

• 研究方法(Research Method) •

潜变量建模视角下属性层级结构的检验与估计*

毛秀珍¹ 黄林超¹ 徐丙慧² 郑巧玉³

(¹ 四川师范大学教育科学学院, 成都 610066) (² 光华学校, 四川 雅安 625000)

(³ 沙湾小学, 四川 乐山 614900)

摘要 属性层级结构(Attribute Hierarchical Structure, AHS)刻画了属性之间的先决关系,在测验设计、教学设计与评估、学习路径规划和个性化学习推荐方面具有重要意义。基于潜变量建模方式,结合假设检验、贝叶斯结构学习和参数估计视角,本文系统梳理了AHS的估计和检验方法,分析了各类方法的基本思想、内在联系及差异、优势与不足,总结了其发展脉络与趋势,并提出了实践中方法选择的参考建议。本研究进一步优化方法的关键要素、丰富实验条件、拓展应用场景,以及创新研究思路等方面提出了未来的研究方向。

关键词 潜变量建模,属性层级结构,假设检验,贝叶斯结构学习,参数估计

分类号 B841

1 引言

认知诊断评估(Cognitive Diagnostic Assessment, CDA)通过对特定领域中认知属性、项目与被试特征的关系建模,能够精准评估个体知识结构的优劣。与经典测验理论相比,CDA不仅关注学生“学得好不好”,更关注学生“学会了什么、尚未掌握什么、为什么没有掌握以及如何促进掌握”等关键问题。随着《深化新时代教育评价改革总体方案》和《进一步加强中小学日常考试管理的通知》等文件的相继颁布,强化过程评价、改进结果评价、优化考试结果的评价用途已成为教育评价改革的重要方向。在此背景下,CDA日益凸显了其在丰富评价信息、优化教学设计和促进个性化发展等方面的应用价值。目前,CDA在教育、心理以及医学等领域已得到了广泛的研究与应用(Lim et al., 2024; Tu et al., 2019; 詹沛达, 2022)。

属性及其层级结构共同描述了特定领域的认知元素及内在结构,是CDA的重要组成部分。其中,属性指测验所考察的、用以解释被试作答表

现的知识、技能或策略;属性层级结构(Attribute Hierarchical Structure, AHS)描述了属性之间的先决关系。Leighton等人(2004)指出,认知属性并非彼此独立,而是具有某种层级关系,并提出了线性、收敛、发散和无结构型4种层级结构。其中,无结构型可视为发散型中先决属性仅产生单层后继属性的特例,而收敛型兼具发散与收敛的特点。于是,Liu和Huggins-Manley(2016)在保留线性型结构的基础上,将发散型称为倒金字塔结构,收敛型命名为金字塔结构,而兼具收敛与发散特征的称为菱形(钻石)结构,形成了更清晰、互斥的分类体系。

特别地,每一种AHS均可从外部形状与内部结构两个方面进行描述。其中,外部形状通过有向无环图直观呈现属性之间的先决关系,而内部组织则规定了在某一层级结构下可能存在和不可能存在的属性掌握模式(Attribute Mastery Pattern, AMP)。表1展示了4种AHS的外部形状和内部结构。

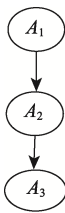
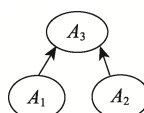
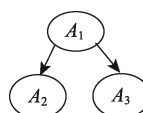
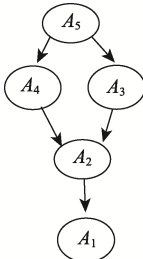
AHS在诊断性测验设计、教学设计与评估、学习路径规划和个性化学习推荐等方面都具有重要的指导意义。合理的AHS具有多重价值。理论上,AHS能够有效减少潜在类别的数量,提高诊断模型参数估计的精度与模型拟合度,并减少获

收稿日期: 2026-03-25

* 四川省教育科研2022年重大牵头项目(SCJG22A005)。

通信作者: 毛秀珍, E-mail: maomao_wanli@163.com

表 1 四种属性层级结构的外部形状和内部结构

线性		金字塔型		倒金字塔型		钻石型	
外部形状	内部结构	外部形状	内部结构	外部形状	内部结构	外部形状	内部结构
	(000)		(000)		(000)		(00000)
	(100)		(100)		(100)		(00001)
	(010)		(010)		(010)		(00011)
	(001)		(001)		(001)		(00101)
	(110)		(110)		(110)		(00111)
	(101)		(101)		(101)		(01111)
	(011)		(011)		(011)		(11111)
	(111)		(111)		(111)		

注：表中加粗的 AMP 代表可能存在的 AMP；钻石结构中因为 AMP 太多，图中仅显示可能的 AHS。

得稳定参数所需的样本量，特别是在高维属性与多题目测验情境中能显著提升计算效率(Tu et al., 2019)。实践中，AHS 通过约束教学路径空间与教学策略的搜索范围，不仅为差异化教学提供重要支撑，还能有效降低个性化智能辅导系统中生成可行教学决策的计算成本(Brunskill & Russell, 2010)。然而，错误的 AHS 可能带来更为隐蔽的风险。Liu (2018)的研究表明：错误的 AHS 会改变模型所允许的 AMP 空间，进而导致学生认知断点和后续补救路径产生系统性偏差。尤其是高层先决关系的错误往往通过层级传递引发级联性误判，增加诊断误差。Liu 还指出，模型拟合指标通常对 AHS 的错误指定并不敏感，即便在层级严重失配的情况下，模型仍可能呈现“拟合良好”的统计假象。大量研究都表明教育与心理领域中普遍存在 AHS (Leighton & Gierl, 2007; Templin & Bradshaw, 2014; Wang & Gierl, 2011)。因此，AHS 并非可忽略的附属设定，而是连接课程目标、测验结果与学习路径的关键环节，其合理性关系到认知诊断的科学性、评估的准确性与教学干预的有效性。

目前，主要有两种途径获得 AHS。一是认知驱动的质性分析方法，即通过分析领域知识与被试作答项目的认知过程，采用任务分析、口语报告和专家判断等方法获得 AHS。该方法虽然可解释性强，但容易受主观理解与判断的影响，可靠性不足；特别是在结构复杂的情境下，不仅费时费力，而且难以获得一致的结果。二是数据驱动的量化分析方法，即依据作答数据，运用假设检验、结构学习和参数估计等视角获得 AHS。该方

法较少受主观经验的影响，能够建构一般性分析框架，可重复性强且适用情境广；但其准确性往往受到样本规模、属性个数和层级结构等因素的影响。尽管如此，量化分析法备受研究者青睐，已得到广泛的关注(Templin & Bradshaw, 2014; Tu et al., 2019; Zhang et al., 2025a)。

现有 AHS 量化分析方法可以从多个角度进行分类：按照目的，可分为 AHS 的验证与估计；按照分析视角，可分为基于假设检验、贝叶斯结构学习和参数估计的方法；按照实现方式，可分为直接和间接分析方法；按照是否依赖模型，可以分为参数化和非参数化方法；按照分析的数据，可分为基于 Q 矩阵(Tatsuoka,1995)、作答数据和被试 AMP 的方法。特别地，梳理它们的建构思路可知，潜变量模型是参数化方法的出发点，决定了方法的目的、分析视角和实现方式。纵观其发展路径，又体现了对灵活建模和数据驱动方法的日益关注：在目的上，形成了由验证 AHS 到估计 AHS、联合估计属性个数与 AHS、再到联合估计 AHS、 Q 矩阵和项目参数；在模型基础上，由测量模型、结构模型和潜类别模型逐步向模型融合发展。

综上，本文将基于潜变量建模方式，结合假设检验、贝叶斯结构学习和参数估计视角系统梳理 AHS 的估计和检验方法，以期揭示方法的基本思想与应用场景，为方法的选用提供依据和新方法的构建提供思路。本文的结构如下：第一部分介绍认知诊断中潜变量建模的一般框架、测量模型和结构模型；第二部分分别针对 AHS 的检验和估计方法，介绍各类方法的理论依据、思路与过

程, 评析方法间的联系和区别、优势与不足; 最后进行总结与展望。

2 认知诊断评估中潜变量建模

认知诊断评估以潜变量建模为基本框架, 通常包括结构模型和测量模型。前者用于刻画潜在变量之间的依赖关系及其分布特征; 后者用于描述观测变量与潜在变量之间的映射机制。在这一框架中, 潜类别模型为刻画潜变量与观测变量的关系提供了一般形式; 诊断分类模型属于测量模型, 通过项目反应函数表征作答项目的认知过程; 而贝叶斯网络结构模型则以概率图方式对结构模型进行直观表达与结构化建模。

为统一表述, 约定符号如下: 用 I 、 J 和 M 分别表示被试、项目、属性和潜在类的数量; i 、 j 、 k 和 m 分别对应被试 i 、项目 j 、属性 k 和潜在类 m ; $\mathbf{q}_j$ 表示项目 j ($j=1, 2, \dots, J$) 考查的属性向量, $\boldsymbol{\alpha}$ 表示被试的知识状态, ε 表示属性层级结构; $\mathbf{X}_{I \times J}$ 和 x_{ij} 分别表示 I 个被试在 J 个项目上的作答反应矩阵和被试 i 在第 j 题上的作答反应。潜变量建模中包括项目参数与结构参数。项目参数用于刻画项目反应概率与潜类别的关系。例如, 潜类别模型中项目参数为每个潜类别正确作答项目 j 的概率向量 $\boldsymbol{\theta}_j = (\theta_{j1}, \theta_{j2}, \dots, \theta_{jM})$ (Ma et al., 2022); GDINA 模型中项目参数表示为各阶效应参数向量 $\boldsymbol{\delta}_j = (\delta_{j0}, \delta_{j1}, \dots, \delta_{j, kk'}, \dots, \delta_{j, (12 \dots K)})$ 。结构参数则用于描述潜类别的分布特征, 例如第 m 个潜类别占总体的比例 π_m (Zhang et al., 2025a)。

2.1 潜类别模型(Latent Class Model, LCM)

在教育测量中, 潜类别模型代表了被试的潜在类别与项目作答反应之间的关系。对属于潜在类别 m 的被试而言, 其正确作答项目 j 的条件概率表示为 θ_{jm} , 即 $P(x_{ij} = 1 | m) = \theta_{jm}$ 。于是, 给定反应矩阵 $\mathbf{X}_{I \times J}$ 、项目参数 $\boldsymbol{\theta}_{J \times M}$ 和结构参数 $\boldsymbol{\pi}$, 在局部独立性假设下, 全体被试反应的边际对数似然函数为:

$$\log L(\boldsymbol{\theta}, \boldsymbol{\pi} | \mathbf{X}) = \sum_{i=1}^I \log \left[\sum_{m=1}^M \pi_m \prod_{j=1}^J \theta_{jm}^{x_{ij}} (1 - \theta_{jm})^{1-x_{ij}} \right]. \quad (1)$$

从统计建模角度看, LCM 通过有限混合分解将总体划分为若干具有异质性作答概率的潜在类别以解释群体作答差异。该模型具有较强的灵活性, 参数估计方法完善, 无需预先设定 $\mathbf{Q}$ 矩阵和

AHS, 在探索性分析中具有一定优势。然而, LCM 本身是一种纯粹的统计分类模型, 并不预设潜类别所对应的认知结构, 项目参数 θ_{jm} 也不反映被试与项目特征之间的交互作用, 解释性有限。此外, 随着潜类别数量增加, 模型参数迅速扩张, 在样本量有限的情况下可能导致模型识别困难、估计不稳定或过拟合问题。于是, 研究者通常在 LCM 框架下引入结构性约束、设定 $\mathbf{Q}$ 矩阵、施加单调性假设等条件, 以提升模型的解释力与可操作性。

2.2 诊断分类模型(Diagnostic Classification Models, DCMs)

在 DCMs 中, 潜在类别常用 K 个属性的掌握状态来表征, 并用统计模型对被试知识状态、项目特征和作答反应的关系建模, 可解释性强, 是一类验证性模型。

熟知的 DCMs 包括 DINA 模型、DINO 模型、广义 DINA 模型(Generalized Deterministic Inputs, Noisy “And” gate model, GDINA)、对数线性认知诊断模型(Log-Linear Cognitive Diagnosis Model, LCDM)和层级诊断分类模型(Hierarchical Diagnostic Classification Model, HDCM) (de la Torre, 2011; Henson et al., 2009; Templin & Bradshaw, 2014)等等。其中, DINA 和 DINO 模型简单易懂, 但难以刻画复杂的认知过程; 相比而言, GDINA 和 LCDM 具有一般性, 应用更为广泛; HDCM 则在 LCDM 的基础上依据 AHS 对项目参数进行限制, 在反映 AHS 方面具有一定优势。总体而言, 不同 DCMs 在参数设定、假设条件、关系表达及适用情境等方面各具特点, 已形成了较为丰富的测量模型体系。

2.3 叶斯网络模型(Bayesian Network Models, BNMs)

一个完整的贝叶斯网络可以用二元组 $\text{BN} = (G; F)$ 表示。其中, $G = (V, E)$ 由随机变量 V 和有向弧集 E 组成, 并通过有向无环图(Directed Acyclic Graph, DAG)描述变量之间的关系; F 表示各个节点条件概率分布的参数集。贝叶斯网络模型作为一类结合概率论与图论的概率模型, 对 AHS 具有较好的表达力。例如, 图 1 所示的贝叶斯网络结构图展示了 4 个属性的层级结构, 以及它们与 4 个项目的关联。

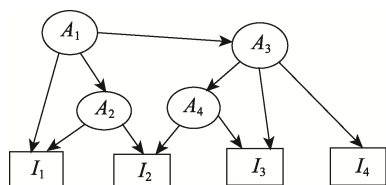


图 1 同时表征 AHS、项目与属性关系的贝叶斯网络结构图

根据局部马尔可夫性质,对于任意的属性节点 N_k ,在给定其父节点集合 $\text{Pa}(N_k)$ 的条件下, N_k 与网络中所有非子孙节点条件独立。于是,基于各节点的条件概率分布 $P(N_k|\text{Pa}(N_k))$,可将各属性掌握状态的联合概率分布分解为:

$$P(N_1, N_2, \dots, N_K) = \prod_{k=1}^K P(N_k | \text{Pa}(N_k))。 \quad (2)$$

其中, $P(N_k|\text{Pa}(N_k))$ 可以用不同方式定义。以三个存在线性结构的属性为例,假设 A_1 是 A_2 的先决属性, A_2 是 A_3 的先决属性。Hu 和 Templin (2020) 提出贝叶斯推理网络(Bayesian Inference Networks, BIN)允许被试在没有掌握父属性的情况下仍有一定概率掌握子属性。于是,结构参数可由 $P(A_1=1), P(A_2=1|A_1=1), P(A_2=1|A_1=0), P(A_3=1|A_2=1)$ 和 $P(A_3=1|A_2=0)$ 五个参数表示。Lee 和 Gu (2024) 提出潜在合取贝叶斯网络(Latent Conjunctive Bayesian Network, LCBN)则假设只有被试掌握了某个属性的所有父属性时,才有一定概率掌握该属性;若至少一个父属性未被掌握,则掌握该属性的概率为 0。这时,只需要 $P(A_1=1), P(A_2=1|A_1=1)$ 和 $P(A_3=1|A_2=1)$ 三个结构参数。近 20 年来,贝叶斯网络作为刻画复杂依赖关系的重要工具,已逐步应用于教育测量领域(Hu & Templin, 2020)。

认知诊断情景中,BNMs 是表征属性层级结构的概率图模型,而 DCMs 可视为 LCM 的具体化形式,三者并非相互独立,而是逐渐趋于融合。除 BIN、LCBN 外,BNMs 融入 CDA 的研究还比如:Zhan 等人(2020)与 Zhan 和 He (2021)运用序列掌握树思想建立 AHS 下条件概率分布参数集,分别提出序列高阶潜在结构模型和层级纵向诊断模型;Li 等人(2022)针对不同父节点数量构建了从父属性掌握状态到子属性掌握概率的贝叶斯网络传播机制,并借助机器学习算法提出基于贝叶斯网络的层级认知诊断框架。

总体上,AHS 研究的演进与潜变量模型的发展密切相关。在认知诊断研究初期,研究者通过模型拟合比较或结构参数的显著性检验考察预设 AHS 的合理性。与此同时,BNMs 在表达 AHS 方面的优势受到关注,贝叶斯结构学习算法被用于估计 AHS。进一步,由于 LCM 弱化了结构模型和测量模型的假设,研究逐渐转向 Q 矩阵及 AHS 的联合估计,通过估计结构参数来恢复 AHS。随着 BNMs 与 DCMs 不断融合,研究进一步向高维属性情况下联合估计 AHS 与模型参数推进。

3 属性层级结构的检验

现有检验 AHS 的思路主要有两种:一是,通过比较嵌套层级结构的模型拟合度来检验 AHS;二是,推断预设 AHS 下所有不可能属性掌握模式的结构参数是否显著为零来检验 AHS。

3.1 基于模型拟合比较的 AHS 检验方法

DCMs 框架下,检验 AHS 的一种常见方法是对预设层级结构 ε_0 和饱和层级结构 ε 的模型拟合度进行似然比检验(Likelihood Ratio Test, LRT)。令 Ω_0 和 Ω 分别表示 ε_0 和 ε 所对应的参数空间。该方法首先计算检验统计量:

$$LR = -2[\sup_{\omega \in \Omega_0} \log L(\omega) - \sup_{\omega' \in \Omega} \log L(\omega')]。 \quad (3)$$

然后,以两模型参数个数的差作为自由度进行渐近卡方检验。若检验结果不显著,则从模型简洁性考虑接受预设 AHS 的合理性。LRT 计算简单且易于实现,常用于检验潜变量模型对 AHS 的表征能力。例如,Templin 和 Bradshaw (2014)通过 LRT 证明,与 DINA 和 DINO 相比,HDCM 能够较好地检验 AHS。又如,Hu 和 Templin (2020)探讨了 HDCM、BIN 和饱和 DCMs 中结构参数的关系,然后通过 LRT 证明 BIN 在表征 AHS 方面优于 HDCM。

Ma 和 Xu (2022)给出了 AHS 可检验的概念和充分条件,并深入考察了 LRT 的性质。他们指出,当 DCMs 中部分参数的真值位于参数空间的边界时,LRT 缺乏正则性条件,导致检验统计量不再服从标准卡方分布。特别地,当项目数量较多或参数接近边界时,统计量向理论极限分布逼近的速度更慢,这导致 LRT 在有限样本量时的检验效能较差。于是,他们提出采用 Bootstrap 方法,通过重复抽取多个与原样本容量相同的样本获得 LRT 统计量的抽样分布,并根据原样本的 LRT 统

计量在该抽样分布中的概率进行统计推断。根据重复抽样的方式, 可以分为参数化和非参数化 Bootstrap LRT 检验方法: 前者基于原样本的参数估计值生成新样本; 后者从原样本中进行有放回抽样。Ma 和 Xu 在不同模型、样本量和项目质量的条件下, 系统比较了上述三种方法的表现。结果表明, 参数化 Bootstrap LRT 的表现最优, 非参数化 Bootstrap LRT 和传统 LRT 在不同实验条件下各有优劣。然而, Bootstrap LRT 方法依赖重复抽样, 其检验效能易受原样本质量的影响, 且计算成本也随样本量增大而增加。

3.2 基于结构参数检验的 AHS 检验方法

属性层级结构决定了可能的 AMP, 与 AHS 不一致的 AMP 对应的结构参数真值为 0。于是, Zhang 等人(2025a)提出对不合理 AMP 的结构参数进行 Wald 检验, 以验证预设 AHS 的合理性。该方法首先在饱和模型下估计项目参数和结构参数, 然后根据预设 AHS 确定需要检验的结构参数并构建约束矩阵 R , 最后利用结构参数估计值的协方差矩阵 $\hat{\Sigma}(\pi, \pi)$ 计算 Wald 检验统计量:

$$Wald = (R\hat{\pi})'(R\hat{\Sigma}(\pi, \pi)R')^{-1}(R\hat{\pi})。 \quad (4)$$

若检验结果不显著, 则表明尚无足够证据拒绝预设 AHS。他们进一步比较了基于经验交叉相乘信息矩阵(Empirical Cross-product Information Matrix, XPD)的 Wald 检验(Wald-XPD)、基于观察信息矩阵(Observed Information Matrix, Obs)的 Wald 检验(Wald-Obs), 以及 LRT 在验证 AHS 中的表现。结果表明, 就运行效率而言, Wald-XPD 方法优于 LRT 方法, Wald-Obs 方法最低; 就经验第 I 类错误率而言, Wald-XPD 在多数情形下表现较为稳健, Wald-Obs 在属性数量较多且项目质量中等时表现最好, 而 LRT 在项目质量较高时过于保守、在项目质量较低时又容易导致第 I 类错误膨胀, 整体稳定性相对较差。

3.3 属性层级结构检验方法的评价

LRT 方法往往因检验统计量的实际分布与渐近分布存在较大偏差, 影响其对第 I 类错误率的控制和检验功效; 参数化和非参数化 Bootstrap LRT 解决了 LRT 渐近分布非标准形式的问题, 但它们在大规模应用中计算成本较高。与 LRT 相比, Wald 检验计算更简单, 在大部分条件下能够更好地控制第 I 类错误率, 检验功效更高且小样本条件下结果更稳定(Zhang et al., 2025a)。然而, 上述

研究考察的属性数量均在 5 个以内, 尚未在更复杂条件下充分验证这些方法的准确性和稳定性。

总体上, LRT 和 Wald 方法考虑 AHS 的整体约束, 思路简单易懂, 但一次只能比较一对嵌套模型。特别地, 针对两种非嵌套关系的层级结构, Ma 和 Xu (2022)建议选择与饱和结构相比显著性概率更大的结构。另外, 当 LRT 和 Wald 检验验证预设 AHS 时, 可以进一步结合 Z 检验或单参数 Wald 检验逐个检验哪些结构参数显著大于 0, 以获得可能的 AMP, 也可通过逐一约束结构参数后比较模型拟合度的变化来探索 AHS。

4 属性层级结构的估计

早期 AHS 的研究主要在既定 DCMs 框架下检验预设结构的合理性。事实上, AHS 往往并非已知。与此同时, HDCM 和 BIN 的提出开始探讨 AHS 的表征问题。于是, AHS 研究逐渐从“验证 AHS”转向“从数据中估计 AHS”。至今, 已形成基于结构参数显著性检验、贝叶斯网络结构学习和参数估计三种估计 AHS 的思路。

4.1 基于结构参数检验的 AHS 估计方法

这类方法的基本思想是通过饱和模型的结构参数逐一进行显著性检验, 获得所有可能的 AMP 之后推导 AHS。例如, Liu 等人(2022)基于经验交叉相乘信息矩阵、观察信息矩阵和三明治协方差矩阵提出了估计结构参数标准误的方法, 并通过结构参数 Z 统计量 $Z = \frac{\hat{\pi}_m}{SE(\hat{\pi}_m)}$ 检验来获得

AMP, 进而估计 AHS。事实上, 不合理 AMP 的结构参数估计值往往接近于 0 甚至等于 0, 容易使得信息矩阵不稳定或奇异, 带来标准误估计偏差, 破坏渐近正态性假设, 从而影响 Z 检验的有效性。针对该问题, Zhang 等人(2025b)提出迭代 Z 检验方法。该方法从饱和模型参数估计出发, 通过“约束—重新估计—检验”的迭代过程, 逐步剔除不可能的 AMP, 直到所有结构参数估计值大于预设阈值且 Z 检验显著, 从而恢复 AHS。

Z 检验提供了简便的估计思路, 但其有效性受限于信息矩阵不稳定或奇异的问题。迭代 Z 检验通过在模型估计过程中逐步识别并约束接近 0 或不显著的结构参数, 有效改善了这一问题。模拟研究显示, 在大多数测验情景下, 其第 I 类错误率和统计功效均优于 Z 检验。但是, 它们通过多

次逐个检验结构参数,过程繁琐、效率不高,且需要考虑多重检验带来的第I类错误率问题。

4.2 基于贝叶斯网络结构学习的AHS估计方法

贝叶斯网络结构学习算法通常将估计的AMP作为观测数据,采用基于评分(score-based)或基于约束(constraint-based)的结构学习算法来估计AHS。K2算法(Cooper & Herskovits, 1992)、爬山算法(Hill Climbing, HC) (Scutari & Denis, 2014)和极大极小爬山算法(Max-Min Hill Climbing, MMHC) (Tsamardinos et al., 2006)是三种经典的贝叶斯网络结构学习算法。它们在初始结构、搜索策略和评分函数上均存在差异。K2算法根据预设属性间的拓扑排序,采用贪婪搜索策略,按顺序为每个节点添加父节点,通过最大化K2评分函数确定最优结构。HC算法将空图、随机网络、完全图或者专家指定的结构作为初始网络,通过加边、删边或反向操作生成备选网络结构,并依据评分函数(如BIC、BDeu)选择当前最优结构,迭代更新至收敛。而MMHC算法先利用极大极小父与子算法构建网络骨架以限制结构搜索空间,随后采用HC算法确定最终的网络结构。

目前,研究者已将K2算法、HC算法和MMHC算法用于估计AHS(马奕帆, 2021; 王军, 2012; 汪玲玲, 孙小坚, 2025; 喻晓锋, 2009; 喻晓锋等, 2021)。总体上,就不同AHS而言,三种算法对线性和发散型结构的判准率最高,对无结构和收敛型的表现较差;就属性条件而言,二分属性下的判准率优于多分属性的结果(马奕帆, 2021)。此外,喻晓锋等人 and 马奕帆一致发现样本量对K2算法影响不大,但评分阈值会影响其表现,且对不同AHS的影响不同。

此外,汪玲玲和孙小坚(2025)将条件独立性检验方法用于联合修正AHS和 Q 矩阵。其基本步骤为:首先根据作答数据、初始 Q 矩阵和AHS估计知识状态;然后从饱和贝叶斯网络模型出发检验各边强度,保留显著的边,得到初步修正的 Q 矩阵和AHS;接下来针对每一个被修正的元素,比较修正前后模型的拟合指标(如BIC、BDeu),保留拟合更优结构。循环这三步,直到模型拟合不再显著改善,从而获得 Q 矩阵和AHS。

结构学习算法建模直观,并依托较成熟的贝叶斯网络结构学习体系。总体上,K2算法最简单表现较好;MMHC和HC的表现相近但不及K2算

法。相比之下,条件独立性检验方法则通过“估计—检验—修正—择优”的迭代机制,考虑到AMP估计与AHS和 Q 矩阵的相互影响,能有效修正 Q 矩阵和AHS,但其稳定性仍受项目数量、项目质量和样本量的影响。这些方法都可以采用参数化或非参数方法估计AMP,并将其作为观测数据来分析AHS。类似地,颜玉枝(2022)基于排序理论提出,首先根据被试AMP计算属性相关强度矩阵(Attribute Correlation Intensity Matrix, ACIM),用于表征“未掌握前提属性却掌握后续属性的比例”;然后,基于ACIM计算层级指标矩阵用于表征属性所在的层级顺序;最后依据临界值判断不同层之间是单独成层还是合并层级,在此基础上恢复属性层级结构,简称为ACIM方法。ACIM方法可不依赖DCMs,属于非参数化方法。该方法不受样本量影响、计算简单,在小样本(50人)中同样表现较好(颜玉枝, 2022; Yan et al., 2025)。另外,Yan等人(2025)研究表明,ACIM方法在不同实验条件下对AHS的判准率和计算效率都显著优于MMHC方法。

4.3 基于参数估计的AHS估计方法

贝叶斯结构学习算法中AMP估计和AHS估计通常被视作两个相对独立的过程,这在一定程度上忽视了二者的内在关联。相比之下,参数估计方法主要基于惩罚边际极大似然估计(Marginal Maximum Likelihood Estimation, MMLE)和贝叶斯联合估计(Bayesian Joint Estimation)两种方法展开。前者在边际似然函数中加入惩罚项控制模型复杂度,并通过EM算法获得结构参数进而估计AHS;后者将AHS、结构参数、AMP、项目参数和项目反应同时纳入贝叶斯框架建立后验分布,并采用MCMC算法估计AHS。

4.3.1 基于惩罚MMLE的AHS估计方法

惩罚MMLE方法的思想是通过在目标函数中加入惩罚项以控制参数数量、避免过拟合和提高估计精度。注意到,DCMs中项目特征参数常与项目考察的属性相关,且一个项目通常不会考察所有属性。对于仅在项目未考察的属性上存在差异的潜类别而言,其在项目上具有相同的反应分布。于是,潜类别比例和项目参数成为边际极大似然函数中施加惩罚的主要对象。Wang和Lu(2021)、Ma等人(2022),以及Lee和Gu(2024)均运用该思想估计AHS。

首先, Wang 和 Lu (2021) 基于 LCM 框架采用正则化潜类别建模 (Regularized Latent Class Modeling, RLCM) 方法 (Chen et al., 2017) 估计结构参数, 分析 AMP 的偏序关系, 恢复 AHS。该方法在对数似然函数 $\log L(\Theta, \pi | X)$ 中对项目参数施加平滑裁剪绝对偏差惩罚 (Smoothly Clipped Absolute Deviation Penalty, SCAD): $-N \sum_{j=1}^J \sum_{m=1}^{M-1} P_{\gamma}^{\text{SCAD}}$

$(\theta_{j(m+1)} - \theta_{jm})$ 。其中, $P_{\gamma}^{\text{SCAD}}(\theta_{j(m+1)} - \theta_{jm})$ 将差异较小的 $\theta_{j(m+1)}$ 和 θ_{jm} 合并为一类, 从而减少不必要的项目参数。注意到 LCM 框架中 θ_{jm} 并未明确项目特征参数, 难以解释项目间的差异。于是, 他们提出在 LCDM 中运用潜变量选择 (Latent Variable Selection, LVS) 方法 (Chen, Liu, et al., 2015; Xu & Shang, 2018), 即在边际极大似然函数中将项目参数 $(\delta_{j0}, \delta_{j1}, \dots, \delta_{jK}, \delta_{j,12}, \dots, \delta_{j,(K-1)K}, \dots, \delta_{j,12 \dots K})$ 作为惩罚对象。具体而言, 通过对项目参数集 Δ 施加非零正则化约束 $\lambda \sum_{j=1}^J \sum_{l=1}^{2^K-1} I(\delta_{jl} \neq 0)$ 建构目标函数:

$$\operatorname{argmin}_{(\Delta, \pi)} [-\log L(\Delta, \pi | X) + \lambda \sum_{j=1}^J \sum_{l=1}^{2^K-1} I(\delta_{jl} \neq 0)], \quad (5)$$

进行参数估计进而探索 AHS。

其次, 借鉴 RLCM 和 LVS 方法对项目参数施加惩罚, Ma 等人 (2022) 基于 LCM 框架提出惩罚似然函数方法 (Penalized Likelihood Approach, PLA)。该方法不仅对项目参数施加分组截断套索惩罚 $J_{\tau, \rho N}(\theta_j)$ (Shen et al., 2012) 实现项目合并模式, 同时对结构参数施加对数型惩罚 $-\lambda_1 \log_{[\rho N]} \pi_m$ (Gu & Xu, 2019) 以使部分结构参数收缩为 0, 得到以下目标函数:

$$\operatorname{argmin}_{(\Theta, \pi)} [-\log L(\Theta, \pi | X) + \lambda_1 \sum_{m=1}^M \log_{[\rho N]} \pi_m + \lambda_2 \sum_{j=1}^J J_{\tau, \rho N}(\theta_j)], \quad (6)$$

通过估计参数恢复 AHS。

再次, 不同于 RLCM、LVS 和 PLA 采用“先正则化估计再恢复 AHS”的思路, Lee 和 Gu (2024) 基于 LCBN 提出了“先结构学习再估计参数”的两阶段分析方法。其中第一阶段在边际似然函数中对结构参数施加对数惩罚建构目标函数:

$$\operatorname{argmin}_{(\Delta, \pi)} \left[-\log L(\Delta, \pi | X) + \gamma \sum_{m=1}^M \log_{[\rho N]} \pi_m \right] \quad (7)$$

通过 EM 算法估计 π_m 并筛选显著 AMP 后恢复

AHS。第二阶段在固定 AHS 的条件下, 采用另一个 EM 算法估计 LCBN 的结构参数和项目参数。

上述 4 种方法除了惩罚项的构建方式不同之外, 还存在以下差异。第一, 在输入信息方面, 除了反应矩阵外, RLCM 还需输入属性个数 K ; PLA 仅需输入潜在类别数量的上界; 而 LVS 和 LCBN 要求给定 Q 矩阵, 且满足模型可识别性条件。第二, 在模型方面, RLCM 和 PLA 基于 LCM、LVS 基于 DCMs、而 LCBN 融合了贝叶斯网络和 DCMs; 第三, 在 AHS 恢复过程方面, RLCM、PLA 和 LCBN 基于结构参数估计潜类别的偏序矩阵, 进而恢复 AHS (详细步骤参见 Ma et al., 2022); LVS 通过设定结构参数阈值筛选可能的潜类别来恢复 AHS。第四, 在结果方面, 当测验考察 4 个属性时, LVS 对 AHS 的判准率均在 90% 及以上, RLCM 的判准率大多低于 10% (Wang & Lu, 2021); PLA 对潜类别数量和 AHS 恢复上的判准率相似, 绝大部分达到 95% 以上 (Ma et al., 2022); 而 LCBN 在测验考察 8 个属性、AHS 结构复杂且样本量较小 (500) 的情形下, 对 AHS 的判准率大都能达到 94% 以上 (Lee & Gu, 2024)。

总体上, RLCM 和 PLA 不对 LCM 施加约束, 具有一定的灵活性和通用性; LVS 指定测量模型具有更强的解释性, 其表现显著优于 RLCM; 与 RLCM 和 LVS 相比, PLA 在已知条件更少的情况下能同时估计属性数量、AHS 和 Q 矩阵, 且表现较好; LCBN 整合了结构模型和测量模型, 减少了结构参数数量, 能处理复杂结构与高维属性情况, 具有较好的应用潜力。此外, 它们都需要结合信息准则或模拟实验对惩罚项的权重进行调优, 调优过程费时费力, 且调优参数会影响方法的稳定性和结果的可重复性。特别地, Ma 等人 (2022) 发现调优参数对结果的影响比样本量的影响更大。因此, 如何设置调优参数、如何选择惩罚函数以及模型融合等都是值得深入研究的问题。

4.3.2 基于贝叶斯参数估计的 AHS 估计方法

Chen 和 Wang (2023) 与 Wang 等人 (2026) 均在 DCMs 框架下运用贝叶斯参数估计, 通过 MCMC 抽样估计 AHS。具体而言, Chen 和 Wang 采用一般贝叶估计方法 (General Bayesian Estimation Method, GBEM) 构建 AHS、结构参数、AMP、项目参数和作答反应的联合分布。其中, 被试 AMP、结构参数和项目参数由 Gibbs 采样更新; AHS 则通过传

递归简策略缩小 DAG 搜索空间,并借助 Metropolis-Hastings 算法构造随机游走进行后验搜索与更新。相比之下,Wang 等人提出新贝叶斯估计方法(Novel Bayesian Estimation Method, NBEM),从研究目的、算法实现和实验设计方面拓展了 Chen 和 Wang 的研究。首先,研究目的从“已知 $\mathbf{Q}$ 矩阵条件下估计 AHS”拓展到“联合估计 $\mathbf{Q}$ 矩阵与 AHS”;其次,算法实现上,他们引入 mini-batch 策略提高计算效率,并采用标签重排技术处理 MCMC 采样中 $\mathbf{Q}$ 矩阵、AHS 和属性标签可能发生转换的问题;再次,实验设计上,除了样本量、AHS 类型和测验长度外,Wang 等人还系统考察了 mini-batch 大小和 AHS 判定截断值对估计结果的影响。

Chen 和 Wang 的研究结果表明,GBEM 方法在已知 $\mathbf{Q}$ 矩阵的条件下能够较好地恢复真实 AHS 或其子图,并较准确地估计被试 AMP 和项目参数。具体来看,在 4 个属性条件下,AHS 的判准率在 20 题时超过 70%、在 40 题时可达 75%~89%;即使增加到 7 个属性、40 个测验项目,其判准率仍在 80%以上。Wang 等人研究发现,即使 $\mathbf{Q}$ 矩阵和 AHS 均未知,NBEM 方法仍能取得较好效果:多数条件下邻接矩阵和 $\mathbf{Q}$ 矩阵的判准率超过 80%,在 1000 人、40 题条件下可达 98%以上。总体上,这类方法能够刻画 AHS 中先修关系的不确定性,为结构解释提供更丰富的信息,但同时也存在参数多、模型复杂和计算负担重等问题。

4.4 AHS 估计方法的评价

AHS 的外部表征、内部结构与条件概率分布是 AHS 的灵魂。结构参数的假设检验、贝叶斯结构学习和参数估计是估计 AHS 的三种主要分析视角。它们在分析思路、结果评价方式和不同 AHS 类型的判准率差异反映了不同方法的特点。

首先,AHS 的分析思路不同。结构学习算法采用“先估计被试 AMP 再学习结构”的思路,但 AMP 与 AHS 相对独立,被试 AMP 的估计误差会影响后续 AHS 的学习,产生类似“插补偏差”的问题。假设检验和参数估计将结构参数和项目参数视作整体:前者依据结构参数估计值建构可能的 AMP 空间来恢复 AHS;后者根据 AHS、AMP、项目特征与作答反应的联合分布来估计 AHS。这两类方法同时处理 AMP、AHS 和项目参数的不确定性,但也面临参数多、收敛困难和计算成本高

等问题。实践中应结合样本量、题目质量、计算资源和结果需求进行方法选择。

其次,AHS 的评价方式不同。假设检验方法重在报告每一类 AMP 的一类错误控制率和统计检验力(Zhang et al., 2025b);贝叶斯结构学习算法则多报告 AHS 估计中平均多弧、少弧和逆向弧数量(喻晓锋等, 2021; 马奕帆, 2021);惩罚 MMLE 方法关注 AMP 中正确偏序的判准率(Ma et al., 2022);贝叶斯参数估计中除报告了 AHS 中各边的后验纳入概率、估计 AHS 属于真实结构子图的比例以及估计 AHS 包含错误边或错误方向的比例(Chen & Wang, 2023) 外,Wang 等人(2026)还报告了邻接矩阵和可达矩阵中真实直接边的检出率 and 不存在直接边的排除率。此外,Yan 等人(2025)还报告了 AHS 中两两属性间关系恢复的准确率、识别出来的属性关系比真实关系多了(漏了)多少的比率。这些指标除了反映方法特征外,还从不同侧面评价了 AHS 估计结果的错误类型和结果稳定性。实践中应选择多维评价指标,以全面分析结果优劣。

再次,对不同 AHS 类型的返真性结果不同。迭代 Z 检验中对线性结构的一类错误率更高;惩罚 MMLE 方法中,PLA 对无结构型的恢复略差;结构学习算法中,K2、HC 和 MMHC 对收敛型或无结构型的表现更差;贝叶斯参数估计中,NBEM 方法在线性结构上恢复能力稍弱。其原因可能是:Z 检验方法易受 AMP 的数量和结构参数标准误的影响;正则化参数估计方法易受 AMP 数量的影响;结构学习算法易受结构复杂度和搜索策略的影响;而 NBEM 方法则易受项目考察的属性模式和被试 AMP 分布均衡度的影响。当然,研究展现出的差异和原因还有待进一步比较分析。

值得关注的是,汪玲玲和孙小坚(2025)、Ma 等人(2022)、Wang 等人(2026)和 Lee 和 Gu (2024)均探讨了“ $\mathbf{Q}$ 矩阵和 AHS”的联合估计或修正问题。特别地, Lee 和 Gu 提出的 LCBN 基于 AHS 对结构参数建模,显著降低了结构参数数量,为处理高维属性提供一种可行的思路。其“先估计 AHS 再估计模型参数”的两阶段策略使结构学习与参数估计的先后关系更符合逻辑。另外,颜玉枝(2022)提出的 ACIM 方法是非参数化方法的代表,不受限于样本量且表现较好。总体上,这些研究都代表了 AHS 分析的发展方向。

5 结论和展望

属性层级结构是认知诊断理论与实践的重要组成部分。合理的 AHS 不仅有助于提高被试的分类准确性和参数估计精度, 也为制定有效的教学策略和个性化学习推荐提供重要依据。本文基于认知诊断潜变量建模方式系统梳理了 AHS 的检验和估计方法, 总结了其研究脉络和发展趋势。总体上, 围绕 AHS 的检验和估计问题, 现有研究逐步形成了三类分析思路。其一是显著性检验, 主要在 DCMs 框架下通过模型拟合检验或结构参数检验来验证或估计 AHS; 其二是贝叶斯结构学习, 主要在 BNMs 框架下借助贝叶斯网络结构学习算法来估计或修正 AHS; 其三是参数估计, 既包括在 LCM、DCMs 或融合 BNMs 与 DCMs 框架下的正则化 MMLE 参数估计方法, 也包括融合 BNMs 与 DCMs 框架下的贝叶斯参数估计方法估计 AHS。AHS 的研究呈现出以下发展脉络: 在目标上, 从验证 AHS 逐步拓展到估计 AHS、联合估计属性个数和 AHS, 以及联合估计 Q 矩阵与 AHS;

在内容上, 从描述 AHS 的外部形状转向刻画内部结构的概率关联; 在模型基础上, 则由 LCM、DCMs 或 BNMs 的单一潜变量建模框架逐渐走向融合建模。

实践中的一个关键的问题是: 如何依据先验信息选择合适的分析方法? 一般而言, 当对测验属性缺乏先验认知时, 可采用 PLA 方法同时估计 K 、潜类别特征及 AHS; 若属性个数已知, 可采用 RLCM 方法估计 AMP 并重建 AHS; 当对 AHS 有初步认识但尚不明确测量模型时, 可借助 $K2$ 、HC、MMHC 或 ACIM 方法; 当对 Q 矩阵及测量模型有一定先验信息时, 可以选用惩罚 MMLE、贝叶斯参数估计、 Z 检验或迭代 Z 检验等方法; 而在 Q 矩阵、AHS 及测量模型均有明确预设时, 可采用 LRT、Bootstrap 近似 LRT 或 Wald 检验来验证 AHS, 也可以运用条件独立性检验来联合修正 Q 矩阵和 AHS。图 2 展示了实践中方法选择的基本思路。未来研究可从优化方法的关键要素、丰富实验条件、拓展应用场景, 以及创新研究思路等方面继续推进。

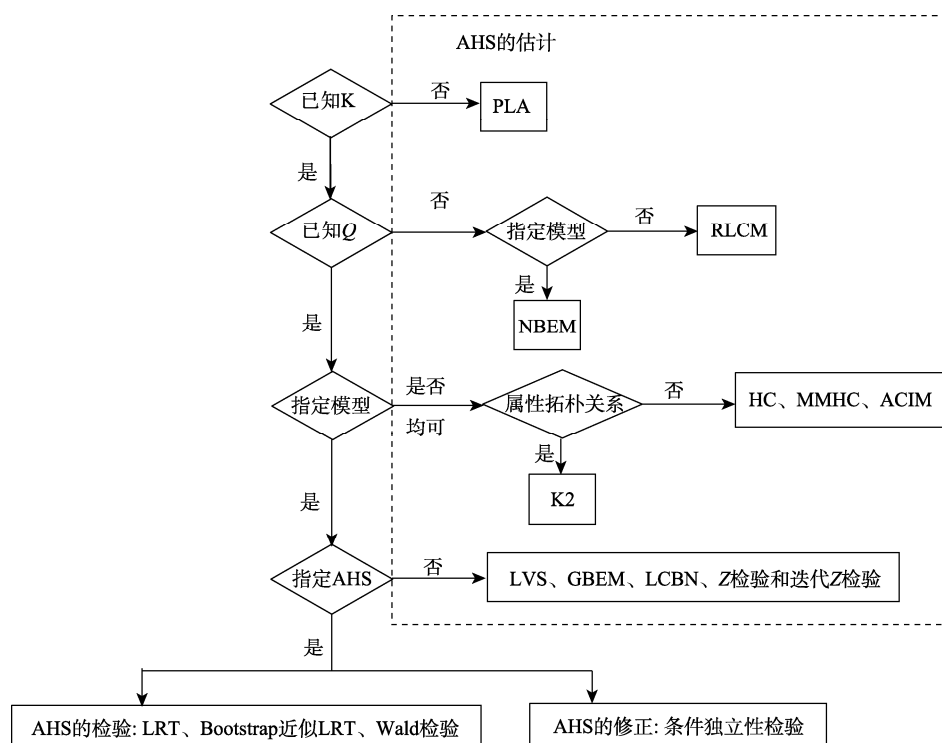


图 2 实践中选用 AHS 分析方法的路径图

5.1 重构方法的关键要素, 改进现有方法

从方法构成来看, 每一类方法都有一些相同的关键要素。例如, 显著性检验中的检验统计量、正则化参数估计中的惩罚项与调优参数、贝叶斯参数估计中的接受阈值、结构学习中的算法设计与评分函数等等。因此, 对它们进行重构或优化是改进现有方法和发展新方法的重要路径。

对于正则化 MMLE 参数估计中的惩罚项, 未来研究除了比较已有惩罚函数的表现外, 还可以进一步考察其他惩罚方式, 如尝试弹性网惩罚 (Zou & Hastie, 2005)、极小极大凹惩罚 (Zhang, 2010) 等。对于迭代 Z 检验、LVS 和 RLCM 中决定结构参数是否显著的阈值、贝叶斯结构学习算法中评分函数的阈值等, 未来研究可进一步探讨其设置原则, 或者参考 Wang 等人 (2026) 的方法, 系统考察阈值变化对实验结果的影响。对于贝叶斯结构学习算法, 未来研究可进一步考察基于连续优化的结构学习方法 (Zheng et al., 2018)、遗传算法、蚁群算法等结构学习算法的表现。

5.2 丰富实验条件, 拓展研究场景, 扩大应用领域

目前, 大部分研究局限在 5 个属性以内, 且样本量不低于 1000、项目特征单一。这无疑限制了方法的实践应用。事实上, 认知诊断中属性数量通常比较多 (Wang et al., 2022)。随着属性个数增加, 潜在 AMP 的数量呈指数级增长, 其规模甚至可能超过样本量, 高维参数给估计带来巨大挑战。因此, 探索高维属性情景下 AHS 的分析方法, 以及借鉴 ACIM 方法提出适合小样本的非参数分析方法, 均具有重要实践价值。

另外, 除马奕帆 (2021) 探讨了多分属性中贝叶斯网络结构学习算法的表现外, 其他研究都局限于二分属性的情况。因此, 未来研究应针对更多实际情况, 比如在多级评分项目、多分属性、多策略作答、纵向追踪、自适应测验等情景下探索 AHS 的分析方法。

5.3 拓展研究思路, 提出创新方法

不同方法在适用条件、计算效率和结果表现等方面各有优劣。首先, 结合已有方法的特点, 取长补短是提出新方法的重要途径。例如, 将正则化参数估计方法与假设检验方法相结合, 先估计结构参数, 再对参数进行显著性检验以确定可能的 AMP, 可避免采用人为主观标准。又如, 在正则化参数估计中结合贝叶斯网络来表征结构参数

也是一种有效的思路。

其次, 探索潜变量建模方法是提出 AHS 分析思路的重要途径。例如, 未来研究除了模型融合、探讨贝叶斯网络中条件概率的设置 (Hu & Templin, 2020; Zhan et al., 2020; Zhan & He, 2021; Li et al., 2022) 外, 还可以借鉴 HDCM 的思路, 探索基于 AHS 表征项目参数方法建构模型, 拓展分析 AHS 的建模思路。21 世纪初期, 越来越多研究开始引入机器学习算法来进行认知诊断 (Gierl et al., 2008; Li et al., 2024; Liu & Cheng, 2018; Tao et al., 2024; Thai-Nghe & Schmidt-Thieme, 2015; Wang et al., 2024), 也有研究从数据挖掘和机器学习角度探索 AHS 的估计 (Chen, Wuillemin, et al., 2015; Chen et al., 2016)。于是, 从机器学习算法视角拓展模型框架、开发 AHS 分析方法无疑是一种重要趋势。

参考文献

- 马奕帆. (2021). 基于三种结构学习算法的属性层级结构估计研究与应用 [硕士学位论文]. 江西师范大学.
- 王军. (2012). 贝叶斯网络在认知诊断中的应用 [硕士学位论文]. 南京师范大学.
- 汪玲玲, 孙小坚. (2025). 认知诊断模型属性层级关系和 Q 矩阵的联合验证方法: 面向实践的视角. *心理学报*, 57(7), 1295-1308.
- 颜玉枝. (2022). 排序理论在认知诊断中辅助标定属性层级关系——基于属性关联强度矩阵 [硕士学位论文]. 江西师范大学.
- 喻晓峰. (2009). 贝叶斯网在认知诊断中的应用 [硕士学位论文]. 江西师范大学.
- 喻晓峰, 马奕帆, 罗照盛, 秦春影. (2021). 基于 K2 算法的属性层级学习研究. *江西师范大学学报(自然科学版)*, 45(4), 376-383.
- 詹沛达. (2022). 引入眼动注视点的联合—交叉负载多模态认知诊断建模. *心理学报*, 54(11), 1416-1438.
- Brunskill, E., & Russell, S. (2010). *RAPID: A reachable anytime planner for imprecisely-sensed domains*. In Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI), Catalina Island, CA, USA.
- Chen, Y., González-Brenes, J. P., & Tian, J. (2016). Joint discovery of skill prerequisite graphs and student models. In *Proceedings of the 9th International Conference on Educational Data Mining (EDM)* (pp. 46-53). International Educational Data Mining Society.
- Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, 110(510), 850-866.
- Chen, Y., Li, X., Liu, J., & Ying, Z. (2017). Regularized latent class analysis with application in cognitive diagnosis. *Psychometrika*, 82(3), 660-692.

- Chen, Y., & Wang, S. (2023). Bayesian estimation of attribute hierarchy for cognitive diagnosis models. *Journal of Educational and Behavioral Statistics*, 48(6), 810–841.
- Chen, Y., Wuillemin, P.-H., & Labat, J.-M. (2015). Discovering prerequisite structure of skills through probabilistic association rules mining. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 117–124). International Educational Data Mining Society.
- Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4), 309–347.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179–199.
- Gierl, M. J., Cui, Y., & Hunka, S. (2008). Using connectionist models to evaluate examinees' response patterns to achievement tests. *Journal of Modern Applied Statistical Methods*, 7(1), 234–245.
- Gu, Y., & Xu, G. (2019). Learning attribute patterns in high-dimensional structured latent attribute models. *Journal of Machine Learning Research*, 20(115), 1–58.
- Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74(2), 191–210.
- Hu, B., & Templin, J. (2020). Using diagnostic classification models to validate attribute hierarchies and evaluate model fit in Bayesian networks. *Multivariate Behavioral Research*, 55(2), 300–311.
- Lee, S., & Gu, Y. (2024). Latent conjunctive Bayesian network: Unify attribute hierarchy and Bayesian network for cognitive diagnosis. *The Annals of Applied Statistics*, 18(3), 1988–2011.
- Leighton, J. P., & Gierl, M. J. (2007). Why cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 3–18). Cambridge University Press. <https://doi.org/10.1017/CBO9780511611186.001>
- Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy Method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, 41(3), 205–237.
- Li, J., Liu, Q., Wang, F., Liu, J., Huang, Z., Yao, F., ... Su, Y. (2024). Towards the identifiability and explainability for personalized learner modeling: An inductive paradigm. In *Proceedings of the ACM on Web Conference*. (pp. 3420–3431). Association for Computing Machinery.
- Li, J., Wang, F., Liu, Q., Zhu, M., Huang, W., Huang, Z., ... Wang, S. (2022). HierCDF: A Bayesian network-based hierarchical cognitive diagnosis framework. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 904–913). Association for Computing Machinery.
- Lim, Y. S., Willey, J. M., & Bangeranye, C. (2024). An exploration of cognitive diagnosis in medical education: Constructing comprehensive feedback for enhanced student learning. *Medical Science Educator*, 34(5), 1117–1122.
- Liu, C., & Cheng, Y. (2018). An application of the support vector machine for attribute-by-attribute classification in cognitive diagnosis. *Applied Psychological Measurement*, 42(1), 58–72.
- Liu, R. (2018). Misspecification of attribute structure in diagnostic measurement. *Educational and Psychological Measurement*, 78(4), 605–634.
- Liu, R., & Huggins-Manley, A. C. (2016). The specification of attribute structures and its effects on classification accuracy in diagnostic test design. In *Quantitative Psychology Research: The 80th Annual Meeting of the Psychometric Society, Beijing, 2015* (pp. 243–254). Springer International Publishing.
- Liu, Y., Xin, T., & Jiang, Y. (2022). Structural parameter standard error estimation method in diagnostic classification models: Estimation and application. *Multivariate Behavioral Research*, 57(5), 784–803.
- Ma, C., Ouyang, J., & Xu, G. (2022). Learning latent and hierarchical structures in cognitive diagnosis models. *Psychometrika*, 88(1), 175–207.
- Ma, C., & Xu, G. (2022). Hypothesis testing for hierarchical structures in cognitive diagnosis models. *Journal of Data Science*, 20(3), 279–302.
- Scutari, M., & Denis, J.-B. (2014). *Bayesian networks: With examples in R*. Chapman and Hall/CRC.
- Shen, X., Pan, W., & Zhu, Y. (2012). Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association*, 107(497), 223–232.
- Tao, J., Zhao, W., Liu, F., Guo, X., Cheng, N., Guo, Q., & Duan, H. (2024). Cognitive diagnosis method via Q -matrix-embedded neural networks. *Applied Sciences*, 14(22), 10380.
- Tatsuoka, K. K. (1995). Architecture of knowledge structure and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327–361). Lawrence Erlbaum Associates.
- Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79(2), 317–339.
- Thai-Nghe, N., & Schmidt-Thieme, L. (2015, October 8–10). Multi-relational factorization models for student modeling in intelligent tutoring systems. In *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)* (pp. 61–66), Hanoi, Vietnam.
- Tu, D., Wang, S., Cai, Y., Douglas, J., & Chang, H. H. (2019). Cognitive diagnostic models with attribute hierarchies: Model estimation with a restricted Q -matrix design. *Applied Psychological Measurement*, 43(4), 255–271.
- Tsamardinos, I., Brown, L. E., & Aliferis, C. F. (2006). The

- max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65(1), 31–78.
- Wang, X., Chen, Y., & Wang, S. (2026). Uncovering latent structures: A Bayesian approach to estimating Q -matrix and attribute hierarchies in cognitive diagnostic models. *Psychometrika*, 91(2), 669–694.
- Wang, C., & Gierl, M. J. (2011). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement*, 48(2), 165–187.
- Wang, F., Liu, Q., Chen, E., Huang, Z., Yin, Y., Wang, S., & Su, Y. (2024). NeuralCD: A general framework for cognitive diagnosis. *IEEE Transactions on Knowledge and Data Engineering*, 36(2), 850–864. <https://doi.org/10.1109/TKDE.2022.3201037>.
- Wang, J., Shi, N., Zhang, X., & Xu, G. (2022). Sequential gibbs sampling algorithm for cognitive diagnosis models with many attributes. *Multivariate Behavioral Research*, 57(5), 840–858.
- Wang, C., & Lu, J. (2021). Learning attribute hierarchies from data: Two exploratory approaches. *Journal of Educational and Behavioral Statistics*, 46(1), 58–84.
- Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*, 113(523), 1284–1295.
- Yan, Y., Dong, S., & Yu, X. (2025). Using ordering theory to learn attribute hierarchies from examinees' attribute profiles. *Journal of Educational and Behavioral Statistics*, 50(6), 985–1013.
- Zhan, P., & He, K. (2021). A longitudinal diagnostic model with hierarchical learning trajectories. *Educational Measurement: Issues and Practice*, 40(3), 18–30.
- Zhan, P., Ma, W., Jiao, H., & Ding, S. (2020). A sequential higher order latent structural model for hierarchical attributes in cognitive diagnostic assessments. *Applied Psychological Measurement*, 44(1), 65–83.
- Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, 38(2), 894–942.
- Zhang, X., Jiang, Y., Xin, T., & Liu, Y. (2025a). Validating attribute hierarchies in cognitive diagnosis models. *Frontiers in Psychology*, 16, 1562807. <https://doi.org/10.3389/fpsyg.2025.1562807>
- Zhang, X., Jiang, Y., Xin, T., & Liu, Y. (2025b). Iterative attribute hierarchy exploration methods for cognitive diagnosis models. *Journal of Educational and Behavioral Statistics*, 50(4), 682–713.
- Zheng, X., Aragam, B., Ravikumar, P. K., & Xing, E. P. (2018, December 3-8). DAGs with NO TEARS: Continuous optimization for structure learning. *Advances in Neural Information Processing Systems*, 31, Montreal, Canada.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2), 301–320.

Methods for testing and learning attribute hierarchical structures from a latent variable modeling perspective

MAO Xiuzhen¹, HUANG Linchao¹, XU Binghui², ZHENG Qiaoyu³

(¹ Institute of Educational Science, Sichuan Normal University, Chengdu 610066, China)

(² Guanghua School, Ya'an 625000, China) (³ Shawan Primary School, Leshan 614900, China)

Abstract: Attribute hierarchical structure (AHS) characterizes the prerequisite relationships among attributes. It plays an important role in test design, instructional design and evaluation, learning path planning, and personalized learning recommendations. From the perspective of latent variable modeling, this study systematically reviews methods for estimating and testing AHS according to three methodological perspectives: hypothesis testing, Bayesian structure learning and parameter estimation. Specifically, existing methods for analyzing attribute hierarchical structures are reviewed in terms of their basic ideas, interrelationships and distinctions, strengths and limitations. Then, this study summarizes the developmental trajectory and further trends of this area, and provides practical suggestions of method selection. Finally, future research directions are proposed from the aspects of optimizing key methodological components, enriching experimental conditions, expanding application scenarios, and developing innovative research perspectives.

Keywords: latent variable modeling, attribute hierarchical structure, hypothesis testing, Bayesian structure learning, parameter estimation