

贝叶斯网络结构学习综述

孟光磊^{1,2,*}, 丛泽林^{1,2}, 宋彬^{1,2}, 李婷琰², 王晨光², 周铭哲^{1,2}

(1. 沈阳航空航天大学 自动化学院, 沈阳 110136; 2. 航空科学技术空战系统技术重点实验室, 沈阳 110136)

摘 要: 贝叶斯网络作为概率论与图论结合的工具, 具备高效处理不确定性推理和数据分析的能力, 被广泛应用于各领域解决复杂工程问题。此外, 还可以结合先验知识和训练样本学习模型, 克服了单纯依靠专家知识建立模型的局限性。基于此, 回顾了贝叶斯网络的发展历程, 分别从基于约束的方法、基于评分搜索的方法、混合约束和评分搜索的方法 3 个方面对已提出的贝叶斯网络结构学习方法进行分类归纳, 并对各类方法研究的现状进行了总结分析。由于现实应用中的数据往往具有非完备性, 从缺失数据处理和隐变量学习 2 个维度阐释了非完备贝叶斯网络结构学习的研究现状。对贝叶斯网络在不同领域中的应用情况进行阐述, 并进行总结, 讨论了未来贝叶斯网络结构学习方法研究的发展趋势。

关 键 词: 机器学习; 人工智能算法; 贝叶斯网络; 结构学习; 隐变量

中图分类号: TP181

文献标志码: A **文章编号:** 1001-5965(2025)09-2829-21

数据挖掘^[1]技术作为一门新兴学科, 融合了当今时代先进的计算机技术、人工智能技术和概率统计学技术, 其对采集到的数据进行自主分析和归纳推理, 发现潜在的数据模型, 从大量的数据中提取有效的、具有潜在价值的信息, 以达到帮助决策者做出正确决定、降低风险的目的。贝叶斯网络(Bayesian network, BN)算法相比于其他决策模型, 将概率推理和图形模型有机结合, 对多元不确定性知识具有可视化的模型推理能力, 已经被证实为是一种可进行不确定性问题推理分析的有效算法模型。

条件独立的随机变量之间的定性和定量关系是贝叶斯网络学习的基本内容。定性内容指描述变量间因果关系的网络结构; 定量内容则主要指描述变量间依赖程度的条件概率分布。在变量结构模型已知的条件下, 通过已知变量间的定性关系来分析各变量间的定量关系, 即从数据样本中学习得到节点间的条件概率分布的过程被称为贝叶斯网

络参数学习; 当变量结构模型未知时, 贝叶斯网络学习同时揭示各变量间的定性和定量关系, 该过程被称为结构学习。军事、医疗、金融、工业生产等很多领域中都积累了海量的数据, 贝叶斯网络学习为用户挖掘数据中潜在的规律, 找出相对于数据集拟合度最高的网络结构模型辅助相关领域决策, 提供了有效的技术途径。

目前, 贝叶斯网络结构学习方法大体可以分为 3 种类型: 基于约束的方法、基于评分搜索的方法及混合约束和评分搜索的方法。基于约束的方法是最早提出的贝叶斯网络结构学习方法, 利用条件独立性(conditional independence, CI)测试^[2]来学习贝叶斯网络结构, 尽可能多地去除变量之间多余的依赖关系, 如 PC 算法^[3-4]、三阶段依赖分析算法(three-phase dependency analysis algorithm, TPDA)^[5]等。由于需要大量样本来测试独立性, 基于约束的方法时间代价较大, 且有时不可靠的 CI 测试会使

收稿日期: 2023-07-07; 录用日期: 2023-08-04; 网络出版时间: 2023-09-05 10:36
网络出版地址: link.cnki.net/urlid/11.2625.V.20230904.1958.001
基金项目: 国家自然科学基金(61973222); 辽宁省“兴辽英才计划”(XLYC2007144); 沈阳市自然科学基金(22-315-6-09)
*通信作者. E-mail: menguanglei@yeah.net

引用格式: 孟光磊, 丛泽林, 宋彬, 等. 贝叶斯网络结构学习综述 [J]. 北京航空航天大学学报, 2025, 51(9): 2829-2849.
MENG G L, CONG Z L, SONG B, et al. A review of Bayesian network structure learning [J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(9): 2829-2849 (in Chinese).

输入级的错误被放大传导到输出级,导致网络推理结果准确性降低。基于评分搜索的方法应用较为广泛,其可以灵活地将专家知识^[6]作为结构先验引入搜索学习过程,利用评分函数评估每个可能的网络结构,如 K2 算法^[7]、BIC 算法^[8]等。随着节点数量增加,基于评分搜索的方法学习网络结构空间数量呈指数级增长,使得最优结构的搜索代价过高。因此,基于评分搜索的方法一般采用启发式策略^[9]来减少搜索空间,但这种方式有时会导致网络陷入局部最优,相对而言,更具有自适应性的基于自然选择的群体智能 (swarm intelligence, SI)^[10] 优化算法成为近年来的研究热点。此外,研究学者也提出了一系列混合方法用以解决基于约束的方法和基于评分搜索的方法的缺陷问题。例如,通过低阶 CI 测试得出节点的排序,减少后续结构搜索代价,再使用 K2 算法学习网络,搜索所有可能网络中的最优结构。

随着贝叶斯网络算法不断改进,数据的处理能力也随之提高,逐步适用于各种复杂技术领域,并发挥重要作用。本文梳理了贝叶斯网络结构学习方法的发展脉络,对现有成果进行总结,为研究人员进行文献追踪、迅速跟进该领域的研究前沿提供参考。首先,简单介绍了贝叶斯网络的发展背景和原理;其次,对贝叶斯网络结构学习方法进行分类,并进行系统性的总结,针对现实应用中变量(如数据缺失)的情况,对非完备信息下贝叶斯网络结构学习进行详细分类阐述;然后,梳理总结了贝叶斯网络结构学习方法在社会各领域的应用情况;最后,总结了未来贝叶斯网络结构学习理论研究的发展趋势。

1 贝叶斯网络

1.1 背景介绍

贝叶斯网络又称信念网络,由图灵奖获得者 Judea Pearl 提出^[11],他也因此被誉为“贝叶斯网络之父”。贝叶斯网络是一种用于计算复杂逻辑、推理因果关系中模糊概率的图模型,能够表示随机变量及变量间的依赖关系,其以因子分解的方式定义了联合概率分布的数据结构,并给出了分布中的一系列 CI 假设。

贝叶斯网络拓扑结构是一个有向无环图 (directed acyclic graph, DAG)。其中,节点表示随机变量,弧表示它们之间的条件概率依赖关系。“有向”指的是用来连接不同节点的边的方向是固定的,起点和终点顺序不能调换,说明由因到果的逻辑关系不能颠倒;“无环”指的是从任意节点出发,

无法经过若干条边再次回到该点,即在图中不存在任何环路。文献 [12] 强调应按照因果关系来决定变量顺序。非条件独立的 2 个随机变量节点之间连接的起始节点为“因”,终点节点为“果”。有向图的结构可以模拟建模域的因果结构,当结构是因果关系时,其为变量之间的相互作用提供了可视化信息,并能预测外部操纵带来的影响。

1.2 数学模型

贝叶斯网络 N 由结构 G 和网络参数 θ 表示为二元组 $N = (G, \theta)$ 。在图论中,可以令有向无环结构图 $G = (X, R)$ 表示一组变量之间的独立关系, X 为图形中所有节点的非空集合 $X = \{X_1, X_2, \dots, X_i, \dots, X_n\}$, X_i 可以为观测变量,也可以为隐变量, R 为 DAG 中不同变量之间有向连接线段的集合, $X_j \rightarrow X_i$ 表示节点之间的直接依赖关系。

$$R = \{X_j \rightarrow X_i | X_j \in \text{pa}(X_i), i = 1, \dots, n\} \quad (1)$$

式中: $\text{pa}(X_i)$ 表示节点 X_i 的“因”,也称为父节点集合。根据马尔可夫独立性,给定父节点集 $\text{pa}(X_i)$, X_i 独立于其非子代节点集 $\text{nd}(X_i)$ 。因此,若干遵循马尔可夫规则的节点 X_i 的联合概率分布可以表示为

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{pa}(X_i)) \quad (2)$$

网络参数 $\theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ 表示每个节点 X_i 在其父节点集 $\text{pa}(X_i)$ 已知时的条件概率表 (conditional probability table, CPT)。当一个贝叶斯网络的网络结构 G 和网络参数 θ 已知时,就能求得节点 X_i 的联合概率分布。贝叶斯网络算法因节点间存在条件独立的性质,所以其计算效率远高于其他计算联合概率的算法。

以糖尿病诊断为例,对贝叶斯网络模型进行直观说明。图 1 模拟了 5 点贝叶斯网络模型图,利用“1 和 0”表示事件发生与否,通过分析节点 $A \sim E$ 之间的独立性和相关性可以得出以下结论。

1) 节点的联合概率分布可以表示为

$$P(A, B, C, D, E) = P(E|C)P(D|B, C)P(C|A)P(B|A)P(A) \quad (3)$$

2) 当 C 为已知条件下, A 和 E 相互独立。

3) 当 D 为未知条件下, B 和 C 相互独立。

结论 1 由式 (2) 可得。结论 2 和结论 3 则可根据有向分离中 3 类节点连接类型(见图 2)推断得出:

1) 顺连。有向连接 $a \rightarrow c \rightarrow b$ 被称为顺连,其中,节点 c 被称为头对尾节点。

2) 分连。有向连接 $a \leftarrow c \rightarrow b$ 被称为分连,其中,节点 c 被称为尾对尾节点。

3) 汇连。有向连接 $a \rightarrow c \leftarrow b$ 被称为汇连,其

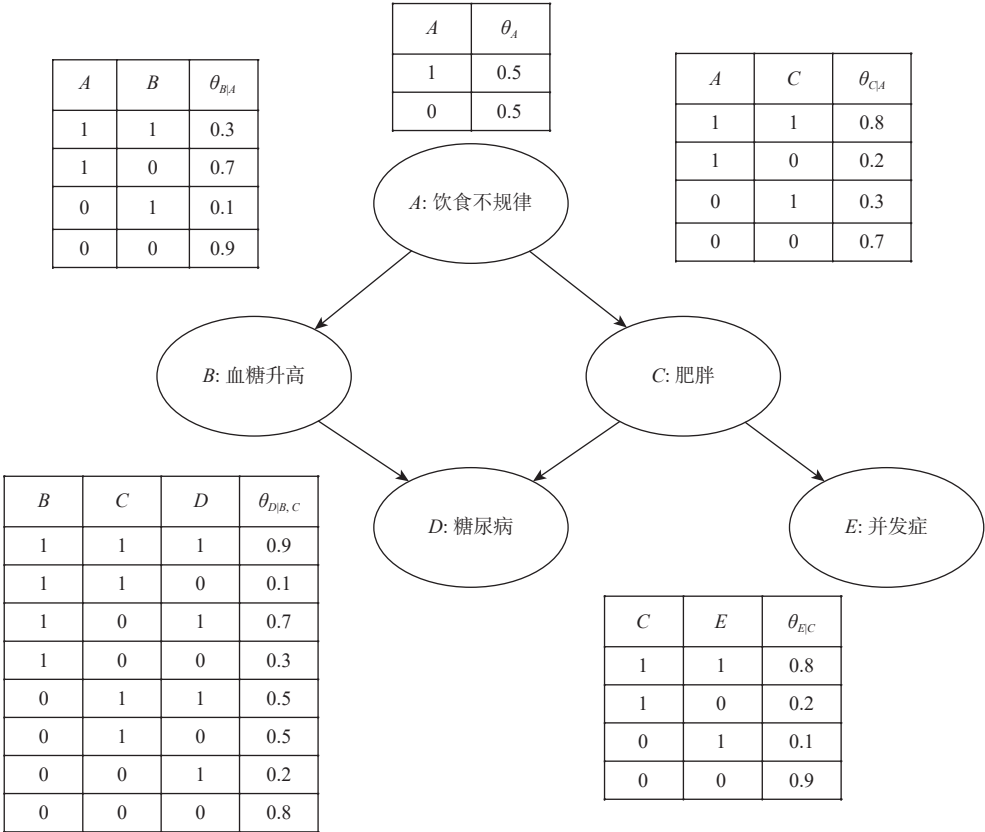


图 1 糖尿病诊断贝叶斯网络模型

Fig. 1 Bayesian network model for diabetes diagnosis

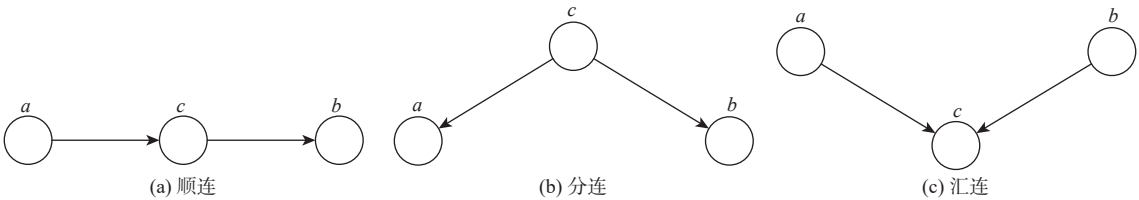


图 2 有向分离中 3 类节点连接类型

Fig. 2 Three types of node connection in d-separation

中, 节点 c 被称为头对头节点。

根据 CI 假设可知, 在顺连中, 当节点 c 未知时, 有: $P(a,b,c) = P(a)P(c|a)P(b|a)$, 但无法推出 $P(a,b) = P(a)P(b)$, 即 a 、 b 之间存在相关性。当节点 c 已知时, 有

$$P(a,b|c) = P(a,b,c)/P(c) = P(a)P(c|a)P(b|c)/P(c) = P(a,c)P(b|c)/P(c) = P(a|c)P(b|c) \quad (4)$$

即 a 、 b 之间被 c 阻断, 称为头对尾条件独立。

同理, 在分连中, 当 c 未知时, a 、 b 节点相关; 当 c 已知时, a 、 b 节点被 c 阻断, 称为尾对尾条件独立。例如, 图 1 中, 当肥胖未知时, 饮食不规律和并发症之间存在相关性; 而当肥胖已知时, 饮食不规律和并发症关于肥胖条件独立 (见结论 2)。

在汇连中, 当节点 c 已知时, 节点 a 、 b 相关, 可以理解为在结果已知的因果关系中, 各原因之间是存在相关性的; 而当节点 c 未知时, a 、 b 关于 c 条件独立, 称为头对头条件独立。例如, 图 1 中, 当糖尿病未知时, 血糖升高和肥胖之间条件独立 (见结论 3)。

2 贝叶斯网络结构学习方法

贝叶斯网络结构构造有专家咨询和数据学习 2 种方法。在贝叶斯网络发展初期, 通常采用专家咨询的构造方法, 依据专家经验得到网络参数, 再手工建立网络结构模型, 通过减少参数的个数来提高咨询结果的质量, 从而得到一些简单的结构模型。因果独立和环境独立是专家咨询中 2 种常见的局部结构^[13]。专家咨询的方式适用于先验知识丰富、模型复杂度低的领域, 多被用于构建没有历史统计数据的网络结构。例如, 文献 [14-15] 通过咨询相应领域的专家经验去判断影响水质各因素之间的联系。对于参数多、复杂度高的大型贝叶斯

网络,仅依靠专家经验是难以建立的。

随着信息技术的进步,很多领域都积累了大量的数据,数据正逐步演变成重要的市场要素,大数据挖掘成为医疗、金融、保险、军事等各社会发展领域前进的驱动力。相比于专家咨询的建模方法,贝叶斯网络结构学习方法可以充分挖掘数据中蕴含的逻辑关联,为建立复杂问题的可解释模型提供了具有潜力的技术途径,受到了学者们的广泛关注。

2.1 基于约束的方法

2.1.1 原理概述

基于约束的方法认为结构学习是约束满足性问题,服从马尔可夫假设和忠诚性假设。基于约束的方法将贝叶斯网络看作是具有独立变量的结构模型,先通过互信息 (mutual information, MI) 或 CI 检验的方式判断变量之间的独立关系,学习变量间的骨架图,再利用 V-结构和定向规则对变量间的无向边进行定向,确定变量间的因果方向,从而学习得到完整的贝叶斯网络结构。

基于约束的方法的精确性通常会随着测试数量和约束集的增多而降低。在构造初始无向图时,通常使用卡方统计量或互信息进行 CI 检验。

1) 节点间的卡方统计量公式为

$$\chi^2 = \sum_{i=1}^k \frac{(W_i - Q_i)^2}{Q_i} \tag{5}$$

式中: W_i 为样本观察值; Q_i 为理论值。

假设数据集近似服从自由度为 $k-1$ 的卡方分布,根据卡方统计量做 CI 假设检验,确定节点间边的存在与否,最终确定贝叶斯网络的初始框架。

2) 节点间的互信息量公式为

$$I(x; y | z) = H(x | z) - H(x | z, y) \tag{6}$$

式中: 条件熵 $H(x | z)$ 和 $H(x | z, y)$ 之差称为给定 z 时, x 和 y 之间的条件互信息。通过评价节点间的互信息值,可以度量在已知 y 的观察值时,对 x 取值的依赖程度。如果互信息量小于给定阈值,反映在真实网络中的 2 个节点极大可能不相邻,在构造初始无向图时,2 个节点间不存在相邻边。

2.1.2 文献分析

1991 年,最早的基于约束的方法,即 SGS 算法^[16]被提出,其利用特定的因果模型来判断变量之间的 CI 假设,不需要预先知道节点顺序。然而,SGS 算法具有不稳定性,容易影响后续过程,产生较大误差,且学习贝叶斯网络结构过程中需要进行指数级的 CI 测试,存在工作效率低的问题。Spirtes 和 Glymour 等在 SGS 算法基础上进行改进,提出了

PC 算法^[3],从顶点集上生成的完全无向图开始,利用 d-分离^[17]判断进行 CI 检测,递归删除边,直至生成需要的网络结构模型。对于稀疏和离散的网络,PC 算法不需要测试高阶独立性关系,大大减少了测试次数,具有计算量小、复杂程度低的特点,算法在保证独立性判断准确的情况下,结果也具有一定程度的可靠性,但缺点是收敛速度慢。

2002 年,文献^[18]提出了 TPDA 算法,通过绘图、增厚和削减 3 种操作,计算各节点之间的互信息量,从而确定节点间的条件独立性来构造 DAG 模型。算法将信息论和统计测试结合,以变量间的互信息检测代替了 CI 检验,可以处理已知节点序或未知节点序的模型,生成非常接近底层逻辑的信念网络,在准确度和时间复杂度方面均优于 PC 算法。

2006 年,文献^[19]提出了贝叶斯网络结构学习的递归自治识别 (recursive autonomy identification, RAI) 算法,从一个完全无向图出发,按照 CI 检验、边的有向性确定、自治子结构识别的顺序来学习贝叶斯网络模型。算法中,CI 测试的顺序与 PC 算法类似,都是高阶测试之前先执行低阶测试(即使用小条件集进行测试),通过优先执行更可靠的测试,从而降低测试失败的概率。与其他基于约束的方法先对结构进行 d-分离,再引导产生无向图的学习方式不同,RAI 算法从开始就将这 2 个过程结合起来,使得结构学习只需要较少的高阶 CI 测试,大大降低了算法的复杂性和运行时间。

2010 年,文献^[20]提出了混合父子 (hybrid parents and children, HPC) 算法,用来学习贝叶斯网络中目标变量的父、子节点,寻找父子节点在贝叶斯网络结构学习、降维和特征选取等应用中也是重要的一步。该算法结合了增量法和分治法的思想,能够处理较大的邻域,在小样本内仍保持良好的性能。

2013 年,文献^[21]提出了轻互最小值 (light mutual min, LMM) 算法,用于解决在小样本数据情况下网络结构学习正确率低的问题,通过考虑节点对的条件独立性来改善对候选边缘节点的评估,并采用基于约束的自适应放宽方式,选择性地删除节点的邻接,从而减少由于多次 CI 测试而导致真实边缘被移除的现象。并行计算技术被引入挖掘基于约束的局部因果关系,从而提高了网络结构学习效率。

2014 年,文献^[22]提出了一种优化的零一阶超结构 (optimized zero-first-order super-structure, Opt01-SS) 算法。采用零一阶 CI 测试解决了 2 类问题: ①小样本条件下近似确定性关系导致的 CI 测试结

果不可靠; ②由于数据稀疏性导致的对于同一组变量测试结果不一致。Opt01SS 算法适用于在有限数据情况下对超级结构进行安全有效的估计, 解释在小样本条件下的近似确定性关系, 解决 CI 测试不一致性的问题。

2016 年, 文献 [23] 提出了一种改进的 PC 算法, 即加速 PC (accclcrated PC, APC) 算法, 采用“分而治之”的思想, 对一系列局部邻域结构进行推导, 获得贝叶斯网络结构。APC 算法继承了 PC 算法优先执行低阶 CI 测试的优点, 能够从已执行的测试中推导相关拓扑信息, 决定测试顺序, 有效避免无效测试, 减少计算量。实验结论得出, 与 PC 算法相比, APC 算法可以减少高达 50% 的计算量, 同时实现了质量相同的学习效果。

2018 年, 文献 [24] 提出了高斯核概率密度估计贝叶斯网络精确学习方法, 在高斯核概率密度的熵估计方法基础上, 利用互信息计算各变量之间的依赖关系。该算法分别描述了离散型、连续型和混合型数据的具体计算过程。报警网络的实验表明, 该算法在不影响学习效率的前提下, 提高了结构学习的精度, 在处理稀疏的连续数据时, 也有较好的学习效果。

2021 年, 文献 [25] 提出了适用于连续时间 PC (continuous-time PC, CTPC) 贝叶斯网结构学习方法, 包含一组适合测试 CI 的统计量和一个基于 PC 的启发式算法。其优势在于: 与传统贝叶斯网络相比, 可以将模拟的基本随机过程通过已知的图形结构进行表达; 有向图中的因果方向表示可以跟随着时间的变化而变化; 专家先验知识可以通过黑名单和白名单的方式整合到结构学习中。贝叶斯网络的无环性对整个网络施加了全局约束, 但 CTPC 算法可以学习独立的父集数据集合, 提高了对局部网络学习的准确性。实验表明, CTPC 算法比连续时间搜索和得分 (continous-time search and score, CTSS) 算法具有更好的结构重构精度, 但随着变量的状态增多, 如变量的状态数大于 20 时, CTPC 算法和 CTSS 算法的性能都会迅速下降。

2022 年, 文献 [26] 提出了一种改进的 PC 算法, 即复杂调查 PC (PC complex surveys, PC-CS) 算法。该算法基于重采样技术与改进的 CI 检验来解释样本选择机制, 解决了复杂调查数据引入的选择偏差。该算法从离散变量数据集进行评估, 并推断网络结构。通过实验比较了 PC 算法和 PC-CS 学习网络的能力, PC-CS 算法的汉明距离仅有 3.4, 表明学习出的网络结构的准确性随着样本数量的增加而提高。

2.1.3 方法总结

在概率分布中, 可能存在 DAG 未反映的条件独立关系, 这种情况称为 DAG 和概率分布彼此不忠诚, 即含有隐藏或潜在变量, 统称为非完备信息。表 1 总结了基于约束的方法的特点及应用场景。

从文献分析可知, 学者们主要从以下角度发展基于约束的贝叶斯网络结构学习方法:

- 1) 改进互信息和 CI 检验方式, 减少无效测试和高阶测试数量, 提高算法学习效率。
- 2) 改进 V-结构的定向规则, 提高算法的准确性。
- 3) 改进算法对于大数据样本或有限小数据样本的灵敏度, 提高模型学习算法的适用性和准确性。

2.2 基于评分搜索的方法

2.2.1 原理概述

基于评分搜索的方法代表了贝叶斯网络结构学习的另一个主要类别, 主要由模型选择和模型优化 2 部分构成。其核心思想是: 将所有可能的结构作为定义域, 将衡量特定结构质量的标准作为评分函数, 将寻找最佳结构的过程看作在定义域范围内寻找评分函数最优值的问题。基于评分搜索的方法由一个初始网络开始, 通过搜索算法对网络结构进行优化操作; 使用评分函数对网络结构打分, 将新网络与旧网络结构的分值进行比较, 若评分高于旧网络, 则对新网络继续进行优化, 直到网络模型不再优化或达到搜索迭代次数限制。

2.2.2 评分函数

基于评分搜索的方法使用评分函数来评估贝叶斯网络结构与给定数据集的拟合程度, 评分函数的适合度直接影响网络模型选择的准确性。评分函数主要包含 2 类: 贝叶斯评分函数和信息论评分函数。

1) 贝叶斯评分函数。对于贝叶斯网络 $N = (G, \theta)$, 关于结构 G 的先验知识概括为概率分布 $P(G)$, 称为结构先验分布; 对于给定的结构 G , 关于参数 θ 的先验知识被概括为另一个概率分布 $P(\theta|G)$, 称为参数先验分布。 N 的先验分布可表达为

$$P(G, \theta) = P(G)P(\theta|G) \tag{7}$$

当给定观测数据集 $K = \{K_1, K_2, \dots, K_n\}$ 后, 计算后验概率分布 $P(G|K)$ 。此时, 只考虑最大的后验概率分布对应搜索空间中的结构模型 G^* :

$$G^* = \underset{G}{\operatorname{argmax}} P(G|K) \tag{8}$$

由于

$$P(G|K) = \frac{P(K|G)P(G)}{P(K)} \tag{9}$$

表 1 基于约束的方法的特点及应用场景

时间	算法	算法特点	算法应用
1991	SGS	误差较大、效率低,具有一定的稳定性,是后续约束型算法的建立基础	只适用于完备的小型数据集
1997	PC	利用d-分离判断进行CI检测,递归删除边,算法较为准确,效率高,稳定性不如SGS算法,是经典的约束型算法	适用于完备的高阶CI测试集(大型父节点集)
2000	FCI	基于PC算法,通过验证部分祖先图中的2个邻接顶点的CI测试,进而寻找超集D-Sep(A,B)来消除隐变量带来的影响,相比于PC算法更具有普适性,是用于学习含有隐变量数据集的经典约束型算法	适用于非完备的大型数据集
2002	TPDA	TPDA基于SGS算法,使用V-结构和定向相位来定向边缘,减少了所需CI测试数量的限制,相较于传统PC算法具有更高的准确性	适用于完备的大型数据集
2006	RAI	CI测试方式与PC算法类似,在高阶数据集中,相较于PC算法具有更快的运算速度	适用于完备的大型数据集
2010	HPC	区别于大多基于约束的全局搜索算法,采用局部搜索方式,CI测试每次添加新变量时都执行消除步骤,能够更好地处理小样本数据集,但在高阶数据集中学习效率低	只适用于完备的小样本集
2011	RFCI	基于FCI算法,仅利用邻接阶段节点父级的条件集进行学习,并且修改了V-结构的定向规则,从而减少了高阶CI测试量,有效提高了FCI算法的运算速度	适用于非完备的大型数据集
2013	LMM	专用于解决小样本数据情况下网络结构学习正确率低的问题	只适用于完备的数据有限的小样本集
2013	FCI+	基于FCI算法,使用邻接节点父级的条件集的同时能够有效识别祖先集,因此,可以在复杂度更高的数据集中更有效地学习稀疏因果图	适用于非完备的大型数据集
2014	PC-Stable	通过CI测试前重新计算邻接概率,降低了传统PC算法由于样本量有限而导致的错误CI测试决策的影响,在高维图中,PC-Stable算法相较于PC算法具有较高的准确性和稳定性,但运算速度有所降低	适用于完备的数据量有限的小样本集,在大型数据集中准确率和运算速度有所下降
2014	Opt01SS	只执行零阶和一阶CI测试,用于学习数据有限情况下小样本数据的结构	只适用于完备的小样本集
2014	OptHPC	建立在局部搜索的HPC算法的基础上,是一种有效的全局超结构恢复方法,与Opt01-SS算法与HPC算法相比,OptHPC算法在大样本量下具有良好的特异性与准确性	适用于完备的大型数据集
2016	APC	继承了PC算法的高数据效率特性,动态调整CI测试顺序,节省了计算量,因此,APC算法相较于PC算法具有更快的运算速度	适用于完备的大型数据集
2018	MGM-FCI-MAX	基于回归的测试来检测不同变量类型的条件独立性,以及使用最高p值的Sepset来识别V-结构,提高了算法的精确度,另外,算法的并行化缩短了其运行时间	适用于非完备的大型数据集
2021	CTPC	包含一组适合测试CI的统计量和一个基于PC的启发式算法,能够学习连续时间的网络结构,具有较好的精确性和学习效率	适用于完备的连续时间数据集;在大型数据集中精度有所下降
2022	PC-CS	利用重采样技术与改进的CI检验,解决了复杂调查数据引入的选择偏差,PC-CS算法相较于PC-Stable算法具有更高的准确性与运算速度	适用于完备的大型数据集

而 $P(K)$ 不依赖于 G ,选择后验概率最大的结构就是选择使如下函数达到最大的结构:

log P(G, K) = log P(K|G) + log P(G) (10)

$\log P(G, K)$ 称为结构 G 的贝叶斯评分,先验分布 $P(G)$ 一般被假设为均匀分布,在评分函数中,对数函数 $\log$ 默认以自然对数 e 为底数,且使用其他底时,通过换底公式转换后,评分函数的相对大小关系不变,因此,不影响网络结构的优化结果。利用贝叶斯评分函数进行结构学习的实质过程是:先找出后验概率最大的结构 G^* ,再对 G^* 进行贝叶斯参数估计。

与贝叶斯评分函数选择模型等价的评分函数有CH评分^[27]、贝叶斯狄利克雷(Bayesian-Dirichlet, BD)评分^[28]、分数等效贝叶斯狄利克雷(Bayesian-Dirichlet equivalent, BDe)评分^[29]和贝叶斯狄利克雷等效一致性(Bayesian-Dirichlet equivalent uniform, BDeu)评分等^[30]。

2) 信息论评分函数。信息论评分函数基于编码长度和信息论的概念将数据集进行压缩,贝叶斯网络的得分与网络对数据集所能实现的压缩程度

有关。信息论评分函数包括最小描述长度(minimum description length, MDL)准则、BIC 准则、Akaike 信息准则(Akaike information criterion, AIC)、归一化极大似然(normalized maximum likelihood, NML)函数和相互信息测试分数(mutual information tests, MIT)^[31]等。

经过大量试验分析,贝叶斯评分函数能够应对样本数较大的DAG结构,而信息论评分函数对于小样本数据搜索学习到的网络结构更优^[32]。对贝叶斯评分函数和信息论评分函数进行梳理,将其基本表达式和特性进行整理归纳,如表2所示。表中: n 为样本变量数目, q_i 为 X_i 父节点个数, r_i 为 X_i 取值数目, m_{ijk} 表示在父节点组合 j 下取值为 k 的次数, $m_{ij*} = \sum_{k=1}^{r_i} m_{ijk}$ 为数据中父节点组合 j 出现的次数, N_{ijk} 为其父节点组合为 j 的样本数, N_{ij} 为总样本数, α_{ijk} 为参数先验分布中的超参数, $\alpha_{ij*} = \sum_{k=1}^{r_i} \alpha_{ijk}$, $\text{MIK}(X_i, \text{paG}(X_i))$ 为 X_i 与其父变量间的互信息, $\chi_{\alpha, d, (r_i, q_i)}$ 表示卡方分布值, $C(G) = 0.5 \log m \sum_{i=1}^n (r_i - 1) q_i$ 为惩罚函数。

表 2 基于评分函数的方法的基本表达式和特性

类别	评分函数	评分表达式	特性
贝叶斯评分函数	贝叶斯评分	$\log P(G, K) = \log P(K G) + \log P(G)$	最优网络结构满足 $P(G K)$ 后验概率分布值最大
	BD	$BD(G K) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij*})}{\Gamma(\alpha_{ij*} + m_{ij*})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + m_{ijk})}{\Gamma(\alpha_{ijk})}$	参数的先验分布符合狄利克雷分布, 但精确度较低
	BDeu	$BDeu(G K) = \log P(G) \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\frac{\alpha}{q_i})}{\Gamma(\frac{\alpha}{q_i} + m_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\frac{\alpha}{r_i q_i} + m_{ijk})}{\Gamma(\frac{\alpha}{r_i q_i})}$	评分算法假设数据 D 是完整的, 服从均匀分布
	K2	$K2(G K) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!$	利用节点序、CH 评分来衡量结构的优劣性, 并使用贪婪搜索得到评分最高的网络结构, 是一种经典的评分函数
	CH	$CH(G K) = \sum_{i=1}^n \sum_{j=1}^{q_i} \left[\log \left(\frac{\Gamma(\alpha_{ij*})}{\Gamma(\alpha_{ij*} + m_{ij*})} \right) + \sum_{k=1}^{r_i} \log \left(\frac{\Gamma(\alpha_{ijk} + m_{ijk})}{\Gamma(\alpha_{ijk})} \right) \right]$	结构先验均匀分布, 需要提前选定超参数 a_{ijk} , 当 $a_{ijk} = 1$ 时, 则 CH 评分转变为 K2 评分
信息论评分函数	BIC	$BIC(G K) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \log \theta_{ijk} - \frac{1}{2} \sum_{i=1}^n q_i (r_i - 1) \log m$	能够度量结构与数据的拟合度, 具有关于模型复杂度的罚项, 用于大样本数据中, 是一种被广泛使用的评分函数
	MDL	$MDL(G K) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \log \left(\frac{m_{ijk}}{m_{ij}} \right) - \frac{1}{2} C(G) \log m$	对数据进行压缩, 从而降低数据的编码长度
	AIC	$AIC(G K) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \log \left(\frac{m_{ijk}}{m_{ij}} \right) - \sum_{i=1}^n (r_i - 1) q_i$	当实例数据服从多项分布时, 对 MDL 进行简化而得, 能够提高模型拟合度
	MIT	$MIT(G K) = \sum_{i=1, \text{paG}(X_i) \neq \emptyset}^n \left(2N \cdot \text{MIK}(X_i, \text{paG}(X_i)) - \max_{\sigma_i} \sum_{j=1}^{q_i} \chi_{\alpha_j \sigma_i, \beta} \right)$	基于信息论和 CI 的评分函数, 通过度量网络结构与观测数据之间的 KL 距离来评价网络结构的优劣

通过评分函数公式可以发现, 评分函数具有分数可分解性, DAG 的分数由每个变量 X_i 与其父集 $\text{pa}(X_i)$ 得到的分数的总和构成:

score(G, K) = $\sum \text{score}(X_i, \text{pa}(X_i), K)$ (11)

2.2.3 搜索策略及文献分析

模型优化是通过搜索策略在所有可行结构组成的空间中寻找评分最高网络结构的过程。实际上, 这是一个 NP 难题, 即使每个节点的最大父节点数为固定值, 搜索空间也是巨大的。有学者引入专家知识指导搜索过程, 以达到对搜索空间进行限制的目的, 如 K2 算法、K3 算法等; 还有学者采取了启发式搜索策略, 利用启发信息引导搜索过程, 减少搜索范围, 降低问题复杂度, 如贪心算法、模拟退火 (simulated annealing, SA)^[33] 算法、粒子群优化 (particle swarm optimization, PSO) 算法等。

1992 年, Cooper 和 Herskovits 提出了著名的 K2 算法^[7], 先根据专家经验确定节点顺序, 通过贝叶斯评分准则评估候选结构, 再使用爬山 (hill-climbing, HC) 搜索策略寻找评分最高的模型。K2 算法的精度取决于节点序的优劣, 其使用先验信息优化搜索过程, 该思路被后续贝叶斯网络结构学习方法广为借鉴。1994 年, Bouckaert 提出 K3 算法^[34], 采用 MDL 准则, 增加了约束网络复杂程度的罚项函数, 避免学习模型时出现过拟合。2022 年, 为解决 K2 算法学习效果强烈依赖于节点序的问题, 一种新的混合结构学习算法——双重 K2 算法^[35] 被提出。首先,

通过 K2 算法的搜索策略得到初始网络结构; 然后, 利用拓扑排序修正节点序; 最后, 使用 K2 算法通过修正后的节点序学习得到最优的网络模型。

在采用启发式搜索策略的算法中, 爬山算法^[36] 是一种局部搜索算法, 其核心思想是每次都选择局部最优解, 分数到达稳定状态后结束搜索, 输出最优网络。该算法的缺点是全局开发能力弱, 容易产生局部最优解。为解决局部最优的问题, 学者们提出了基于单个解的优化算法进行改进, 包括禁忌搜索 (tabu search, TS)^[37]、模拟退火等。

近年来, 基于自然选择的群体智能优化算法引起了学者的广泛关注。这些优化算法包括遗传算法 (genetic algorithm, GA)^[38]、进化规划^[39]、蚁群优化 (ant colony optimization, ACO) 算法^[40]、粒子群优化算法^[41-42]、布谷鸟优化^[43]、象群优化^[44]、水循环优化^[45]、人工蜂群 (artificial bee colony, ABC) 算法^[46] 和细菌觅食优化 (bacterial foraging optimization, BFO) 算法^[47-48] 等。基于群体智能的优化算法的共同特点是: 使用元启发式搜索机制对贝叶斯网络结构空间进行探索, 通过多次计算对局部模型进行状态更新, 在每次计算时都会创建新的种群来寻找局部最优解。随着循环次数的增加, 解的质量不断提高, 最终可以接近最优模型。

表 3^[48] 给出了 5 种算法在 10% 和 20% 噪声数据比例下 Alarm-3000 数据集集中学习的详细对比结果。其中, HKS 表示在所有试验中获得的最高

表 3 5 种算法在不同噪声数据比例下的对比实验结果^[48]

Table 3 Comparative experiment results of five algorithms under different noise data proportions ^[48]										
噪声比/%	算法	得分			结构			时间/s		
		HKS	LKS	AKS	SSD	BSD	ASD	SET	LET	AET
10	ACO-B	-14 804.11	-14 810.40	-14 805.79±2.28	2	13	5.55±2.58	358	567	402.5±86.5
	ABC-B	-14 804.11	-14 810.40	-14 804.67±1.46	2	9	4.25±1.26	117	169	141.9±17.0
	BFO-B	-14 804.11	-14 813.26	-14 805.75±2.44	2	6	3.65±1.06	343	477	401.0±36.0
	TS-B	-14 805.81	-14 805.81	-14 805.81±0	2	2	2.00±0	238	245	240.1±2.5
	HC-B	-14 953.06	-14 953.06	-14 953.06±0	11	11	11.00±0	7	8	7.1±0.2
20	ACO-B	-15 247.18	-15 257.07	-15 248.67±2.18	5	14	8.80±2.44	332	495	389.9±47.8
	ABC-B	-15 247.18	-15 256.65	-15 247.70±2.06	8	13	8.50±1.36	114	182	154.9±18.0
	BFO-B	-15 247.18	-15 256.92	-15 252.30±3.50	6	17	10.35±2.74	324	439	376.5±34.9
	TS-B	-15 293.55	-15 293.55	-15 293.55±0	20	20	20.00±0	251	262	254.7±3.1
	HC-B	-15 324.16	-15 324.16	-15 324.16±0	14	14	14.00±0	7	8	7.1±0.3

K2 分数; LKS 表示在所有试验中获得的最低 K2 分数; AKS 表示在所有试验中获得的平均 K2 分数; BSD 表示在所有试验中获得的最大结构差异; SSD 表示在所有试验中结构差异最小; ASD 表示在所有试验中收敛的平均结构差异; SET 表示在所有试验中获得的最小执行时间; LET 表示在所有试验中获得的最长执行时间; AET 表示所有试验的平均执行时间。结果以 K2 分数为准, 结构差异最小的网络并不对应得分最高的网络。

与非群体智能优化算法相比, 3 种群体智能优化算法在每项指标下几乎一致获得最佳结果(10% 噪声数据下得分-14 804.11; 20% 噪声数据下得分-15 247.18)。基于贪婪搜索的算法性能一般, 在大多数情况下明显不如 3 种群体智能优化算法, 并且在 20% 噪声数据下, SSD 和 BSD 从 2 下降到 20, 说明在噪声的影响下, 结构差异越来越大。基于爬山搜索的算法对噪声的增加不敏感, 几乎在所有情况下性能指标都是最差, 但从执行时间上是最快的, 远远优于其他 4 种算法, 原因是: 群体智能优化算法是一种基于群体的优化算法, 每个个体在每次迭代中都要独立执行优化任务, 并完成个体之间的信息交换, 因此, 时间成本较高, 同时也获得了较好的鲁棒性。

2011 年, 文献 [49] 提出了一种基于改进遗传算法的贝叶斯网络结构学习方法, 增加了优化变异和修正非法图 2 个新的算子, 代替经典遗传算法中的交叉、选择、变异等操作, 新算子不但保持了贝叶斯网络结构学习的多样性和正确性, 还能缩小网络结构的搜索空间, 提高搜索效率。

2012 年, 文献 [50] 提出一种 K2 与模拟退火相结合的贝叶斯网络结构学习方法——K2SA 算法, 该算法的主体为模拟退火算法, 对贝叶斯网络结构

模型的扰动方式进行了修改, 通过随机交换当前节点序中 2 个节点的位置以产生新的节点序, 再利用 K2 算法学习贝叶斯网络结构, 作为模拟退火算法中的新状态, 以此提高模型全局扰动能力。算法记录了在模拟退火过程中得到的最优网络结构, 并在模拟退火过程结束后再利用爬山算法对其作进一步优化。该算法一定程度上改善了模拟退火算法因遍历模型空间能力不足而难以获得全局最优网络结构的问题。实验结果表明, 在样本量较为充足的情况下, 该算法能够获得近似全局最优的网络结构, 具有较好的学习效果。

2021 年, 文献 [51] 针对基于粒子群优化算法的贝叶斯网络结构搜索容易陷入局部最优的缺点, 提出一种结合灰狼算法的粒子群优化算法。首先, 对每次迭代数据中不符合迭代规则的粒子进行优化; 其次, 由于迭代过程中整个粒子群中最优粒子具有极强的引导搜索能力, 对最优粒子进行干扰, 增大寻找全局最优的可能性; 最后, 结合灰狼算法, 引导粒子群进行全局包围搜索, 寻找全局最优解。

2022 年, 文献 [52] 提出了一种基于评分搜索的离散萤火虫优化算法 (discrete firefly algorithm, DFA-B) 用于学习贝叶斯网络结构。该算法使得每个萤火虫在离散空间中基于重定义的策略进行移动, 同时采用变异算子以防算法过早停止而陷入局部最优, 利用局部优化器提高萤火虫的使用效率, 获得最佳可行解。实验表明, DFA-B 算法学习得到的 K2 分数的标准差为 15.2, 随着信用网络数据集样本量的增加, 算法学习到的结构差异性减小。

2022 年, 文献 [53] 提出了一种模型平均爬山 (model averaging hill-climbing, MAHC) 算法。该算法应用修剪搜索策略对 DAG 搜索空间进行预处理, 在爬山搜索空间上应用模型平均确定最佳网络

结构,同时采用 BIC 评分函数对网络进行打分。与传统的爬山算法不同,MAHC 算法通过对比所有邻域图,以此来确定最佳网络结构。实验表明,MAHC 算法在数据噪声存在下的复杂网络中仍具有较好的学习性能。

2.2.4 方法总结

除文献分析外,表 4^[7,47,52-73]总结了部分基于评分搜索的方法使用的搜索空间、评分函数及搜索策略,帮助读者在建立贝叶斯网络时更加灵活及有针对性地选择学习方法。

表 4 基于评分搜索的方法的搜索空间、评分函数及搜索策略^[7,47,52-73]

Table 4 Search space, scoring function, and search strategy based on method of scoring search^[7,47,52-73]

时间	算法	搜索空间	评分函数	搜索策略
1968	Chow-Liu ^[54]	Tree	Entropy	
1992	K2 ^[7]	DAG	K2	爬山
1995	TS ^[55]	DAG	K2 & BIC	禁忌搜索
1995	HC ^[56]	DAG	BDe	爬山
1996	GA ^[57]	Orderings	K2	遗传算法
2001	K2SN ^[58]	Orderings	K2	爬山
2002	GES ^[59]	Equivalence	BDeu	贪婪搜索
2002	ACO-B ^[60]	DAG	K2	ACO+局部搜索
2005	SGA ^[61]	Orderings	K2	遗传算法
2025	Hybrid MCMC ^[62]	DAG	BDeu	MCMC
2008	ACO-K2SN ^[63]	Orderings	K2	ACO+局部搜索
2009	BN-BQPSO ^[64]	DAG	BIC	粒子群优化
2009	I-ACO-B ^[65]	DAG	K2	ACO+局部搜索
2010	IBPSO ^[66]	DAG	MDL	粒子群优化
2012	GIES ^[67]	Equivalence	BIC	贪婪搜索
2013	PSO-K2 ^[68]	Orderings	K2	粒子群优化
2017	FGES ^[69]	Equivalence	BIC	贪婪搜索
2016	BFO-B ^[47]	DAG	K2	细菌觅食优化
2017	Partition MCMC ^[70]	DAG	BIC	MCMC
2020	CIGAR ^[71]	Orderings	BDe	遗传算法
2020	K2-I ^[72]	DAG	BIC	爬山
2023	FGES-M ^[73]	Equivalence	BIC	贪婪搜索
2022	MAHC ^[53]	DAG	BIC	爬山
2022	DFA-B ^[52]	DAG	K2	萤火虫优化算法

基于评分搜索的方法与基于约束的方法中利用 CI 测试及 V-结构定向的学习策略不同,需要确定具体的搜索策略与评分函数。目前,学者们主要从以下方面提升算法性能:

- 1) 改进经典启发式搜索策略,如禁忌搜索、爬山算法等,使其更好地避免陷入局部最优值,进而学习得到最佳的网络结构。
- 2) 融合不同类型群体智能优化算法,如粒子群优化算法、蚁群优化算法等,达到优势互补的目的,优化算法性能,克服收敛速度慢、精确度低等主要问题。
- 3) 评分函数选择方面,综合考虑模型对于数据样本的拟合度和模型复杂度,以提高模型质量和学习效率。
- 4) 搜索策略设计方面,综合考虑算法的全局搜

索能力和时间复杂度,以减小在搜索过程中出现局部最优解的几率,提高算法的搜索效率。

2.3 混合约束和评分搜索的方法

2.3.1 原理概述

事实上,基于约束的方法实现相对简单,但性能随着约束集的增多而降低,对于高阶 CI 检验,需要足够多的样本数据才能保证学习结果的可靠性。基于评分搜索的方法虽然目前应用较多,但存在计算复杂、收敛速度较慢等问题。随着各种智能优化算法的发展,涌现了一批混合约束和评分搜索的贝叶斯网络结构学习方法。

混合方法可以结合基于约束和基于评分搜索方法的优点,通过 CI 检验缩小、限制搜索空间,并采用基于评分搜索的方法学习最优网络结构,对大型网络学习速度有明显提高。

2.3.2 文献分析

早期的混合方法在每次迭代中都有一个约束条件和分数最大化的步骤, Singh 等提出了混合方法中的第 1 个方法——CB 算法, 利用 PC 算法确定节点次序及条件设置, 再最大化步骤定向无向边, 得到邻接节点最高 K2 评分, 重复限制和最大化步骤, 增加条件设置大小, 直到生成的 DAG 分数不再提高, 完成网络结构学习^[74]。随后, 基本图搜索 (essential graph search, EGS) 算法^[75] 被提出, 使用 PC 算法搜索多个基础图, 将每个基础图随机转换为 DAG, 并使用贝叶斯度量对 DAG 进行评分, 最终使用贝叶斯评分度量搜索评分最大的 DAG。类似地, 重启 I-Map(I-Map restart, IMAPR) 算法^[76] 使用 CI 测试来寻找一个更优的 DAG 生成起点, 以重新开始局部结构搜索。

最著名的混合方法是最大-最小爬山 (max-min hill climbing, MMHC) 算法^[77]。在约束阶段, MMHC 算法采用基于最大-最小父子 (max-min parents and children, MMPC) 算法使变量进入目标节点的候选父子 (candidate parents and children, CPC) 节点集之后, 再移除多余变量, 得到每个节点的父子节点集, 从而构建出网络框架, 并在最大化阶段融合禁忌搜索和 K2 评分得到最优 DAG 结构。

近年来, 学者们提出了一些新型混合方法应用于贝叶斯网络结构学习。2017 年, 文献 [78] 提出分离与重聚 (separation and reunion, SAR) 算法, 学习大型贝叶斯网络的任务时, 先将其分解为相对较小的贝叶斯网络, 确切地说, 只使用少部分 CI 检验构造前视部分图^[79], 再将其分解成若干大小合适的小子图, 使用基于评分搜索的方法从每个子图中学习一个小型贝叶斯网络, 通过将这些小型贝叶斯网络统一起来, 形成实际的贝叶斯网络整体。

2018 年, 文献 [80] 提出了一种连续变量的自适应限制贪婪等价搜索 (adaptively restricted GES, ARGES) 算法, 使用 MMPC 算法作为限制步骤, 并使用改进的贪婪等价搜索 (greedy equivalence search, GES) 算法作为最大化步骤, 构建网络框架。ARGES 算法允许在学习完全部分有向无环图 (completed partially directed acyclic graph, CPDAG) 的同时在 V-结构上添加额外边, 随着迭代次数的增加, 这些额外的屏蔽边缘会自动删除, 从而不会出现在最终的 CPDAG 中, 因此, 该算法被称为“自适应限制”。实验表明, “自适应限制”意味着 ARGES 算法具有高维的一致性。

2019 年, 文献 [81] 提出了 2 种数据驱动的对称校正方法, 即基于分数的方法 symG 和基于约束的方法 symC, 用于确定节点对之间的对称校正问题,

并给出了一种结构学习与数据驱动对称修正方法相结合的混合框架, 用于提高贝叶斯网络的网络性能, 解决非对称的相邻节点对和马尔可夫毯在局部结构之间的冲突问题。

2020 年, 文献 [82] 提出一种基于局部拓扑信息分解的贝叶斯网络结构学习方法, 主要运用模型分解的思想, 且考虑了局部邻域结构的知识。算法主要步骤分为 4 个阶段: ①使用马尔可夫覆盖发现方法绘制无向独立图; ②将整个网络拆分为一系列子图; ③从观测数据中学习结构规模较小的贝叶斯网络; ④通过组合这些小贝叶斯网络来获得最终的贝叶斯网拓扑结构。

2021 年, 文献 [83] 将 PC 算法和粒子群优化算法结合为贝叶斯网络结构学习提供合理的解决方案。首先, 使用 PC 算法生成初始 DAG; 然后, 在搜索过程中结合粒子群优化算法和遗传算法找到最优结构。该算法使用了 2 种新的变异算子, 其中一种被称为添加和删除均匀变异算子, 另一种被称为均匀交叉变异算子。该算法将 PC 算法的输出作为先验结构, 改进了贝叶斯网络结构学习的初始解。实验表明, 该算法在 CANCER 网络中的 BIC 评分为 -1 091.88, 接近标准 BIC 评分, 能够得到较优的网络结构。

2023 年, 文献 [84] 提出了基于 PC-SSA 算法的贝叶斯网络混合结构学习方法。首先, 采用 PC 算法生成初始网络图作为结构先验, 并基于它们生成初始解; 然后, 采用 BIC 评分和麻雀搜索算法 (sparrow search algorithm, SSA) 进行贝叶斯网络结构寻优。与其他启发式搜索算法相比, 麻雀搜索算法的时间代价小, 设置参数少, 在标准网络上的测试分数误差仅为 0.2, 具有较高的精度。

2.3.3 方法总结

混合方法可以提高模型学习质量和效率, 是优化贝叶斯网络结构学习方法的有效手段。通过文献分析, 混合方法研究主要集中在以下方面:

- 1) 探索不同人工智能算法的融合能力, 发挥智能算法优势, 突破传统算法学习短板, 提高贝叶斯网络结构学习效率与学习精度。
- 2) 强化对大规模复杂网络的学习能力, 利用贝叶斯网络本身特点与数据所提供的信息, 缩小搜索空间, 用启发式信息指导搜索过程, 减少搜索盲目性, 提高学习效率。

3 非完备信息下贝叶斯网络结构学习方法

结构学习方法在一般情况下通常假设数据集

是完整的, 但真实的数据集往往存在 2 类情况: ①部分观测变量数据缺失的情况; ②存在无法被观测的变量, 这类变量被称为隐结构变量。针对上述 2 种情况, 学者们研究了非完备信息下的贝叶斯网络结构学习方法。由于现实各领域积累的数据往往具有非完备性, 该方向的研究具有良好的工程应用前景。

3.1 数据缺失的贝叶斯网络结构学习

3.1.1 原理概述

无论是因为偶然性缺失还是系统性缺失, 缺失的数据都可以分为 3 类: 随机缺失 (missing at random, MAR)、完全随机缺失 (missing completely at random, MCAR) 和非随机缺失 (missing not at random, MNAR)^[85]。MAR 类型数据的缺失依赖于其他已知变量。MCAR 类型数据的缺失是随机发生的, 且丢失概率与数据变量已知或未知无关。MNAR 类型数据缺失的概率取决于已观测到和未观测到的数据本身。

期望最大化 (expectation maximization, EM) 算法^[86]是非完备信息下使用最广泛的网络模型学习算法之一。该算法从网络参数 θ 的某个初始值 θ_0 出发开始迭代, 假设进行了 t 次迭代得到估计 θ_t , 则 $t+1$ 次迭代过程由 2 个步骤组成: 首先, “E-step” 利用估计值 θ_t 对数据进行补全; 然后, “M-step” 利用补全后的完整数据计算网络参数 θ 的最大似然估计, 得出 θ_{t+1} 。EM 算法重复以上 2 个步骤, 直至模型不再优化或达到最大迭代次数。

如何将缺失数据合理完备化是非完备信息下贝叶斯网络结构学习的关键, 具体学习流程如图 3 所示。

3.1.2 文献分析

EM 算法具有对初始化数据敏感、计算复杂、收敛较慢、不适用于大规模数据集和高维数据、容易出现陷入局部最优的问题。为避免以上问题的出现, Friedman 在 EM 算法基础上提出了结构期望最大化 (structural expectation maximization, SEM)^[87] 算法, 将 EM 算法与结构搜索模型相结合, 学习更优的结构和更加精确的参数。文献 [88] 表明, SEM 算法采用 BIC 和 MDL 评分函数取得了比 EM 算法更好的模型学习结果。

2008 年, 文献 [89] 将进化算法 (evolution algorithm, EA) 与传统 EM 算法相结合, 提出了一种处理具有缺失值的数据库的新方法, 并应用进化算法搜索更优的贝叶斯网络, 实验结果证明了进化贝叶斯网络 (evolutionary Bayesian network, EBN) 结构学习方法可以从非完备的数据集中学习得到更优的网

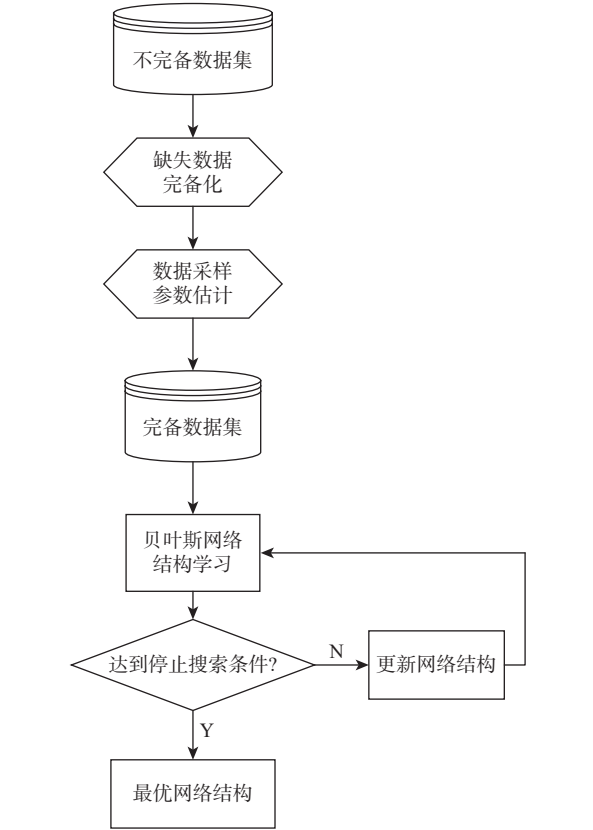


图 3 数据缺失的贝叶斯网络结构学习流程

Fig. 3 Flow of Bayesian network structure learning with incomplete data

络结构, 同时还探索了 EBN 方法在数据挖掘中应用的可行性。

2018 年, 文献 [90] 提出融合使用 SEM 算法作为贝叶斯网络结构学习的输入, 同时在 EM 算法的最大化步骤中采用 k-MAX 算法, 用于学习有界树宽贝叶斯网络, 实验证明, 采用 k-MAX 算法作为 SEM 算法中的结构学习方法与基于随机森林的方法具有相同估算精度, 但迭代速度更快, 使得 SEM 算法可以应用到含有更多变量的复杂问题中。

除了在 EM 算法框架上进行改进, 学者们近期还提出了其他一些从含有缺失值的数据中进行的结构学习方法。2019 年, 文献 [91] 提出一种基于 Gibbs 采样^[92] 的马尔可夫链蒙特卡罗 (Markov chain Monte Carlo, MCMC)^[93] 算法。Gibbs 采样是一种在节点间局部交互的图形结构中执行的采样过程, 用于在难以直接采样时从某一多变量概率分布中近似抽取样本序列, 计算其联合概率分布或积分。该算法用于检测变化点的经验贝叶斯信息准则 (empirical Bayesian information criterion, embIC), 能同时估计变化点配置和相关参数。首先, 使用经验贝叶斯和最大似然原理推导出模型选择或变化点选择标准; 当观测的后验分布复杂且高维时, EM 算法只能得到后验分布的众数, 此时可以利用 Gibbs

采样对贝叶斯网络的后验分布进行抽样,得到 Gibbs 序列,使其收敛到一个独立于初始值的平稳分布,从而求得如后验均值、方差和分位线等更多的后验量。然后,通过在变化点指示器上诱导多变量离散分布,并使用 Gibbs 采样从诱导分布生成马尔可夫链来估计最佳变化点配置。该算法相比于 EM 算法具有更快的收敛速度,且不易陷入局部最优。

2021 年,文献 [94] 提出了一种基于评分搜索学习下的离散节点平均似然 (node-average likelihood, NAL) 的方法,该方法在结构精度与计算效率上与基于 SEM 的算法相比更有竞争力,对离散贝叶斯网络模型具有可识别性,并证明了 NAL 具有可辨识性和一致性。

2022 年,文献 [95] 提出了基于爬山搜索评分的结构学习与配对删除 (pairwise deletion, PD) 和逆概率加权 (inverse probability weighting, IP-W) 方法之间的融合算法。该算法对缺失数据集加权的情况下执行配对删除,并利用原始配对删除的数据集和标准评分函数计算当前最优 DAG 和相邻 DAG 之间的得分差,能够显著降低由数据缺失引起的偏差。通过实验对比可知,该算法在结构学习效率上优于 SEM 算法,且具有较高的学习准确性。

3.2 含有隐变量的贝叶斯网络结构学习

3.2.1 原理概述

隐变量指在贝叶斯网络结构学习中隐藏或潜在的变量,这些变量可以是实际存在但不能被观测的变量,也可以是为了贝叶斯网络结构学习的合理

性而被提出的虚拟变量。例如,在某次医学案例中,训练数据集包含医生观察到的症状(如发烧、头痛、血压等)和处方药物,这些已知变量被称为观测变量;而病人自身不能被观测到的症状就是隐变量。这些隐变量与大多数观测变量是相互独立的,却又能够对患病“原因”和治疗“结果”的解释产生深远的影响。因此,可以通过学习隐藏变量,避免产生对数据的过拟合,提高贝叶斯网络的可解释性,从而更有效地进行推理。

隐变量包括 2 种:①取值个数已知但数值未知的变量,称作值隐变量;②取值个数与数值均未知的变量,称作势隐变量。包含至少一个隐变量的贝叶斯网络模型称为隐变量模型。在实际应用中,往往需要考虑多个隐变量存在的情况,并通过它们与观测变量及自身之间的关系,通过数据分析得到模型中隐变量的个数和势、网络模型结构、模型参数,这一过程称为隐结构模型学习。

含有隐变量的网络结构学习大多采用数据降维的方法,将具有复杂依赖关系的节点变量分解为少数无法直接观测到的公共因子(隐藏变量),以此来确定网络中含有的隐变量和与之对应的观测变量的位置、数量;再通过 EM 算法等补全隐变量的参数。在复杂网络结构中研究隐含变量,能够在简化网络的同时,更深入地了解节点间相互影响的方式。

含有隐变量的贝叶斯网络结构学习方法特性总结如图 4 所示。

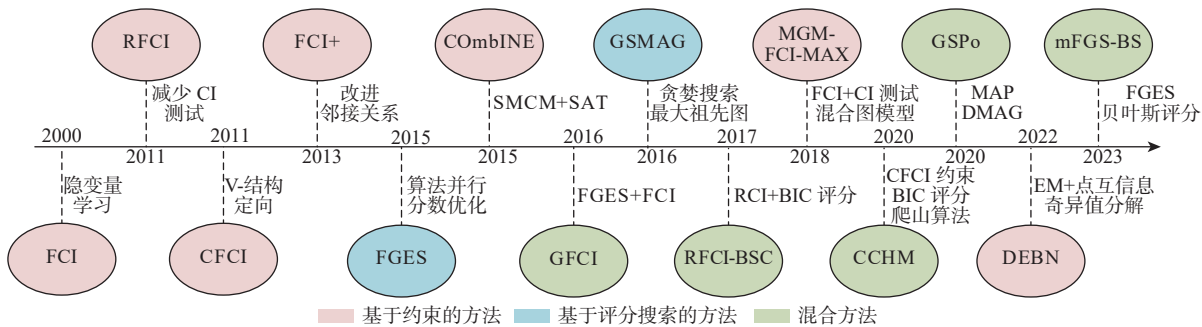


图 4 含有隐变量的贝叶斯网络结构学习方法特性

Fig. 4 Characteristics of Bayesian network structure learning algorithm with hidden variables

3.2.2 基于约束的隐变量学习方法文献分析

最初,在含有隐变量的 DAG 搜索算法中,通常将结构分解算法^[96]和信息瓶颈算法^[97]与结构搜索、参数估计结合在一起,用于学习含有隐变量的贝叶斯网络。但是,这些算法只是启发式搜索,不能准确确定贝叶斯网络中所含有的隐变量个数及位置,它们的输出取决于输入的初始 DAG。

2000 年,Spirtes 等提出的快速因果推理 (fast causal inference, FCI) 算法^[98]是一种基于约束的方

法,其以样本数据和可选背景知识作为输入,能够从大样本中学习含有隐变量的等价类因果贝叶斯网络,但该算法难以保证邻接节点识别和有向弧定向的准确性。2011 年,Colombo 等提出的非常真正快速因果推理 (really fast causal inference, RFci) 和保守快速因果推理 (conservative fast causal inference, CFCI) 算法^[99],以及 2013 年 Claasen 等提出的 FCI+ 算法^[100],都是基于 FCI 的改进算法,通过更精确的初始图来提高 FCI 算法的邻接阶段和定向阶段的

准确性, 并且可以输出一类对观测变量具有马尔可夫等效的图形对象和称为部分祖先图 (partial ancestry graph, PAG) 的图形模型。PAG 表示等价类中所有 DAG 的共有特性, 在大样本情况下, 给予了无反馈的因果马尔可夫假设和因果忠实假设^[101], 可以输出一个含有实际隐变量的 PAG。FCI 早期相关算法的主要限制是在小样本中算法性能差、最终输出网络会出现不同程度的偏差、在大样本中算法运行时间过长等问题。

2015 年, 文献 [102] 提出了一种基于 CFCI 的 COmbINE 算法, 以学习不同数据集的共同特征和 CI 测试结果。COmbINE 算法将数据间的依赖性和独立性转换为半马尔可夫因果模型 (semi Markov causal model, SMCM) 中的路径约束, 并编码为布尔可满足性 (Boolean satisfiability, SAT) 实例。实验结果表明, 在相对稀疏的数据集中, COmbINE 算法具有较好的精度和召回率。

2018 年, 文献 [103] 提出了一种结合混合图形的最大快速因果推理 (mixed graphical models-fast causal inference-max, MGM-FCI-MAX) 算法。算法引入了回归测试, 对邻接节点的所有可能子集执行 CI 检测, 将 CI 检测中 p 值最大的子集作为真正的分离集来识别 V-结构, 以此提高了边的定向精度。该算法还使用了混合图形模型无向图作为邻接阶段的输入, 用于表示混合变量类型之间的条件独立关系。结合以上创新, 通过实验表明, MGM 相比 CFCI 和 FCI, 能够更好地平衡网络学习精度和召回率, 缩短运算时间。

2022 年, 文献 [104] 提出了一种隐变量学习的动态嵌入技术在贝叶斯网络参数学习中的应用。首先, 重建 EM 算法中的 E-step、M-step, 并提出使用动态嵌入来计算分数样本的权重, 以降低网络学习的计算复杂度; 然后, 构建点互信息 (point mutual information, P-MI) 矩阵, 表示从更新的参数转换而来的有向加权图 (directed weighted graph, DWG); 最后, 采用增量奇异值分解 (singular value decomposition, SVD) 生成动态嵌入值, 同时捕获更新的参数, 从而形成贝叶斯网络的图形结构。实验结果表明, 该方法在含有隐变量的贝叶斯网络结构学习上具有较好的性能。

3.2.3 基于评分搜索的隐变量学习方法文献分析

2015 年, Ramsey 提出了一种基于分数的贪婪搜索算法——快速贪婪搜索 (fast greedy search, FGES) 算法^[105], 通过算法并行和缓存分数优化 GES 算法的增添、删除边阶段, 提高算法精度。

2016 年, 文献 [106] 提出了一种基于贪婪的最

大祖先图搜索 (greedy search for maximal ancestral graph, GSMAG) 算法。首先, 在多元正态性假设下, 采用贪婪搜索策略, 从连续数据中识别出最适合的最大祖先图 (maximal ancestral graph, MAG), 其中, 有向和双向边在邻接之间执行添加、删除、反转或转换操作。然后, 使用 Richardson 等提出的用于 MAG 因子分解的 BIC 评分函数^[107], 检查评分最低的网络的局部邻域, 避免重新拟合整个 MAG。在含有隐变量的随机合成高斯网络上评估了 GSMAG 算法与基于 FCI 和 CFCI 约束的算法, 实验表明, GSMAG 算法具有更好的召回率和精度, 但运算速度略有下降。

3.2.4 基于混合的隐变量学习方法文献分析

从观测数据中推断因果 DAG 的马尔可夫等价类问题的主要困难是: 解决在观测变量之间存在的隐变量时, 产生额外的邻接边以及边的定向错误。2016 年, 文献 [108] 提出由 FGES 和 FCI 组成的贪婪快速因果推理 (greedy fast causal inference, GFCI) 算法, 使用 CI 测试删除初始框架中的无关邻接边, 通过修改 FCI 定向规则以产生 PAG 结构图。实验表明, GFCI 算法相较于 FCI 算法提高了学习精度和学习效率, 并在小样本容量下仍具有较好的性能。

2017 年, 文献 [109] 提出了一种基于 RFCI 算法的因果贝叶斯约束评分 (Bayesian scoring of constraints, BSC) 算法的混合方法, 用于学习可能包含隐变量的因果贝叶斯网络 (causal Bayesian network, CBN), 该算法调整了 RFCI 算法, 以使用 BSC 评分进行模型评分搜索。当 RFCI 算法请求 CI 测试时, 利用 BSC 来确定独立性成立的概率, 并以对独立性是否成立进行采样, 能同时对多个可能含有隐变量的因果模型进行评分, 并将该结果返回给 RFCI 算法, 最终生成 PAG。将该算法与 RFCI 算法进行对比, 在小样本限制下, 输出模型结构的汉明距离相较于 RFCI 算法有明显改善。

因果不足的结构学习方法倾向于重构贝叶斯网络的原始图, 其中, 有向边表示节点间直接原始关系。2020 年, 文献 [110] 将基于约束的方法与爬山算法相结合提出了保守规则和因果效应爬坡 (conservative rule and causal effect hill-climbing, CCHM) 算法。CCHM 算法假设贝叶斯网络是线性结构方程模型, 其中, 数据遵循多元高斯分布及 CFCI 的约束算法, 最大化 BIC 等效分数。采用 Pearl 的微积分来观测因果效应, 通过爬山搜索结合因果效应定位 MAG 中的无向边。实验表明, CCHM 算法在重建原始图方面具有较好的高效性和准确性。

2020 年,文献 [111] 针对因果关系不足的问题,提出了贪婪稀疏偏序集 (greedy sparsest poset, GSPo) 算法,偏序集是与有向最大祖先图 (directed maximal ancestral graph, DMAG) 关联的偏节点排序。GSPo 算法提供了从偏序集到最小 I-MAP-DMAG 的映射,并通过遍历偏序集空间关联至新的 DMAG 集,每次新的关联都会产生相同或具有更少边的 DMAG。实验表明, GSPo 算法在高斯网络中的表现比 FCI 和 FCI+算法具有更好的结构精度和更快的运行速度。

2023 年,文献 [112] 提出一种混合多规则及快速贪婪等价搜索与贝叶斯评分 (majority rule and fast greedy equivalence search with Bayesian scoring, mFGS-BS) 算法,能够从缺乏因果性的离散数据集 中进行含有隐变量的结构学习,并生成 PAG。利用 mFGS-BS 算法进行结构学习依赖于混合方法和贝叶斯评分,将该算法与多种算法进行对比,实验结果表明, mFGS-BS 算法能够快速学习得到更精确的网络结构。

3.3 非完备信息下贝叶斯网络结构学习方法总结

由于现实应用中使用的数据通常具有非完备性,而非完备信息的处理极大增加了贝叶斯网络结构学习方法设计的难度,该领域的研究引起了学者们的广泛关注。通过文献分析,非完备信息条件下的贝叶斯网络结构学习主要集中于以下方面:

1) 关于缺失数据学习,采用对缺失数据进行补全的思路进行网络学习,补全数据主要采用 EM 算法及其基础框架上衍生出的一系列算法。目前,通过文献可以证实,缺失数据类型下的贝叶斯网络结构学习方法较为成熟,但如何克服已有算法短板,继续改进 EM 算法框架或提出准确度更高的融合算法,将成为未来该领域的主要研究方向。

2) 关于隐变量学习,主要针对在确定隐变量个数及位置信息后,贝叶斯网络节点信息在因果不足下产生错误的邻接关系这一问题进行改进学习。通过改进约束限制、评分条件、定向规则等方式,提高网络结构学习的精确性和完整性。目前,对于隐变量的结构学习,仍存在算法运行效率低、复杂网络学习效果差的问题。混合方法使用可以一定程度上提高算法处理效率及准确度,而融合其他智能算法进行含有隐变量的贝叶斯网络结构学习,为未来学者们进一步的研究提供了新的解决思路。

4 贝叶斯网络结构学习方法应用

贝叶斯网络可以直观地反映出复杂数据模型中各变量之间的因果概率关系,具有良好的可解释

性。在工程领域广泛运用贝叶斯网络进行问题域建模,应用一系列推理算法来分析计算模型中各个节点变量之间的关系及状态变化。目前,贝叶斯网络算法的典型应用领域包括军事领域、医学领域、生态及生物信息学应用、工业应用等。

4.1 军事领域

在军事作战领域中,通常因为电子通信干扰、传感器噪声等原因造成部分信息缺失。贝叶斯网络具备在非完备数据、不确定信息情况下进行双向推理的能力,适用于战场作战环境,因此,被广泛应用于战场态势评估、目标毁伤效能评估、智能攻击决策、目标识别等领域。

文献 [113] 为处理空战环境中的时序模糊信息序列并有效评估态势,构建了一种直觉模糊动态贝叶斯网络 (intuitive fuzzy dynamic Bayesian network, IFDBN)。在模型中,将直觉模糊推理机作为虚拟节点嵌入到动态贝叶斯网络中,并提出一种将直觉模糊推理输出转化为概率的新方法,该概率可以作为虚拟信息输入到动态贝叶斯网络中,并给出了详细的推理算法,以空战态势评估为例,验证了模型的可行性。

文献 [114] 利用贝叶斯网络处理不确定复杂问题的优点,将动态贝叶斯网络与威胁评估模型相结合,建立了用于评估威胁等级的模型。仿真结果表明,动态贝叶斯网络能够较好地地对具有实时性、不确定性和不可预知性的战场信息进行威胁等级评估。

文献 [115] 构建了一种自主机动决策系统,该系统将空战博弈视为马尔可夫过程,利用贝叶斯推理理论对空战态势进行了有效的计算,根据态势评估结果,自适应调整机动决策因子权重,使目标函数更加合理,保证无人战机的优势态势,实验结果证明了所设计机动决策方法的优越性。

4.2 医学领域

近年来,贝叶斯网络已在医学研究中的多个领域得到应用,包括诊断筛选^[116]、临床试验^[117]、Meta 分析^[118]等,尤其是临床试验已经贯穿了包括先验信息的优化设计、样本量估算、监测、中期分析和预测等全过程。

2016 年, *Machine Learning* 杂志刊登了一期关于“医疗保健和医学机器学习”的特刊^[119],讨论了如何开发利用已知的健康监测数据改善对患者的护理模式。机器学习 (machine learning, ML) 和人工智能 (artificial intelligence, AI) 方法已被提出用于从遗传学和基因组学^[120]到治疗方法选择^[121]及结果预测^[122]的各种医学问题。与来自数据的纯 ML 方法相比,贝叶斯网络的显著优势是:在数据有限的情况下,可以利用

专家知识对数据进行补充,从已知的健康监测数据中进行数据分析,帮助医生实现精准的临床试验决策,这可能是目前最有效的精确医疗的方法之一。

文献 [123] 提出一个动态的高维贝叶斯推理框架用于学习 COVID-19 模型异质性。使用关于确诊、住院和死亡病例数量的报告数据作为先验概率,建立不规则的房室模型,该模型考虑了 COVID-19 在长期护理 (long term care, LTC) 设施内外的时变传播和严重程度的异质性,提出了一种与维数无

关的投影 Stein 变分梯度下降方法用于解决高维推理的维数问题,专门用于在 COVID-19 传播严重程度不确定条件下进行建模、推理,得出了基于后验概率的预测和验证结果。该方法用于对 COVID-19 未来传播轨迹进行预测,以及在风险规避、随机最优控制问题中寻找关于 COVID-19 的缓解策略方面进行了有益的探索。

表 5^[124-131] 介绍了贝叶斯网络在各种不同类型疾病中的病情预测及治疗决策等方面的应用。

表 5 贝叶斯网络在各种疾病中的应用^[124-131]
Table 5 Application of Bayesian networks in various diseases^[124-131]

年份	疾病情况	贝叶斯应用实例	文献引用
2017	中枢神经损伤	使用智能分类器算法预测脊髓损伤患者的疼痛类型	[124]
2018	慢性疾病(糖尿病、关节炎等)	使用DBN的逐步隐变量方法预测慢性疾病并发症	[125]
2015	癌症	深度信念网络用于乳腺癌的预测与诊断	[126]
2017	癌症	使用贝叶斯网络和支持向量机进行乳腺癌手术后生存概率预测	[127]
2016	癌症	使用贝叶斯网络建立乳房病变模型图预测良性肿瘤癌变概率	[128]
2015	心脏疾病	通过电子健康记录数据建立贝叶斯网络模型预测心血管病风险	[129]
2017	心脏疾病	高维数据集集中使用贝叶斯网络建立因高血压引起的并发症模型	[130]
2018	传染病	使用贝叶斯网络建立登革热感染诊断模型	[131]

4.3 生态及生物信息学应用

文献 [132] 将水、能源与生物生态关系耦合到贝叶斯网络中,提供了系统联合概率行为的形式化表示,分析了过去 50 年锡尔河流域的用水状况,为水资源管理和决策提供了优化方案,并证明了该技术能够缓解流域内的生态危机。

文献 [133] 从生态数据中学习贝叶斯网络,对岩石海岸生态系统的物种数据应用贝叶斯网络推理,建立具有记录链接的系统,将贝叶斯网络评估作为一种变量选择方法,训练网络中每个分组物种的神经网络,证明了贝叶斯网络在解释生态网络方面具有很大潜力,以及在处理复杂非线性数据时的强大预测能力。

文献 [134] 提出一种对联结构造贝叶斯网络(pair-copula constructions Bayesian network, PCCBN)模型,对在生物学和生物医学中广泛应用的基因调

控网络 (gene regulatory network, GRN) 进行建模。在基因正态性假设下,通过考虑基因间的条件独立性,对逆协方差矩阵 (或精度矩阵) 的非零元素进行评估,构建了关于 GRN 的贝叶斯网络模型,并将 PCCBN 算法与高斯贝叶斯网络进行了比较。

4.4 工业应用

近年来,贝叶斯网络还与其他数据分析方法相结合,广泛应用于预测分析、可靠性分析或故障诊断等领域,用于保障工业系统正常运行及提升系统的安全性。例如,利用贝叶斯网络进行故障诊断,是贝叶斯网络在工业中的典型应用。对每个故障源及系统建立对应的贝叶斯网络故障诊断模型,利用贝叶斯网络对系统进行反向推理,可以推算出系统出现故障时,每个故障源发生故障的概率。贝叶斯网络在故障诊断、风险评估典型的应用案例如表 6^[135-139] 所示。

表 6 贝叶斯网络在故障诊断、风险评估等工业应用^[135-139]
Table 6 Industrial applications of Bayesian networks in fault diagnosis and risk assessment^[135-139]

年份	故障类型	概述	文献引用
2015	智能故障诊断	提出了一种两级流程操作故障诊断技术,利用贝叶斯网络分析当存在未被监控的流程变量时,无法准确定位故障原因的限制条件	[135]
2016	智能故障诊断	通过结合历史和在线信息,开发了一种贝叶斯故障诊断系统,以解决异步测量问题,可用于事件树分析	[136]
2018	工业故障诊断	通过故障树和事件序列图(FT-ESD)分析,建立了系统的紧急断开故障和后果模型图,将FT-ESD模型映射到贝叶斯网络中,提出了一种利用贝叶斯网络进行紧急断开操作故障概率分析的方法	[137]
2018	工业故障诊断	提出了一种冷冻机组故障诊断方法DR-BN,并将该方法用于评价特征选取结果	[138]
2016	动态风险评估	提出了一种基于贝叶斯网络对液化天然气运输船卸载过程动态风险评估的新方法,能够识别卸载操作中的危险事件,并对关键事件进行失效分析建模	[139]

4.5 贝叶斯网络结构学习方法应用总结

贝叶斯网络被证实为是一种表达、推理不确定

知识的强有力工具,其应用已经涉及预测、智能推理、诊断、决策、可靠性分析等众多领域。经文献

分析可知,未来贝叶斯网络将广泛应用于人类生产生活等众多方面,这对于贝叶斯网络的深入研究和经济社会发展都具有重要意义。

5 结束语

贝叶斯网络因其“白盒性”、工程易用性引起了学者和工程技术人员广泛关注。回顾贝叶斯网络的发展历程,针对贝叶斯网络在复杂问题域中存在的模型维数高、学习难度大、易陷入局部最优解等难题,学者们从不同角度提出了多种算法进行改进,并在多个领域的应用中验证了贝叶斯网络的研究和推广价值。

本文回顾了贝叶斯网络的发展历程,介绍了其基本原理。将贝叶斯网络结构学习方法归为 3 类:基于约束的方法、基于评分搜索的方法及混合方法,并对其各自发展情况进行了梳理和分析。概括来说,基于约束的方法以变量间的 CI 检验为基础,识因果依赖关系,在数据样本完备的条件下,能够学习获得质量较高的贝叶斯网络模型;基于评分搜索的方法设计评分函数作为贝叶斯网络模型选择指标,多以启发式知识指导模型搜索过程,通过多轮迭代持续优化模型;混合方法可以结合基于约束和基于评分搜索的方法的优点,通过 CI 检验缩小搜索空间,并采用基于评分搜索的方法进行网络结构学习,对大型网络结构学习速度有明显提高。针对现实应用中的数据往往具有非完备性,从缺失数据处理和隐变量学习 2 个维度阐释了非完备贝叶斯网络结构学习的研究现状。总结了贝叶斯网络在多个领域中的工程应用情况,揭示了其研究意义和具有的广阔应用前景。

综合分析,未来贝叶斯网络结构学习具有如下发展趋势:

1) 非完备信息条件下的贝叶斯网络结构学习。在军事决策、大型设备故障诊断等复杂应用领域,非完备信息的存在极大增加了贝叶斯网络结构学习的难度,如何准确地将缺失数据完备化,以及提高含有隐变量模型的学习精确度是未来贝叶斯网络结构学习研究的热点之一。

2) 大型或高维度贝叶斯网络结构学习。目前,贝叶斯网络结构学习涉及范围逐渐广泛,复杂领域的需求使得贝叶斯网络算法逐渐向着善于学习大型网络模型的方向发展。网络中节点的增加会导致贝叶斯网络结构学习准确性降低。因此,在高维度贝叶斯网络算法研究中通常会出现局部最优解的情况,导致最终结果偏差较大。如何解决这些问题,寻求适应大型贝叶斯网络结构的学习方法也是

未来的研究重点。

3) 融合人工智能算法的贝叶斯网络结构学习。随着人工智能技术的进步,很多学者已经尝试研究融合不同智能算法的贝叶斯网络结构混合学习方法,以发挥不同智能算法的优点,达到学习复杂程度更高的模型、提升算法性能的目的。可以预期,随着智能算法研究的不断深入,通过融合先进算法来不断完善贝叶斯网络结构学习方法将成为未来研究的发展趋势。

参考文献 (References)

- [1] LEE S J, SIAU K. A review of data mining techniques[J]. *Industrial Management & Data Systems*, 2001, 101(1): 41-46.
- [2] HECKERMAN D, MAMDANI A, WELLMAN M P. Real-world applications of Bayesian networks[J]. *Communications of the ACM*, 1995, 38(3): 24-26.
- [3] SPIRITES P, GLYMOUR C, SCHEINES R. Reply to Humphreys and Freedman’s review of causation, prediction, and search[J]. *The British Journal for the Philosophy of Science*, 1997, 48(4): 555-568.
- [4] COLOMBO D, MAATHUIS M H. Order-independent constraint-based causal structure learning[EB/OL]. (2013-09-27)[2023-07-01]. <https://arxiv.org/abs/1211.3295v2>.
- [5] LIU J X, TIAN Z L. Verification of three-phase dependency analysis Bayesian network learning method for maize carotenoid gene mining[J]. *BioMed Research International*, 2017, 2017: 1813494.
- [6] MASEGOSA A R, MORAL S. An interactive approach for Bayesian network learning using domain/expert knowledge[J]. *International Journal of Approximate Reasoning*, 2013, 54(8): 1168-1181.
- [7] COOPER G F, HERSKOVITS E. A Bayesian method for the induction of probabilistic networks from data[J]. *Machine Learning*, 1992, 9(4): 309-347.
- [8] NEATH A A, CAVANAUGH J E. The Bayesian information criterion: background, derivation, and applications[J]. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2012, 4(2): 199-203.
- [9] LIU H, CAO Y H. Study of heuristic search and exhaustive search in search algorithms of the structural learning[C]//*Proceedings of the 2nd International Conference on Multimedia and Information Technology*. Piscataway: IEEE Press, 2010: 169-171.
- [10] CHAKRABORTY A, KAR A K. Swarm intelligence: a review of algorithms[M]//PATNAIK S, YANG X S, NAKAMATSU K. *Nature-inspired computing and optimization*. Berlin: Springer, 2017: 475-494.
- [11] GEFFNER H, DECHTER R, HALPERN J Y. Probabilistic and causal inference: the works of Judea Pearl[M]. New York: ACM, 2022.
- [12] PEARL J. *Causality: models, reasoning and inference*[M]. Cambridge: Cambridge University Press, 2000.
- [13] DAGUM P, GALPER A. Additive belief-network models[C]//*Proceedings of the Uncertainty in Artificial Intelligence*. Amsterdam: Elsevier, 1993: 91-98.
- [14] BATCHELOR C, CAIN J. Application of belief networks to water management studies[J]. *Agricultural Water Management*, 1999,

- 40(1): 51-57.
- [15] TICEHURST J L, NEWHAM L T H, RISSIK D, et al. A Bayesian network approach for assessing the sustainability of coastal lakes in New South Wales, Australia[J]. *Environmental Modelling & Software*, 2007, 22(8): 1129-1139.
- [16] SPIRITES P, GLYMOUR C, SCHEINES R. From probability to causality[J]. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 1991, 64(1): 1-36.
- [17] GEIGER D, VERMA T, PEARL J. D-separation: from theorems to algorithms[M]//HENRION M, SHACHTER R D, KANAL L N, et al. Uncertainty in artificial intelligence. Amsterdam: Elsevier, 1990: 139-148.
- [18] CHENG J, GREINER R, KELLY J, et al. Learning Bayesian networks from data: an information-theory based approach[J]. *Artificial Intelligence*, 2002, 137(1-2): 43-90.
- [19] YEHEZKEL R, LERNER B. Bayesian network structure learning by recursive autonomy identification[C]//Proceedings of the Structural, Syntactic, and Statistical Pattern Recognition. Berlin: Springer, 2006: 154-162.
- [20] DE MORAIS S R, AUSSEM A. An efficient and scalable algorithm for local Bayesian network structure discovery[C]//Proceedings of the Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2010: 164-179.
- [21] MAHDI R, MEZEY J. Sub-local constraint-based learning of Bayesian networks using a joint dependence criterion[J]. *Journal of Machine Learning Research*, 2013, 14(1): 1563-1603.
- [22] VILLANUEVA E, MACIEL C D. Efficient methods for learning Bayesian network super-structures[J]. *Neurocomputing*, 2014, 123: 3-12.
- [23] MINN S, 傅顺开. 贝叶斯网络结构加速学习算法[J]. *计算机科学*, 2016, 43(2): 263-268.
- MINN S, FU S K. Accelerating structure learning of Bayesian network[J]. *Computer Science*, 2016, 43(2): 263-268(in Chinese).
- [24] JIANG Y L, LIANG Z Z, GAO H, et al. An improved constraint-based Bayesian network learning method using Gaussian kernel probability density estimator[J]. *Expert Systems with Applications*, 2018, 113: 544-554.
- [25] BREGOLI A, SCUTARI M, STELLA F. A constraint-based algorithm for the structural learning of continuous-time Bayesian networks[J]. *International Journal of Approximate Reasoning*, 2021, 138: 105-122.
- [26] MARELLA D, VICARD P. Bayesian network structural learning from complex survey data: a resampling based approach[J]. *Statistical Methods & Applications*, 2022, 31(4): 981-1013.
- [27] ZGUROVSKII M Z, BIDYUK P I, TERENT'EV A N. Methods of constructing Bayesian networks based on scoring functions[J]. *Cybernetics and Systems Analysis*, 2008, 44(2): 219-224.
- [28] SIMONS K T, KOOPERBERG C, HUANG E, et al. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions[J]. *Journal of Molecular Biology*, 1997, 268(1): 209-225.
- [29] YANG S L, CHANG K C. Comparison of score metrics for Bayesian network learning[J]. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 2002, 32(3): 419-428.
- [30] SCUTARI M. An empirical-Bayes score for discrete Bayesian networks[C]//Proceedings of the Conference on Probabilistic Graphical Models. [S. l.]: PMLR, 2016: 438-448.
- [31] DE CAMPOS L M, FRIEDMAN N. A scoring function for learning Bayesian networks based on mutual information and conditional independence tests[J]. *Journal of Machine Learning Research*, 2006, 7(10): 2149-2187.
- [32] DE CAMPOS L M, FERNÁNDEZ-LUNA J M, GÁMEZ J A, et al. Ant colony optimization for learning Bayesian networks[J]. *International Journal of Approximate Reasoning*, 2002, 31(3): 291-311.
- [33] VAN LAARHOVEN P J M, AARTS E H L. Simulated annealing [M]//VAN LAARHOVEN P J M, AARTS E H L. Simulated annealing: theory and applications. Berlin: Springer, 1987: 7-15.
- [34] BOUCKAERT R R. Properties of Bayesian belief network learning algorithms[C]//Proceedings of the 10th Annual Conference on Uncertainty in Artificial Intelligence. Amsterdam: Elsevier, 1994: 102-109.
- [35] 李晓晴, 于海征. 贝叶斯网络结构学习的双重 K2 算法[J]. *科学技术与工程*, 2022, 22(24): 10602-10610.
- LI X Q, YU H Z. Double K2 algorithm for Bayesian network structure learning[J]. *Science Technology and Engineering*, 2022, 22(24): 10602-10610(in Chinese).
- [36] SELMAN B, GOMES C P. Hill-climbing search[J]. *Encyclopedia of Cognitive Science*, 2006, 81: 82.
- [37] GLOVER F, LAGUNA M. Tabu search[M]//DU D Z, PARDALOS P M. Handbook of combinatorial optimization. Berlin: Springer, 1998: 2093-229.
- [38] 刘大有, 王飞, 卢奕南, 等. 基于遗传算法的 Bayesian 网络结构学习研究[J]. *计算机研究与发展*, 2001, 38(8): 916-922.
- LIU D Y, WANG F, LU Y N, et al. Research on learning Bayesian network structure based on genetic algorithms[J]. *Journal of Computer Research and Development*, 2001, 38(8): 916-922(in Chinese).
- [39] CUI G, WONG M L, LUI H K. Machine learning for direct marketing response models: Bayesian networks with evolutionary programming[J]. *Management Science*, 2006, 52(4): 597-612.
- [40] GÁMEZ J A, PUERTA J M. Searching for the best elimination sequence in Bayesian networks by using ant colony optimization [J]. *Pattern Recognition Letters*, 2002, 23(1-3): 261-277.
- [41] GHEISARI S, MEYBODI M R. BNC-PSO: structure learning of Bayesian networks by particle swarm optimization[J]. *Information Sciences*, 2016, 348: 272-289.
- [42] SAHIN F, DEVASIA A. Distributed particle swarm optimization for structural Bayesian network learning[M]//CHAN F, TIWARI M. Swarm intelligence, focus on ant and particle swarm optimization. London: IntechOpen Limited, 2007: 505-532.
- [43] ASKARI M B A, AHSAAE M G. Bayesian network structure learning based on cuckoo search algorithm[C]//Proceedings of the 6th Iranian Joint Congress on Fuzzy and Intelligent Systems. Piscataway: IEEE Press, 2018: 127-130.
- [44] KAREEM S W, OKUR M C. Evaluation of Bayesian network structure learning using elephant swarm water search algorithm [M]//SHI C. Handbook of research on advancements of swarm intelligence algorithms for solving real-world problems. Hershey:

- Engineering Science Reference, 2020: 139-159.
- [45] WANG J Y, LIU S Y. Novel binary encoding water cycle algorithm for solving Bayesian network structures learning problem[J]. *Knowledge-Based Systems*, 2018, 150: 95-110.
- [46] JI J Z, WEI H K, LIU C N. An artificial bee colony algorithm for learning Bayesian networks[J]. *Soft Computing*, 2013, 17(6): 983-994.
- [47] YANG C C, JI J Z, LIU J M, et al. Structural learning of Bayesian networks by bacterial foraging optimization[J]. *International Journal of Approximate Reasoning*, 2016, 69: 147-167.
- [48] JI J Z, YANG C C, LIU J M, et al. A comparative study on swarm intelligence for structure learning of Bayesian networks[J]. *Soft Computing*, 2017, 21(22): 6713-6738.
- [49] 张亮, 章兢. 改进遗传优化的贝叶斯网络结构学习[J]. *计算机系统应用*, 2011, 20(9): 68-72.
- ZHANG L, ZHANG J. Structure learning of BN based on improved genetic algorithm[J]. *Computer Systems & Applications*, 2011, 20(9): 68-72(in Chinese).
- [50] 金焱, 胡云安, 张瑾, 等. K2 与模拟退火相结合的贝叶斯网络结构学习[J]. *东南大学学报 (自然科学版)*, 2012, 42(增刊 1): 82-86.
- JIN Y, HU Y A, ZHANG J, et al. Bayesian network structure learning based on K2 and simulated annealing[J]. *Journal of Southeast University (Natural Science Edition)*, 2012, 42(Sup 1): 82-86(in Chinese).
- [51] 潘成胜, 张斌, 吕亚娜, 等. 改进灰狼优化算法的 K-Means 文本聚类[J]. *计算机工程与应用*, 2021, 57(1): 188-193.
- PAN C S, ZHANG B, LYU Y N, et al. K-Means text clustering based on improved gray wolf optimization algorithm[J]. *Computer Engineering and Applications*, 2021, 57(1): 188-193(in Chinese).
- [52] WANG X C, REN H J, GUO X X. A novel discrete firefly algorithm for Bayesian network structure learning[J]. *Knowledge-Based Systems*, 2022, 242: 108426.
- [53] CONSTANTINOU A C, LIU Y, KITSON N K, et al. Effective and efficient structure learning with pruning and model averaging strategies[J]. *International Journal of Approximate Reasoning*, 2022, 151: 292-321.
- [54] CHOW C, LIU C. Approximating discrete probability distributions with dependence trees[J]. *IEEE Transactions on Information Theory*, 1968, 14(3): 462-467.
- [55] BOUCKAERT R R. Bayesian belief networks: from construction to inference[D]. Utrecht: Utrecht University, 1995.
- [56] HECKERMAN D, GEIGER D, CHICKERING D M. Learning Bayesian networks: the combination of knowledge and statistical data[C]//Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence. Amsterdam: Elsevier, 1994: 293-301.
- [57] LARRANAGA P, KUIJPERS C M H, MURGA R H, et al. Learning Bayesian network structures by searching for the best ordering with genetic algorithms[J]. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 1996, 26(4): 487-493.
- [58] DE CAMPOS L M, PUERTA J M. Stochastic local and distributed search algorithms for learning belief networks[C]//Proceedings of the III International Symposium on Adaptive Systems: Evolutionary Computation and Probabilistic Graphical Model. Berlin: Springer, 2001: 109-115.
- [59] CHICKERING D M. Optimal structure identification with greedy search[J]. *Journal of Machine Learning Research*, 2002, 3: 507-554.
- [60] DE CAMPOS L M, GÁMEZ J A, PUERTA J M. Learning Bayesian networks by ant colony optimisation: searching in two different spaces[J]. *Mathware & Soft Computing*, 2002, 9: 2-3.
- [61] SHETTY S, SONG M. Structure learning of Bayesian networks using a semantic genetic algorithm-based approach[C]//Proceedings of the 3rd International Conference on Information Technology: Research and Education. Piscataway: IEEE Press, 2005: 454-458.
- [62] CHAN L S H, CHU A M Y, SO M K P. A hybrid Markov chain Monte Carlo approach for structural learning in Bayesian networks based on variable blocking[EB/OL]. [2023-07-01]. <https://doi.org/10.1214/25-BA1521>.
- [63] SALEHPOUR A A, MIRMOBIN B, AFZALI-KUSHA A, et al. An energy efficient routing protocol for cluster-based wireless sensor networks using ant colony optimization[C]//Proceedings of the International Conference on Innovations in Information Technology. Piscataway: IEEE Press, 2008: 455-459.
- [64] ZHAO J, SUN J, XU W B, et al. Structure learning of Bayesian networks based on discrete binary quantum-behaved particle swarm optimization algorithm[C]//Proceedings of the 5th International Conference on Natural Computation. Piscataway: IEEE Press, 2009: 86-90.
- [65] JI J Z, ZHANG H X, HU R B, et al. A Bayesian network learning algorithm based on independence test and ant colony optimization[J]. *Acta Automatica Sinica*, 2009, 35(3): 281-288.
- [66] LI X L. A particle swarm optimization and immune theory-based algorithm for structure learning of Bayesian networks[J]. *International Journal of Database Theory and Application*, 2010, 3(2): 61-69.
- [67] HAUSER A, BÜHLMANN P. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs[J]. *Journal of Machine Learning Research*, 2012, 13(1): 2409-2464.
- [68] AOUAY S, JAMOSSI S, AYED Y B. Particle swarm optimization based method for Bayesian network structure learning[C]//Proceedings of the 5th International Conference on Modeling, Simulation and Applied Optimization. Piscataway: IEEE Press, 2013: 1-6.
- [69] RAMSEY J, GLYMOUR M, SANCHEZ-ROMERO R, et al. A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images[J]. *International Journal of Data Science and Analytics*, 2017, 3(2): 121-129.
- [70] SCANAGATTA M, CORANI G, ZAFFALON M. Improved local search in Bayesian networks structure learning[C]//Proceedings of the Advanced Methodologies for Bayesian Networks. [S. l.]: PM-LR, 2017: 45-56.
- [71] JOSE S, LOUIS S J, DASCALU S M, et al. Bayesian network structure learning using case-injected genetic algorithms[C]//Proceedings of the IEEE 32nd International Conference on Tools with Artificial Intelligence. Piscataway: IEEE Press, 2020: 572-579.

- [72] BEHJATI S, BEIGY H. Improved K2 algorithm for Bayesian network structure learning[J]. *Engineering Applications of Artificial Intelligence*, 2020, 91: 103617.
- [73] BERNAOLA N, MICHIELS M, LARRAÑAGA P, et al. Learning massive interpretable gene regulatory networks of the human brain by merging Bayesian networks[J]. *PLoS Computational Biology*, 2023, 19(12): e1011443.
- [74] SINGH M, VALTORTA M. Construction of Bayesian network structures from data: a brief survey and an efficient algorithm[J]. *International Journal of Approximate Reasoning*, 1995, 12(2): 111-131.
- [75] DASH D, DRUZDZEL M J. A hybrid anytime algorithm for the construction of causal models from sparse data[EB/OL]. (2013-01-23)[2023-07-01]. <https://arxiv.org/abs/1301.6689>.
- [76] DE CAMPOS L M, FERNÁNDEZ-LUNA J M, PUERTA J M. An iterated local search algorithm for learning Bayesian networks with restarts based on conditional independence tests[J]. *International Journal of Intelligent Systems*, 2003, 18(2): 221-235.
- [77] TSAMARDINOS I, BROWN L E, ALIFERIS C F. The max-min hill-climbing Bayesian network structure learning algorithm[J]. *Machine Learning*, 2006, 65(1): 31-78.
- [78] LIU H, ZHOU S G, LAM W, et al. A new hybrid method for learning Bayesian networks: separation and reunion[J]. *Knowledge-Based Systems*, 2017, 121: 185-197.
- [79] CASTELO R, ROVERATO A. A robust procedure for Gaussian graphical model search from microarray data with p larger than n [J]. *Journal of Machine Learning Research*, 2006, 7(12): 2621-2650.
- [80] NANDY P, HAUSER A, MAATHUIS M H. High-dimensional consistency in score-based and hybrid structure learning[J]. *The Annals of Statistics*, 2018, 46(6A): 3151-3183.
- [81] ZHAO J J, HO S S. Improving Bayesian network local structure learning via data-driven symmetry correction methods[J]. *International Journal of Approximate Reasoning*, 2019, 107: 101-121.
- [82] DAI J G, REN J, DU W C. Decomposition-based Bayesian network structure learning algorithm using local topology information[J]. *Knowledge-Based Systems*, 2020, 195: 105602.
- [83] SUN B D, ZHOU Y, WANG J J, et al. A new PC-PSO algorithm for Bayesian network structure learning with structure priors[J]. *Expert Systems with Applications*, 2021, 184: 115237.
- [84] 全兆景, 李金香, 乔征端. 一种具有结构先验的贝叶斯网络结构学习算法 [J]. *电子科技*, 2023, 36(11): 1-7.
TONG Z J, LI J X, QIAO Z R. A structural learning algorithm for Bayesian networks with structural prior[J]. *Electronic Science and Technology*, 2023, 36(11): 1-7(in Chinese).
- [85] RESSEGUIER N, GIORGI R, PAOLETTI X. Sensitivity analysis when data are missing not-at-random[J]. *Epidemiology*, 2011, 22(2): 282.
- [86] DEMPSTER A P, LAIRD N M, RUBIN D B. Maximum likelihood from incomplete data via the EM algorithm[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1977, 39(1): 1-22.
- [87] FRIEDMAN N. The Bayesian structural EM algorithm[EB/OL]. (2013-01-30)[2023-07-01]. <http://arxiv.org/abs/1301.7373>.
- [88] FRIEDMAN N. Learning belief networks in the presence of missing values and hidden variables[C]//Proceedings of the Fourteenth International Conference on Machine Learning. New York: ACM, 1997: 125-133.
- [89] WONG M L, GUO Y Y. Learning Bayesian networks from incomplete databases using a novel evolutionary algorithm[J]. *Decision Support Systems*, 2008, 45(2): 368-383.
- [90] SCANAGATTA M, CORANI G, ZAFFALON M, et al. Efficient learning of bounded-treewidth Bayesian networks from complete and incomplete data sets[J]. *International Journal of Approximate Reasoning*, 2018, 95: 152-166.
- [91] QIAN G Q, WU Y H, XU M. Multiple change-points detection by empirical Bayesian information criteria and Gibbs sampling induced stochastic search[J]. *Applied Mathematical Modelling*, 2019, 72: 202-216.
- [92] GELFAND A E. Gibbs sampling[J]. *Journal of the American Statistical Association*, 2000, 95(452): 1300.
- [93] BROOKS S. Markov chain Monte Carlo method and its application[J]. *Journal of the Royal Statistical Society: Series D (the Statistician)*, 1998, 47(1): 69-100.
- [94] BODEWES T, SCUTARI M. Learning Bayesian networks from incomplete data with the node-average likelihood[J]. *International Journal of Approximate Reasoning*, 2021, 138: 145-160.
- [95] LIU Y, CONSTANTINOU A C. Greedy structure learning from data that contain systematic missing values[J]. *Machine Learning*, 2022, 111(10): 3867-3896.
- [96] ELIDAN G, LOTNER N, FRIEDMAN N, et al. Discovering hidden variables: a structure-based approach[C]//Proceedings of the Advances in Neural Information Processing Systems. [S.l.]: NIPS, 2000.
- [97] ELIDAN G, FRIEDMAN N. Learning hidden variable networks: the information bottleneck approach[J]. *Journal of Machine Learning Research*, 2005, 6: 81-127.
- [98] SPIRITES P, GLYMOUR C N, SCHEINES R, et al. Causation, prediction, and search[M]. Cambridge: MIT Press, 2000.
- [99] COLOMBO D, MAATHUIS M H, KALISCH M, et al. Learning high-dimensional DAGs with latent and selection variables[C]//Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence. [S. l.]: DBLP, 2011: 850-850.
- [100] CLAASSEN T, MOOIJ J, HESKES T. Learning sparse causal models is not NP-hard[EB/OL]. (2013-09-26)[2023-07-01]. <http://arxiv.org/abs/1309.6824>.
- [101] SCHEINES R. An introduction to causal inference[EB/OL]. [2023-07-01]. https://www.academich.edu/3135379/An_introduction_to_causal_inference.
- [102] TRIANTAFILLOU S, TSAMARDINOS I. Constraint-based causal discovery from multiple interventions over overlapping variable sets[J]. *The Journal of Machine Learning Research*, 2015, 16(1): 2147-2205.
- [103] RAGHU V K, RAMSEY J D, MORRIS A, et al. Comparison of strategies for scalable causal discovery of latent variable models from mixed data[J]. *International Journal of Data Science and Analytics*, 2018, 6(1): 33-45.
- [104] QI Z W, YUE K, DUAN L, et al. Dynamic embeddings for efficient parameter learning of Bayesian network with multiple latent variables[J]. *Information Sciences*, 2022, 590: 198-216.
- [105] RAMSEY J D. Scaling up greedy causal search for continuous

- variables[EB/OL]. (2015-10-11)[2023-07-01]. <https://arxiv.org/abs/1507.07749v2>.
- [106] TRIANTAFILLOU S, TSAMARDINOS I. Score-based vs constraint-based causal learning in the presence of confounders[C]//Proceedings of the Uncertainty in Artificial Intelligence. New York:[s.n.], 2016: 59-67.
- [107] DRTON M, EICHLER M, RICHARDSON T S. Computing maximum likelihood estimates in recursive linear models with correlated errors[J]. *Journal of Machine Learning Research*, 2009, 10: 2329-2348.
- [108] OGARRIO J M, SPIRITES P, RAMSEY J. A hybrid causal search algorithm for latent variable models[C]//Proceedings of the Conference on probabilistic graphical models. [S. l.]: PMLR, 2016: 368-379.
- [109] JABBARI F, RAMSEY J, SPIRITES P, et al. Discovery of causal models that contain latent variables through Bayesian scoring of independence constraints[C]//Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2017: 142-157.
- [110] CHOBTHAM K, CONSTANTINOU A C. Bayesian network structure learning with causal effects in the presence of latent variables[C]//Proceedings of the International Conference on Probabilistic Graphical Models. [S. l.]: PMLR, 2020: 101-112.
- [111] BERNSTEIN D, SAEED B, SQUIRES C, et al. Ordering-based causal structure learning in the presence of latent variables[C]//Proceedings of the International Conference on Artificial Intelligence and Statistics. [S. l.]: PMLR, 2020: 4098-4108.
- [112] CHOBTHAM K, CONSTANTINOU A C, KITSON N K. Hybrid Bayesian network discovery with latent variables by scoring multiple interventions[J]. *Data Mining and Knowledge Discovery*, 2023, 37(1): 476-520.
- [113] WANG L, YANG H Y, ZHANG S W, et al. Intuitionistic fuzzy dynamic Bayesian network and its application to terminating situation assessment[J]. *Procedia Computer Science*, 2019, 154: 238-248.
- [114] 晏师励, 李德华. 基于动态贝叶斯网络的空战目标威胁等级评估[J]. *计算机与数字工程*, 2015, 43(12): 2150-2154.
- YAN S L, LI D H. Threat level assessment of the air combat target based on DBN[J]. *Computer & Digital Engineering*, 2015, 43(12): 2150-2154(in Chinese).
- [115] HUANG C Q, DONG K S, HUANG H Q, et al. Autonomous air combat maneuver decision using Bayesian inference and moving horizon optimization[J]. *Journal of Systems Engineering and Electronics*, 2018, 29(1): 86-97.
- [116] BOURS M J. Bayes' rule in diagnosis[J]. *Journal of Clinical Epidemiology*, 2021, 131: 158-160.
- [117] 刘桂芬, 孟海英, 张岩波. Bayes 线性混合效应模型多中心临床试验应用[J]. *中国卫生统计*, 2005, 22(4): 200-203.
- LIU G F, MENG H Y, ZHANG Y B. Bayes analysis of mixed model applied in a multi-center trials[J]. *Chinese Journal of Health Statistics*, 2005, 22(4): 200-203(in Chinese).
- [118] 周旭毓, 方积乾. Meta 分析中随机效应模型的 Gibbs 抽样及其应用[J]. *中国卫生统计*, 2002, 19(4): 204-207.
- ZHOU X Y, FANG J Q. Gibbs sampling in random effects model for meta-analysis with application[J]. *Chinese Journal of Health Statistics*, 2002, 19(4): 204-207(in Chinese).
- [119] WIENS J, WALLACE B C. Special issue on machine learning for health and medicine[J]. *Machine Learning*, 2016, 102(3): 305-307.
- [120] LIBBRECHT M W, NOBLE W S. Machine learning applications in genetics and genomics[J]. *Nature Reviews Genetics*, 2015, 16(6): 321-332.
- [121] ALIZADEHSANI R, ABDAR M, JALALI S M J, et al. Comparing the performance of feature selection algorithms for wart treatment selection[C]//Proceedings of the International Workshop on Future Technology. [S.l.: s.n.], 2019: 6-18.
- [122] PARMAR C, GROSSMANN P, RIETVELD D, et al. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer[J]. *Frontiers in Oncology*, 2015, 5: 272.
- [123] CHEN P, WU K Y, GHATTAS O. Bayesian inference of heterogeneous epidemic models: application to COVID-19 spread accounting for long-term care facilities[J]. *Computer Methods in Applied Mechanics and Engineering*, 2021, 385: 114020.
- [124] NASRHEIDARABADI N, HAKEMI L, KOLIVAND P, et al. Comparing performances of intelligent classifier algorithms for predicting type of pain in patients with spinal cord injury[J]. *Electronic Physician*, 2017, 9(7): 4847-4852.
- [125] YOUSEFI L, TUCKER A, AL-LUHAYBI M, et al. Predicting disease complications using a stepwise hidden variable approach for learning dynamic Bayesian networks[C]//Proceedings of the IEEE 31st International Symposium on Computer-Based Medical Systems. Piscataway: IEEE Press, 2018: 106-111.
- [126] KHADEMI M, NEDIALKOV N S. Probabilistic graphical models and deep belief networks for prognosis of breast cancer[C]//Proceedings of the IEEE 14th International Conference on Machine Learning and Applications. Piscataway: IEEE Press, 2015: 727-732.
- [127] ALJAWAD D A, ALQAHTANI E, AL-KUHAILI G, et al. Breast cancer surgery survivability prediction using Bayesian network and support vector machines[C]//Proceedings of the International Conference on Informatics, Health & Technology. Piscataway: IEEE Press, 2017: 1-6.
- [128] REFAI A, MEROUANI H F, AOURAS H. Maintenance of a Bayesian network: application using medical diagnosis[J]. *Evolving Systems*, 2016, 7(3): 187-196.
- [129] BANDYOPADHYAY S, WOLFSON J, VOCK D M, et al. Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data[J]. *Data Mining and Knowledge Discovery*, 2015, 29(4): 1033-1069.
- [130] BUKHANOV N, BALAKHONTCEVA M, KOVALCHUK S, et al. Multiscale modeling of comorbidity relations in hypertensive outpatients[J]. *Procedia Computer Science*, 2017, 121: 446-450.
- [131] SA-NGAMUANG C, HADDAWY P, LUVIRA V, et al. Accuracy of dengue clinical diagnosis with and without NS1 antigen rapid test: comparison between human and Bayesian network model decision[J]. *PLoS Neglected Tropical Diseases*, 2018, 12(6): e0006573.
- [132] SHI H Y, LUO G P, ZHENG H W, et al. Coupling the water-energy-food-ecology nexus into a Bayesian network for water resources analysis and management in the Syr Darya River Basin[J]. *Journal of Hydrology*, 2020, 581: 124387.

[133] HUI E, STAFFORD R, MATTHEWS I M, et al. Bayesian networks as a novel tool to enhance interpretability and predictive power of ecological models[J]. *Ecological Informatics*, 2022, 68: 101539.

[134] CHATRABGOUN O, HOSSEINIAN-FAR A, DANESHKHAH A. Constructing gene regulatory networks from microarray data using non-Gaussian pair-copula Bayesian networks[J]. *Journal of Bioinformatics and Computational Biology*, 2020, 18(4): 2050023.

[135] YU H Y, KHAN F, GARANIYA V. Modified independent component analysis and Bayesian network-based two-stage fault diagnosis of process operations[J]. *Industrial & Engineering Chemistry Research*, 2015, 54(10): 2724-2742.

[136] JIANG Q C, HUANG B, DING S X, et al. Bayesian fault diagnosis with asynchronous measurements and its application in networked distributed monitoring[J]. *IEEE Transactions on Industrial Electronics*, 2016, 63(10): 6316-6324.

[137] CHANG Y J, CHEN G M, WU X F, et al. Failure probability analysis for emergency disconnect of deepwater drilling riser using Bayesian network[J]. *Journal of Loss Prevention in the Process Industries*, 2018, 51: 42-53.

[138] WANG Z W, WANG Z W, GU X W, et al. Feature selection based on Bayesian network for chiller fault diagnosis from the perspective of field applications[J]. *Applied Thermal Engineering*, 2018, 129: 674-683.

[139] YEO C, BHANDARI J, ABBASSI R, et al. Dynamic risk analysis of offloading process in floating liquefied natural gas (FLNG) platform using Bayesian network[J]. *Journal of Loss Prevention in the Process Industries*, 2016, 41: 259-269.

A review of Bayesian network structure learning

MENG Guanglei^{1,2,*}, CONG Zelin^{1,2}, SONG Bin^{1,2}, LI Tingting², WANG Chenguang², ZHOU Mingzhe^{1,2}

(1. School of Automation, Shenyang Aerospace University, Shenyang 110136, China;
2. Aviation Science and Technology Key Laboratory of Air Combat System Technology, Shenyang 110136, China)

Abstract: Bayesian networks, as a tool combining probability theory and graph theory, have the ability to efficiently handle uncertain reasoning and data analysis, and are widely used in various fields to solve complex engineering problems. Furthermore, the model can be learned by combining prior knowledge and training samples, overcoming the limitations of establishing the model solely relying on expert knowledge. Based on this, the development history of Bayesian networks was reviewed. The proposed Bayesian network structure learning algorithms were classified and summarized from three aspects: constraint-based methods, rating-based methods, and hybrid search algorithms respectively, and the current research status of various algorithms was summarized and analyzed. Since the data in practical applications often have incompleteness, the research status of incomplete Bayesian network structure learning is explained from two dimensions: missing data processing and latent variable learning. The application of Bayesian networks in different fields is expounded and summarized, and the development trend of future research on Bayesian network structure learning algorithms is discussed.

Keywords: machine learning; artificial intelligence algorithm; Bayesian network; structure learning; hidden variable

Received: 2023-07-07; Accepted: 2023-08-04; Published Online: 2023-09-05 10:36

URL: link.cnki.net/urlid/11.2625.V.20230904.1958.001

Foundation items: National Natural Science Foundation of China (61973222); Liaoning Talent Program (XLYC2007144); Shenyang Natural Science Foundation (22-315-6-09)

* Corresponding author. E-mail: mengguanglei@yeah.net