/CVPR2021/html/Li_图像到图像翻译_通过_层次化风格_解耦_CVPR_论文.html [97] Zhongyang Li、Xiao Ding、Ting Liu、J. Edward Hu 和 Benjamin Van Durme。2020年。因果关系引导的生成。载于第二十九届国际人工智能联合会议论文集，IJCAI 2020，Christian Bessiere（编）。ijcai.org，3629–3636。https://doi.org/10.24963/ijcai.2020/502 [98] Xiao Liu、Pedro Sanchez、Spyridon Thermos、Alison Q. O'Neil 和 Sotirios A. Tsaftaris。2021年。成像领域解耦表示学习教程。《CoRR》abs/2108.12043（2021）。arXiv:2108.12043 https://arxiv.org/abs/2108.12043[99] Francesco Locatello、Stefan Bauer、Mario Lucic、Gunnar Rätsch、Sylvain Gelly、Bernhard Schölkopf 和 Olivier Bachem。2019年。挑战无监督解耦表示学习中的常见假设。载于《机器学习中的可复现性》，ICLR 2019研讨会，美国路易斯安那州新奥尔良，2019年5月6日。OpenReview.net。https://openreview.net/forum?id=Byg6VhUp8V [100] Francesco Locatello、Ben Poole、Gunnar Rätsch、Bernhard Schölkopf、Olivier Bachem 和 Michael Tschannen。2020年。无妥协的弱监督解耦。载于第三十七届国际机器学习会议，ICML 2020，2020年7月13日至18日，线上会议（《机器学习研究论文集》，第119卷）。PMLR，6348–6359。http://proceedings.mlr.press/v119/locatello20a.html [101] Francesco Locatello、Michael Tschannen、Stefan Bauer、Gunnar Rätsch、Bernhard Schölkopf 和 Olivier Bachem。2020年。使用少量标签解耦变化因素。载于第八届国际学习表征会议，ICLR 2020，埃塞俄比亚亚的斯亚贝巴，2020年4月26日至30日。OpenReview.net。https://openreview.net/forum?id=SygagpEKwB [102] Wei-Yin Loh。2011年。分类与回归树。《WIREs数据挖掘与知识发现》1，1（2011），14–23。https://doi.org/10.1002/widm.8 [103] 2016。arXiv:https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.8。Warren E. Lord、Jie Sun、Nicholas T. Ouellette 和 Erik M. Bollt。推断群体动物行为中的因果信息流。《IEEE分子、生物与多尺度通信汇刊》2，1（2016），107–116。https://doi.org/10.1109/TMBMC.2016.2632099 [104] Christos Louizos、Uri Shalit、Joris M. Mooij、David A. Sontag、Richard S. Zemel 和 Max Welling。2017年。利用深度潜变量模型进行因果效应推断。载于《神经信息处理系统进展》第30卷：神经信息处理系统进展年度会议 2017，2017年12月4日至9日，美国加利福尼亚州长滩，Isabelle Guyon、Ulrike von Luxburg、Samy Bengio、Hanna M. Wallach、Rob Fergus、S. V. N. Vishwanathan 和 Roman Garnett（编）。6446–6456。https://proceedings.neurips.cc/paper/2017/hash/94b5bde6de888ddf9cde6748ad2523d1-Abstract.html[105] Josef Ludescher、Maria Martin、Niklas Boers、Armin Bunde、Catrin Ciemer、Jingfang Fan、Shlomo Havlin、Marlene Kretschmer、Jürgen Kurths、Jakob Runge 等。2021年。基于网络的气候现象预测。《美国国家科学院院刊》118，47（2021），e1922872118。

[106] Jianxin Ma、Peng Cui、Kun Kuang、Xin Wang 和 Wenwu Zhu。2019年。解耦图卷积网络。载于《第36届国际机器学习会议论文集》，ICML 2019，2019年6月9日至15日，美国加利福尼亚州长滩（《机器学习研究论文集》第97卷），Kamalika Chaudhuri 和 Ruslan Salakhutdinov（编）。PMLR，4212–4221。http://proceedings.mlr.press/v97/ma19a.html [107] John F Magnotti、Kristen B Smith、Marcelo Salinas、Jacqunae Mays、Lin L Zhu 和 Michael S Beauchamp。2018年。协同发音增强多感官整合的因果推断解释。《科学报告》8，1（2018），1–10。[108] Nishant Malik、Bodo Bookhagen、Norbert Marwan 和 Jürgen Kurths。2012年。利用复杂网络分析南亚季风降雨的时空极端事件。《气候动力学》39，3（2012），971–987。[109] Puneet Mangla、Mayank Singh、Abhishek Sinha、Nupur Kumari、Vineeth N. Balasubramanian 和 Balaji Krishna- murthy。2020年。描绘正确流形：用于少样本学习的流形混合。载于《IEEE冬季计算机视觉应用会议》，WACV 2020，美国科罗拉多州斯诺马斯村，2020年3月1日至5日。IEEE，2207–2216。https://doi.org/ 10 1109 45572 2020 9093338 . /WACV . . Emile [110] Mathieu、Tom Rainforth、N. Siddharth 和 Yee Whye Teh。2019年。在变分自编码器中解开解耦问题。载于《第36届国际机器学习会议论文集》，ICML 2019，2019年6月9日至15日，美国加利福尼亚州长滩（《机器学习研究论文集》第97卷），Kamalika Chaudhuri 和 Ruslan Salakhutdinov（编）。PMLR，4402–4412。http://proceedings.mlr.press/v97/mathieu19a.html[111] Yutaka Matsuo、Yann LeCun、Maneesh Sahani、Doina Precup、David Silver、Masashi Sugiyama、Eiji Uchibe 和 Jun Morimoto。2022年。深度学习、强化学习与世界模型。《神经网络》152（2022），267–275。https://doi.org/ 10 1016 2022 03 037 . /j.neunet. . . [112] 2015 Mariusz Maziarz。《格兰杰因果关系谬误述评》。《哲学经济学杂志：经济与社会问题思考》8，2（2015），86–105。[113] Christopher Meek。1995年。基于背景知识的因果推断与因果解释。载于《UAI '95：第十一届人工智能不确定性年会论文集》，加拿大魁北克省蒙特利尔，1995年8月18日至20日，Philippe Besnard 和 Steve Hanks（编）。Morgan Kaufmann，403–410。https://dslpitt.org/uai/displayArticleDetails. jsp?mmnu =1 =2 _ =458 _ =11 &smnu &article id &proceeding id Christopher Meek。《图 [114] 1997 模型：因果模型与统计模型的选择》。博士论文。卡内基梅隆大学。[115] Valentyn Melnychuk、Dennis Frauen 和 Stefan Feuerriegel。2022年。用于估计反事实结果的因果Transformer。载于《国际机器学习会议》，ICML 2022，2022年7月17日至23日，美国马里兰州巴尔的摩（《机器学习研究论文集》第162卷），Kamalika Chaudhuri、Stefanie Jegelka、Le Song、Csaba Szepesvári、Gang Niu 和 Sivan Sabato（编）。PMLR，15293–15329。https://proceedings.mlr.press/v162/melnychuk22a.html [116] Jeff T. Mohl、John M. Pearson 和 Jennifer M. Groh。2020年。猴子和人类运用因果推断，同时定位听觉和视觉刺激。《神经生理学杂志》124，3（2020），715–727。https: // doi.org/10.1152/jn.00046.2020 arXiv:https://doi.org/10.1152/jn.00046.2020 PMID: 32727263。[117] Joris M. Mooij、Sara Magliacane 和 Tom Claassen。2020年。基于多个情境的联合因果推断。《机器学习研究杂志》21（2020），99:1–99:108。http://jmlr.org/papers/v21/17-123.html[118] Raha Moraffah、Paras Sheth、Mansooreh Karami、Anchit Bhattacharya、Qianru Wang、Anique Tahir、Adrienne Raglin 和 Huan Liu。2021年。时间序列分析中的因果推断：问题、方法与评估。《知识与信息系统》（2021）63，12，3041–3085。https://doi.org/10.1007/s10115-021-01621-0 [119] 2021 Milan Mossé、Duligur Ibeling 和 Thomas Icard。《因果推理是否比概率推理更难？》CoRR abs/ 2111 13936 2021 2111 13936 2111 13936 。（）。arXiv：。https://arxiv.org/abs/。[120] Ana Rita Nogueira、Andrea Pugnana、Salvatore Ruggieri、Dino Pedreschi 和 João Gama。2022年。因果发现与因果推断的方法和工具。《WIREs数据挖掘与知识发现》12，2（2022）。https://doi.org/10.1002/widm.1449[121] Juan Miguel Ogarrio、Peter Spirtes 和 Joe Ramsey。2016年。针对潜变量模型的混合因果搜索算法。载于《概率图模型——第八届国际会议论文集》，PGM 2016，瑞士卢加诺，2016年9月6日至9日（《JMLR研讨会与会议论文集》第52卷），Alessandro Antonucci、Giorgio Corani 和 Cassio Polpo de Campos（编）。JMLR.org，368–379。http://proceedings.mlr.press/v52/ogarrio16.html [122] Jong-Hoon Oh、Kentaro Torisawa、Canasai Kruengkrai、Ryu Iida 和 Julien Kloetzer。2017年。用于回答"为什么"问题的因果性注意力多列卷积神经网络。载于《第十届ACM国际网络搜索与数据挖掘会议论文集》，WSDM 2017，英国剑桥，2017年2月6日至10日，Maarten de Rijke、Milad Shokouhi、Andrew Tomkins 和 Min Zhang（编）。ACM，415–424。https://doi.org/ 10 1145 3018661 3018737 . / . Milan Palus、Vladimir Komarek、Zbynek Hrncír 和 Katalin Sterbová。[123] 2001年。将同步视为信息速率的调整：从二元时间序列中检测同步。《物理评论E：统计、非线性与软物质物理》63（2001年5月），046211。https://doi.org/10.1103/PhysRevE.63.046211

[124] Angeliki Papana、Catherine Kyrtsou、Dimitris Kugiumtzis 和 Cees Diks。2017年。基于格兰杰因果关系的金融网络：案例研究。《Physica A：统计力学及其应用》482（2017），65–73。[125] Judea Pearl。1989年。《智能系统中的概率推理——合理推断网络》。Morgan Kaufmann。[126] Judea Pearl。1995年。实证研究中的因果图。《Biometrika》82, 4（1995年12月），669–688。https://doi.org/10. 1093/biomet/82.4.669 arXiv:https://academic.oup.com/biomet/article-pdf/82/4/669/698263/82-4-669.pdf [127] 2009 2 10 1017 9780511803161 Judea Pearl。[128] 2014 2, 2 2014 115 129 10 《因果性》（版）。剑桥大学出版社。https://doi.org/ . /CBO Judea Pearl。《因果推断的演绎方法》。（），–。https://doi.org/ .1515/jci- 2014-0016[129] Xingchao Peng、Zijun Huang、Ximeng Sun 和 Kate Saenko。2019年。利用解耦表示进行领域无关学习。载于《第36届国际机器学习会议论文集》，ICML 2019，2019年6月9日至15日，美国加利福尼亚州长滩（《机器学习研究论文集》第97卷），Kamalika Chaudhuri 和 Ruslan Salakhutdinov（编）。PMLR，5102–5112。http://proceedings.mlr.press/v97/peng19b.html [130] Jonas Peters、Dominik Janzing 和 Bernhard Schlkopf。2017年。《因果推断要素：基础与学习算法》。麻省理工学院出版社。[131] Edoardo Maria Ponti、Goran Glavas、Olga Majewska、Qianchu Liu、Ivan Vulic 和 Anna Korhonen。2020年。XCOPA：用于因果常识推理的多语言数据集。载于《2020年自然语言处理实证方法会议论文集》，EMNLP 2020，线上，2020年11月16日至20日，Bonnie Webber、Trevor Cohn、Yulan He 和 Yang Liu（编）。计算语言学协会，2362–2376。https://doi.org/10.18653/v1/2020.emnlp-main.185[132] Hoifung Poon 和 Pedro M. Domingos。2011年。和-积网络：一种新的深度架构。载于UAI 2011，《第二十七届人工智能不确定性会议论文集》，西班牙巴塞罗那，2011年7月14日至17日，Fábio Gagliardi Cozman 和 Avi Pfeffer（编）。AUAI出版社，337–346。https://dslpitt.org/uai/displayArticleDetails. jsp?mmnu =1 =2 _ =2194 _ =27 &smnu &article id &proceeding id Jiaxin Qi、Yulei Niu、Jianqiang Huang 和 Hanwang Zhang。2020年。改进视觉对话的两条因果原则。载于2020年IEEE/CVF计算机视觉与模式识别会议，CVPR，美国华盛顿州西雅图，2020年6月13日至19日。计算机视觉基金会/电气电子工程师学会，10857–10866。https://doi.org/10.1109/CVPR.2020.01087[134] Nasim Rahaman、Muhammad Waleed Gondal、Shruti Joshi、Peter Gehler、Yoshua Bengio、Francesco Locatello 和 Bernhard Schölkopf。2021年。利用神经解释器进行动态推断。（2021）。arXiv:2110.06399 http: //arxiv.org/abs/ 2110 06399 Aditya Ramesh、Prafulla Dhariwal、Alex Nichol、Casey Chu 和 Mark Chen。2022年。利用CLIP潜变量进行分层文本条件图像生成。CoRR abs/2204.06125（2022）。https://doi.org/10.48550/arXiv.2204.06125 arXiv:2204.06125[136] Joseph Ramsey、Madelyn Glymour、Ruben Sanchez-Romero 和 Clark Glymour。2017年。百万个及更多变量：用于学习高维图形因果模型的快速贪心等价搜索算法及其在功能性磁共振图像中的应用。《国际数据科学与分析期刊》3，2（2017），121–129。[137] Joseph D. Ramsey、Jiji Zhang 和 Peter Spirtes。2006年。邻接忠实性与保守因果推断。载于UAI '06，《第22届人工智能不确定性会议论文集》，美国马萨诸塞州剑桥，2006年7月13日至16日。AUAI出版社。https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu =1 =2 _ =1259 _ =22 &smnu &article id &proceeding id [135] [138] 1956 65 Hans Reichenbach。《时间方向》。第卷。加利福尼亚大学出版社。[139] Ronan Riochet、Mario Ynocente Castro、Mathieu Bernard、Adam Lerer、Rob Fergus、Véronique Izard 和 Emmanuel Dupoux。2018年。IntPhys：视觉直觉物理推理的框架与基准。CoRR abs/1803.07616（2018 1803 07616 1803 07616）。arXiv：。http://arxiv.org/abs/。James M. Robins、Andrea Rotnitzky 和 Lue Ping Zhao。1994年。当某些回归变量并非总能被观测到时回归系数的估计。《美国统计学会杂志》89，427（1994），846–866。https://doi.org/10.1080/01621459。[140] 1994 10476818 10 1080 01621459 1994 10476818 。arXiv:https://doi.org/ . / . . Melissa Roemmele、Cosmin Adrian Bejan 和 Andrew S. Gordon。2011年。合理备选项的选择：对常识因果推理的评估。载于《常识推理的逻辑形式化》，2011年AAAI春季研讨会论文集，技术报告SS-11-06，美国加利福尼亚州斯坦福，2011年3月21日至23日。AAAI。http://www.aaai.org/ocs/index.php/SSS/SSS 11 2418 / paper/view/ Tim Rohe 和 Uta Noppeney。皮层层级在多感官知觉中执行贝叶斯因果推断。[141] 2015 [142] 《PLoS Biology》13，2（2015），e1002073。[143] Raanan Y. Rohekar、Shami Nisimov、Yaniv Gurwicz 和 Gal Novik。2021年。在可能存在潜在混杂因素和选择偏差的情况下进行迭代因果发现。载于《神经信息处理系统进展第34卷：2021年神经信息处理系统年会论文集》，NeurIPS 2021，2021年12月6日至14日，线上，Marc'Aurelio Ranzato、Alina Beygelzimer、Yann N. Dauphin、Percy Liang 和 Jennifer Wortman Vaughan（编）。2454–2465。https://proceedings.neurips.cc/paper/2021/hash/144a3f71a03ab7c4f46f9656608efdb2-Abstract.html

[144] Paul R. Rosenbaum 和 Donald B. Rubin。1983年。倾向得分在观察性研究因果效应中的核心作用。Biometrika 70, 1（04 1983年），41–55。https://doi.org/10.1093/biomet/70.1.41 arXiv:https://academic.oup.com/biomet/article-pdf/70/1/41/662954/70-1-41.pdf [145] Dominik Rothenhäusler、Christina Heinze、Jonas Peters 和 Nicolai Meinshausen。2015年。BACKSHIFT：从未知移位干预中学习因果循环图。载于《神经信息处理系统进展》第28卷：2015年神经信息处理系统年度会议，2015年12月7日至12日，加拿大魁北克省蒙特利尔，Corinna Cortes、Neil D. Lawrence、Daniel D. Lee、Masashi Sugiyama 和 Roman Garnett（编）。1513–1521。https: // proceedings.neurips.cc/paper/2015/hash/92262bf907af914b95a0fc33c3f33bf6-Abstract.html [146] Donald B Rubin。1974年。随机研究与非随机研究中治疗的因果效应估计。《教育心理学杂志》66, 5（1974年），688。[147] Mikail Rubinov 和 Olaf Sporns。2010年。脑连接的复杂网络测度：用途与解释。NeuroImage 52, 3（2010年），1059–1069。[148] Jakob Runge、Peer Nowack、Marlene Kretschmer、Seth Flaxman 和 Dino Sejdinovic。2019年。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（2019年），eaau4996。https://doi.org/ 10.1126/sciadv.aau4996 arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.aau4996[149] Chitwan Saharia、William Chan、Saurabh Saxena、Lala Li、Jay Whang、Emily Denton、Seyed Kamyar Seyed Ghasemipour、Burcu Karagol Ayan、S. Sara Mahdavi、Rapha Gontijo Lopes、Tim Salimans、Jonathan Ho、David J. Fleet 和 Mohammad Norouzi。2022年。基于深度语言理解的逼真文本到图像扩散模型。CoRR abs/ 2205 11487 2022 10 48550 2205 11487 2205 11487 。（）。[150] https://doi.org/ . /arXiv. . arXiv: . Francesco Sambo、Barbara Di Camillo 和 Gianna Toffolo。。CNET：因果基因网络逆向工程算法。NETTAB2008，意大利瓦伦纳（2008年）。[151] Raquel Sánchez-Cauce、Iago Paris 和 Francisco Javier Diez。2022年。和-积网络：综述。IEEE Trans. Pattern Anal. Mach. Intell. 44, 7 2022 3821 3839 10 1109 2021 3062898（），–。[152] https://doi.org/ . /TPAMI. . Pablo Sánchez-Martin、Miriam Rateike 和 Isabel Valera。。VACA：为因果查询设计变分图自编码器。载于第三十六届AAAI人工智能大会、AAAI 2022，第三十四届人工智能创新应用大会、IAAI 2022，第十二届人工智能教育进展研讨会、EAAI 2022线上会议，2022年2月22日至3月1日。AAAI出版社，8159–8168。https: //ojs.aaai.org/index.php/AAAI/article/view/[153] Franco Scarselli、Marco Gori、Ah Chung Tsoi、Markus Hagenbuchner 和 Gabriele Monfardini。2009年。图神经网络模型。IEEE Trans. Neural Networks 20, 1（2009年），61–80。https://doi.org/10.1109/TNN.2008.2005605[154] Eric E Schadt、John Lamb、Xia Yang、Jun Zhu、Steve Edwards、Debraj GuhaThakurta、Solveig K Sieberts、Stephanie Monks、Marc Reitman、Chunsheng Zhang 等。2005年。通过整合基因组学方法推断基因表达与疾病之间的因果关联。Nature Genetics 37, 7（2005年），710–717。[155] Nino Scherrer、Anirudh Goyal、Stefan Bauer、Yoshua Bengio 和 Nan Rosemary Ke。2022年。论因果模型的泛化与适应性能。CoRR abs/2206.04620（2022年）。https://doi.org/10.48550/arXiv.2206.04620 arXiv:2206.04620[156] Jürgen Schmidhuber。2015年。论思维学习：用于新型强化学习控制器组合与循环神经世界模型的算法信息论。CoRR abs/1511.09249（2015年）。arXiv:1511.09249 http://arxiv.org/abs/1511.09249[157] Bernhard Schölkopf、Francesco Locatello、Stefan Bauer、Nan Rosemary Ke、Nal Kalchbrenner、Anirudh Goyal 和 Yoshua Bengio。2021年。迈向因果表征学习。Proc. IEEE 109, 5（2021年），612–634。https://doi.org/10. 1109/JPROC.2021.3058954[158] Bernhard Schölkopf 和 Julius von Kügelgen。2022年。从统计学习到因果学习。CoRR abs/2204.00607（2022年）。https://doi.org/ 10 48550 2204 00607 2204 00607 . /arXiv. . arXiv: . Thomas Schreiber。。测量信息传递。[159]《物理评论快报》85（7 2000年），461–464。第2期。https: //doi.org/ 10 1103 85 461 . / 2000 PhysRevLett. . Anil K Seth。。行为期间进化神经网络的因果连接。Network: Computation in Neural Systems 16, 1（2005年），35–54。[161] Anil K Seth、Adam B Barrett 和 Lionel Barnett。2015年。神经科学与神经影像学中的格兰杰因果分析。Journal of Neuroscience 35, 8（2015年），3293–3297。[162] Ladan Shams 和 Ulrik R Beierholm。2010年。感知中的因果推断。Trends in Cognitive Sciences 14, 9（2010年），425–432。[163] Claude E. Shannon。1948年。通信的数学理论。Bell Syst. Tech. J. 27, 3（1948年），379–423。https://doi.org/10.1002/j.1538-7305.1948.tb01338.x [164] Shohei Shimizu、Patrik O. Hoyer、Aapo Hyvärinen 和 Antti J. Kerminen。2006年。用于因果发现的线性非高斯无环模型。《机器学习研究杂志》7（2006年），2003–2030。http://jmlr.org/papers/v7/shimizu06a.html [160] 2005

[165] Bill Shipley。1999年。生物个体生物学中的因果解释检验：因果关系、相关性与结构方程建模。Oikos（1999年），374–382。[166] Christopher A Sims。1972年。货币、收入与因果性。《美国经济评论》62, 4（1972年），540–552。[167] Peter Spirtes、Clark Glymour 和 Richard Scheines。1989年。从概率推断因果性。（1989年）。[168] Peter Spirtes、Clark Glymour 和 Richard Scheines。2000年。《因果关系、预测与搜索》，第二版。麻省理工学院出版社。[169] Jerzy Splawa-Neyman、D. M. Dabrowska 和 T. P. Speed。1990年。概率论在农业实验中的应用。原理论文。第9节。5, 4（1990年），465–472。https://doi.org/10.1214/ss/1177012031 出版者：数理统计学会。[170] Elizabeth A Stuart。2010年。因果推断的匹配方法：综述与展望。《统计科学》：数理统计学会评论期刊25, 1（2010年），1。[171] George Sugihara、Robert May、Hao Ye、Chih-hao Hsieh、Ethan Deyle、Michael Fogarty 和 Stephan Munch。2012年。在复杂生态系统中检测因果性。《科学》338, 6106（2012年），496–500。[172] Yuan Sui、Shanshan Feng、Huaxiang Zhang、Jian Cao、Liang Hu 和 Nengjun Zhu。2022年。面向知识图谱多跳问答的因果性感知增强模型。《知识系统》250（2022年），108943。https: //doi.org/ 10 1016 2022 108943 . /j.knosys. . Jie Sun、Dane Taylor 和 Erik M. Bollt。[173] 2015年。通过最优因果熵进行因果网络推断。《SIAM应用动力系统杂志》14, 1 2015 73 106 10 1137 140956166 （），–。https://doi.org/ . / Robert E. Tillman 和 [174] Peter Spirtes。 。从多个具有重叠变量的数据集中学习带有潜变量和选择变量的无环模型等价类。载于第十四届人工智能与统计国际会议论文集，AISTATS 2011，2011年4月11日至13日，美国劳德代尔堡（《机器学习研究杂志》会议论文集，第15卷），Geoffrey J. Gordon、David B. Dunson 和 Miroslav Dudik（编）。JMLR.org，3–15。2011 http://proceedings.mlr.press/v15/ tillman11a/tillman11a.pdf [175] William Timberlake。1997年。以动物为中心的因果系统方法：理解和控制行为。《应用动物行为科学》53, 1（1997年），107–129。https://doi.org/10.1016/S0168-1591(96)01154-9 动机与认知的基础和应用方面。[176] Naftali Tishby、Fernando C. N. Pereira 和 William Bialek。2000年。信息瓶颈方法。CoRR physics/ 0004057 2000 0004057 （）。http://arxiv.org/abs/physics/ Sofia Triantafillou 和 [177] Ioannis Tsamardinos。 。基于约束的因果发现：来自多个干预且变量集相互重叠的情形。《机器学习研究杂志》2015 16 2015 2147 2205 10 5555 2789272 2886819 （），–。[178] https://doi.org/ . / . Ioannis Tsamardinos、Laura E. Brown 和 Constantin F. Aliferis。 。最大最小爬山贝叶斯网络结构学习算法。《机器学习》2006 65, 1 2006 31 78 10 1007 10994 006 6889 7 （），–。[179] https://doi.org/ . /s - - - Pulakesh Upadhyaya、Kai Zhang、Can Li、Xiaoqian Jiang 和 Yejin Kim。 。可扩展因果结构学习：生物医学领域的新机遇。arXiv预印本 arXiv:2110.07785（2021年）。[180] Vasily A. Vakorin、Olga A. Krakovska 和 Anthony R. McIntosh。2009年。间接连接的混杂效应对因果性估计的影响。《神经科学方法杂志》184, 1（2009年），152–160。https://doi.org/10.1016/j.jneumeth. 2009.07.014[181] Sjoerd van Steenkiste、Francesco Locatello、Jürgen Schmidhuber 和 Olivier Bachem。2019年。解耦表示是否有助于抽象视觉推理？载于《神经信息处理系统进展》第32卷：2019年神经信息处理系统年度会议，NeurIPS 2019，2019年12月8日至14日，加拿大不列颠哥伦比亚省温哥华，Hanna M. Wallach、Hugo Larochelle、Alina Beygelzimer、Florence d’Alché-Buc、Emily B. Fox 和 Roman Garnett（编）。14222–14235。https://proceedings.neurips.cc/paper/2019/hash/bc3c4a6331a8a9950945a1aa8c95ab8a-Abstract.html[182] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N. Gomez、Lukasz Kaiser 和 Illia Polosukhin。2017年。注意力就是一切。载于《神经信息处理系统进展》第30卷：2017年神经信息处理系统年度会议，2017年12月4日至9日，美国加利福尼亚州长滩，Isabelle Guyon、Ulrike von Luxburg、Samy Bengio、Hanna M. Wallach、Rob Fergus、S. V. N. Vishwanathan 和 Roman Garnett（编）。5998–6008。https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html [183] Thomas Verma 和 Judea Pearl。1990年。因果模型的等价与综合。载于UAI ’90：第六届人工智能不确定性年度会议论文集，美国马萨诸塞州剑桥麻省理工学院，1990年7月27日至29日，Piero P. Bonissone、Max Henrion、Laveen N. Kanal 和 John F. Lemmer（编）。爱思唯尔，255–270。https://dslpitt.org/uai/ displayArticleDetails.jsp?mmnu =1 =2 _ =1918 _ =1006 &smnu &article id &proceeding id [184] Vikas Verma、Alex Lamb、Christopher Beckham、Amir Najafi、Ioannis Mitliagkas、David Lopez-Paz 和 Yoshua Bengio。2019年。流形混合：通过插值隐藏状态获得更优表征。载于第三十六届国际机器学习会议论文集，ICML 2019，2019年6月9日至15日，美国加利福尼亚州长滩（《机器学习研究论文集》，第97卷），Kamalika Chaudhuri 和 Ruslan Salakhutdinov（编）。PMLR，6438–6447。http://proceedings.mlr.press/v97/verma19a.html

[185] Ioannis Vlachos 和 Dimitris Kugiumtzis。2010年。非均匀状态空间重构与耦合检测。《物理评论E》 82 7 2010 016207 1 10 1103 82 016207 ( )，。第期。https://doi.org/ 。/PhysRevE. 。
[186] Charles Vorbach、Ramin M. Hasani、Alexander Amini、Mathias Lechner 和 Daniela Rus。2021年。通过连续时间神经网络进行因果导航。载于《神经信息处理系统进展第34卷》：神经信息处理系统进展2021年年度会议，NeurIPS 2021，2021年12月6日至14日，线上会议，Marc'Aurelio Ranzato、Alina Beygelzimer、Yann N. Dauphin、Percy Liang 和 Jennifer Wortman Vaughan（编）。12425–12440。https: // proceedings.neurips.cc/paper/2021/hash/67ba02d73c54f0b83c05507b7fb7267f-Abstract.html [187] Matthew J. Vowels、Necati Cihan Camgöz 和 Richard Bowden。2023年。你喜欢有向无环图吗？关于结构学习与因果发现的综述。《ACM计算调查》 55, 4 2023 82 1 82 36 10 1145 3527154 ( )，：–：。
[188] https://doi.org/ 。/ Christoph Völter 和 Josep Call。 2017 。动物中的因果与推断推理。–。https://doi.org/ .1037/ 643 671 0000012-029 10 [189] Anna Waismeyer、Andrew N Meltzoff 和 Alison Gopnik。2015年。24个月大儿童从概率事件中进行因果学习：一种行动测量。《发展科学》18，1（2015），175–182。[190] Lan Wang 和 Vishnu Naresh Boddeti。2022年。学习到的表征是否遵循因果关系? CoRR abs/ 2204 00762 2022 10 48550 2204 00762 2204 00762 。（）。https://doi.org/ 。/arXiv. 。
[191] arXiv: 。William Whitney。 2016 。神经模型中的解耦表示。CoRR abs/1602.02383（2016）。arXiv:1602.02383 http://arxiv.org/abs/ 1602 02383 。Marc Wiedermann、Alexander Radebach、Jonathan F Donges、
[192] Jürgen Kurths 和 Reik V Donner。2016年。基于气候网络的指数：区分不同类型的厄尔尼诺和拉尼娜。《地球物理研究快报》43，13（2016），7176–7185。[193] James Woodward。2005年。促成事物发生：因果解释理论。牛津大学出版社。[194] Sewall Wright。1921年。相关性与因果关系。（1921年）。[195] Aming Wu、Rui Liu、Yahong Han、Linchao Zhu 和 Yi Yang。2021年。用于领域不变目标检测的向量分解解耦。载于2021年IEEE/CVF国际计算机视觉大会，ICCV 2021，加拿大魁北克省蒙特利尔，10月 *10 17, 2021* 9322 9331 10 1109 48922 2021 00921 -。IEEE，–。https://doi.org/ 。/
[196] ICCV 。 2021 。Pengzhou Wu 和 Kenji Fukumizu。 。通过深度可识别模型实现有原则的因果效应估计。CoRR abs/ 2109 15062 2021 2109 15062 2109 15062 。（）。arXiv: 。https://arxiv.org/abs/ 。Kevin
[197] Xia、Kai-Zhan Lee、Yoshua Bengio 和 Elias Bareinboim。 2021 。因果—神经联系：表达能力、可学习性与推断。载于《神经信息处理系统进展第34卷》：神经信息处理系统进展2021年年度会议，NeurIPS 2021，2021年12月6日至14日，线上会议，Marc' Aurelio Ran-zato、Alina Beygelzimer、Yann N. Dauphin、Percy Liang 和 Jennifer Wortman Vaughan（编）。10823–10836。https://proceedings.neurips.cc/paper/2021/hash/5989add1703e4b0480f75e2390739f34-Abstract.html [198] Sang Michael Xie、Aditi Raghunathan、Percy Liang 和 Tengyu Ma。2022年。将上下文学习解释为隐式贝叶斯推断。载于第十届国际学习表征会议，ICLR 2022，线上会议，4月25日至29日，2022年。OpenReview.net。https://openreview.net/forum?id= RdJVFCHjUMI [199] Kazuko Yamasaki、Avi Gozolchiani 和 Shlomo Havlin。2008年。全球气候网络显著受到厄尔尼诺影响。《物理评论快报》100，22（2008），228501。[200] Xu Yang、Hanwang Zhang、Guojun Qi 和 Jianfei Cai。2021年。面向视觉语言任务的因果注意力。载于IEEE计算机视觉与模式识别会议，CVPR 2021，线上会议，2021年6月19日至25日。计算机视觉基金会/ IEEE, 9847 9857 10 1109 46437 2021 00972 –。https://doi.org/ 。/CVPR 。 。
[201] Can-Zhong Yao、Ji-Nan Lin、Qing-Wen Lin、Xu-Zhou Zheng 和 Xiao-Feng Liu。2016年。基于格兰杰网络的工业用电量因果结构与动力学研究。Physica A: Statistical Mechanics and its Applications 462（2016），297–320。[202] Liuyi Yao、Zhixuan Chu、Sheng Li、Yaliang Li、Jing Gao 和 Aidong Zhang。2021年。因果推断综述。《ACM知识发现与数据挖掘汇刊》 15, 5 2021 74 1 74 46 10 1145 3444944 ( )，：–：。https://doi.org/ 。/ Kexin Yi、
[203] Chuang Gan、Yunzhu Li、Pushmeet Kohli、Jiajun Wu、Antonio Torralba 和 Joshua B. Tenenbaum。2020年。CLEVRER: 用于视频表征与推理的碰撞事件。载于第八届国际学习表征会议，ICLR 2020，埃塞俄比亚亚的斯亚贝巴，2020年4月26日至30日。OpenReview.net。https://openreview.net/forum? id=HkxYzANYDB[204] Jing Yu、V Anne Smith、Paul P Wang、Alexander J Hartemink 和 Erich D Jarvis。2004年。改进贝叶斯网络推断，从观察性生物数据生成因果网络。《生物信息学》20，18（2004），3594–3603。[205] Zhongqi Yue、Tan Wang、Qianru Sun、Xian-Sheng Hua 和 Hanwang Zhang。2021年。反事实零样本与开放集视觉识别。载于IEEE计算机视觉与模式识别会议，CVPR 2021，线上会议，2021年6月 *19 25, 2021* 15404 15414 10 1109 46437 2021 01515 -。计算机视觉基金
[206] 会/IEEE，–。https://doi.org/ 。/CVPR 。 。Zhongqi Yue、Hanwang Zhang、Qianru Sun 和 Xian-Sheng Hua。 2020 。干预式少样本学习。载于《神经信息处理系统进展第33卷》：神经信息处理系统进展2020年年度会议，NeurIPS 2020，2020年12月6日至12日，线上会议，Hugo Larochelle、Marc'Aurelio Ranzato、Raia Hadsell、Maria-Florina Balcan，

以及 Hsuan-Tien Lin（编）。https://proceedings.neurips.cc/paper/2020/hash/1cc8a8ea51cd0adddf5dab504a285915-Abstract.html[207] Alessio Zanga、Elif Ozkirimli 和 Fabio Stella。2022年。因果发现综述：理论与实践。《国际近似推理杂志》 151 2022 101 129 10 1016 2022 09 004 ( ), –。https://doi.org/ 。/J.IJAR. 。 。Matej [208] Zecevic、Devendra Singh Dhami 和 Kristian Kersting。 。神经因果推断的可处理性。CoRR abs/ 2110 12052 2021 2110 12052 2110 12052 。（ ）。arXiv: 。https://arxiv.org/abs/ 。Matej [209] Zecevic、Devendra Singh Dhami、Petar Velickovic 和 Kristian Kersting。2021年。将图神经网络与结构因果模型联系起来。CoRR abs/ 2109 04173 2021 2109 04173 2109 04173 。（ ）。 [210] arXiv: 。https://arxiv.org/abs/ 。Matej Zecevic、Moritz Willig、Devendra Singh Dhami 和 Kristian Kersting。 。因果鹦鹉：大语言模型或许能谈论因果性，但并非因果的。CoRR abs/2308.13067（2023年）。https://doi.org/10.48550/ARXIV.2308. 13067 arXiv:2308.13067[211] Chi Zhang、Baoxiong Jia、Mark Edmonds、Song-Chun Zhu 和 Yixin Zhu。2021年。ACRE：超越协变性的抽象因果推理。载于IEEE计算机视觉与模式识别会议，CVPR 2021，25，2021 19 10643 10653 10 1109 46437 2021virtual，2021年6月-。计算机视觉基金会/IEEE，–。https://doi.org/ 。/CVPR 。 。 01050 [212] 2018 Hongyi Zhang、

Moustapha Cissé、Yann N. Dauphin 和 David Lopez-Paz。 。mixup：超越经验风险最小化。载于第六届国际学习表征会议，ICLR 2018，加拿大不列颠哥伦比亚省温哥华，2018年4月30日至5月3日，会议论文集。OpenReview.net。 https://openreview.net/forum?id=r1Ddp1-Rb [213] Hanlin Zhang、Yi-Fan Zhang、Weiyang Liu、Adrian Weller、Bernhard Schölkopf 和 Eric P. Xing。2021年。面向领域泛化的有原则解耦。CoRR abs/2111.13839（2021年）。arXiv:2111.13839 https: // arxiv.org/abs/ 2111 13839 。Kun Zhang 和 Aapo Hyvärinen。2009年。后非线性因果模型的可识别性。载于 [214] UAI 2009，即第二十五届不确定性人工智能会议论文集，加拿大魁北克省蒙特利尔，2009年6月18日至21日，647 655 - Jeff A. Bilmes 和 Andrew Y. Ng（编）。AUAI出版社，–。https://dslpitt.org/uai/displayArticleDetails.jsp? mmnu =1 =2 _ =1606 _ =25 &smnu &article id &proceeding id Kun Zhang 和 Aapo Hyvärinen。 。 [215] 2010 使用非线性无环因果模型区分原因与结果。载于《因果性：目标与评估》。PMLR，157–164。[216] Xun Zheng、Bryon Aragam、Pradeep Ravikumar 和 Eric P. Xing。2018年。无撕裂DAG：用于结构学习的连续优化。载于《神经信息处理系统进展》第31卷：神经信息处理系统进展2018年年度会议，NeurIPS 2018，2018年12月3日至8日，加拿大蒙特利尔，Samy Bengio、Hanna M. Wallach、Hugo Larochelle、Kristen Grauman、Nicolò Cesa-Bianchi 和 Roman Garnett（编）。9492–9503。 https://proceedings.neurips.cc/paper/2018/hash/e347c51419ffb23ca3fd5050202f9c3d-Abstract.html

表3. 因果性的应用研究综述。

领域	研究问题	文章
动物行为 A.1	动物中的原因归因	[116, 175, 188]
	动物的因果模型	[20, 51, 103]
自动化推理 A.2	图像识别	[133, 200, 205, 206]
	问答	[122, 172, 211]
	现实世界导航	[30, 186]
	因果常识推理	[19, 33, 56, 97, 131, 141]
生物学 A.3	生态系统建模	[94, 171, 194]
	基因表达的因果解释	[32, 74, 150, 154, 204]
	发现连接器官的共同原因-ism 表型	[5, 165]
气候学 A.4	气候机制建模	[22, 38, 108, 192, 199]
经济学 A.5	金融趋势预测	[13, 124]
	经济建模	[59, 67, 166, 201]
健康A.6	治疗效应估计	[71, 146, 179]
	《流行病学》	[61, 70]
神经科学 A.7	脑中因果推断的兴起	[47, 91, 142, 162]
	大脑的因果建模	[64]
	神经系统的连通性	[45, 147, 160, 161]
《物理学》 A.8	物理机制建模	[9, 35, 58, 139, 203]
社会科学 A.9	人类的原因归因	[16, 54, 116, 189]
	大脑的因果建模	[31, 107]

A 更多应用

因果性领域的发展与科学的发展密切相关。因果性的早期出现之一可追溯到亚里士多德所著的《物理学》第二卷 [42]。因此，因果性在科学的众多广泛领域中都有应用，从生物学到经济学。本节将简要概述运用因果性方法的领域。表3 总结了我们讨论的主要文章。

A.1 动物行为

大量研究通过实验探讨动物的因果推理。这些研究并未提供用于发现因果关系的通用框架，而是逐案分析，或研究人类和动物的特定能力 [116, 175, 188]。少数文章运用第2 节介绍的因果关系理论，来发现动物行为的结构。Lord 等人[103] 采用一种基于传递熵（最优因果熵）的方法，对蠓群中的信息流进行建模。类似地，Chen 等人 [20] 运用因果熵推断鸽子的运动。Gendron 等人 [51] 将针对时间序列的因果发现与图神经网络相结合，为狐獴行为构建结构因果模型。

A.2 自动推理

如前几节所述，因果性可用于提升深度学习模型的性能。一些文章将因果关系理论应用于视觉任务，以提升模型性能：例如视觉对话 [133], 、少样本和零样本设置下的图像识别 [205, 206], , 或视觉问答 [200]。

深度学习中的许多问答任务都依赖文本理解、记忆或常识。有些任务还需要复杂的推理或逻辑结构。因果关系理论有助于应对这些任务。Oh 等人 [122]将因果注意力用于原因问答，而 Sui 等人[172] 则使用因果滤波器来减少知识图谱问答（KGQA）任务中的伪相关数量。ACRE数据集 [211] 是一项需要检索因果机制的视觉推理任务。

因果推理的另一个应用场景是自动驾驶和导航。de Haan 等人 [30] 研究了控制游戏和自动驾驶汽车中用于模仿学习的因果归因。Vorbach 等人 [186] 尝试使用液态时间常数网络（LTCs）来表征无人机自主导航中的环境因果结构。

因果常识推理旨在解决一种更接近人类直觉理解的因果推理问题。由于解决这些查询所需的信息并未明确给出，因此很难为这一任务生成因果图。SemEval-2010 Task 7 [56] 是一个用于从句子中抽取原因与结果的评测数据集。尽管该数据集发表于十年前，它仍是为数不多的因果常识推理基准测试之一。另一个值得一提的数据集是 COPA [141], , 它由一组前提以及可能的原因或结果组成。受测系统需要运用常识找出正确答案。XCOPA 是 COPA [131]的多语言扩展。CausalBank数据集 [97] 是一个因果句语料库，每个句子都基于一张因果图。Event StoryLine [19] 数据集包含新闻文档，并为信息抽取和链接任务奠定了基础。e-CARE数据集 [33] 引入了一组因果查询：一个原因和多个可能的结果句，并为正确答案提供因果解释。它与 COPA 的主要区别在于包含因果解释。

A.3 生物学

因果性在生物学中的应用多种多样。首先，因果性有助于对生态系统等动力系统进行建模 [94, 171, 194]。另一项影响深远的应用，是将个体的基因表达归因于某种表型属性。一些研究为基因表达或其他生物通路构建调控网络模型 [32, 74, 150, 204]。另一些研究则尝试将基因表达与肥胖等疾病的出现联系起来 [154]。第三，建立表型各种性状之间的因果联系，有助于了解物种的进化过程并预测物种分布 [5, 165]。

A.4 气候学

因果性在气候学中的主要用途，是将气候事件建模为网络。例如，Wiedermann等人 [192], Yamasaki等人 [199] 提出了用于研究厄尔尼诺影响的模型。厄尔尼诺是太平洋中暖水的周期性流动，对该地区各国的气温和降雨有重要影响。Malik等人 [108] 利用网络表征研究印度的降雨，而Ludescher等人 [105] 则利用网络预测极端天气事件。气候学任务颇具挑战，因为涉及许多因子，且存在复杂的非线性交互。一些研究指出，准确归因原因十分困难，基于相关性的方法也存在局限 [22, 38]。

A.5 经济学

早期针对时间序列的因果关系理论源于经济应用 [59, 67, 166],，尤其是计量经济学。计量经济学是指运用统计方法分析经济数据，以揭示数据中的关系。这些研究试图运用因果性提取关系，并构建理论经济模型。还有一些研究继续运用因果方法探讨经济学的理论基础 [201]。另一个重要应用

是金融领域的因果性，以及对股票市场趋势和危机的预测。为此，常用的方法基于格兰杰因果关系或熵的度量 [13, 124]。

A.6 健康

由于格兰杰框架最初是为经济领域开发的，Rubin因果模型则是为医学应用而创建的 [146], ，尤其用于研究随机试验的影响。因果推理能够减少此类研究中的偏差和混杂效应，研究人员认为，应更广泛地采用因果方法，以减少导致错误结论的伪相关 [71]。Rubin 模型广泛应用于医学领域，但并非唯一的模型。Upadhyaya 等人 [179]概述了因果结构发现技术在医疗保健领域的应用。流行病学中也存在因果模型，用于识别传播疾病的协变量和混杂因素，并对人群中的流行病进行建模 [61, 70]。

A.7 神经科学

大量研究致力于探究人脑如何整合来自感觉输入的信息，尤其是人脑如何进行因果推断 [47, 91, 142, 162]。另一些研究试图在人脑结构中发现因果模式，并据此构建因果模型 [64]。连接组是对大脑神经连接的绘制，而连接组学则是研究连接组的领域。了解连接组中发生的神经机制，有助于加深对大脑的认识。已有研究尝试基于格兰杰因果关系开展因果连接分析 [45, 147, 160]。在神经影像学领域，研究连接组的变化有助于深入了解认知和个体行为。一些研究利用时间序列中的因果网络对这些变化进行建模 [161]。

A.8 《物理学》

因果方法可用于模拟世界的物理机制。循环独立机制 [58]利用深度学习来预测物理对象的动力学。研究者提出了基准测试，用于训练系统进行物理推理，例如预测碰撞力学 [9, 139, 203]。这些方法旨在回答干预或反事实查询。大多数这类方法试图复现人类进行直觉物理推理的方式 [35]。

A.9 社会科学

与动物研究类似，一些研究通过实验考察人类进行因果推断的能力。Mohl 等人 [116] 比较了人类和猴子对刺激位置的判断，Bramley 等人[16] 研究了时间如何影响因果归因，而最后一类研究 [54, 189] 则探讨儿童如何学习进行因果推理。另一方面，其他研究领域试图更深入地理解人脑的一些因果行为，尤其是与感知有关的行为[31, 107]。