

文章编号: 1000-5862(2021)04-0376-08

基于 K2 算法的属性层级结构学习研究

喻晓锋¹, 马奕帆¹, 罗照盛¹, 秦春影²

(1. 江西师范大学心理学院, 江西 南昌 330022; 2. 南昌师范学院数学与计算机系, 江西 南昌 330032)

摘要: 诊断测验所考察的属性之间往往存在某种层级关系, 然而基于专家经验获得的属性层级关系易出现分歧或错误. 该文将属性掌握模式作为输入, 考察 K2 算法在不同阈值条件下学习得到属性层级结构的准确性. 模拟研究和实证数据分析的结果表明: K2 算法对属性层级结构的学习有较高的成功率, 并且 K2 算法对于 4 种基本层级结构有不同的敏感性, 其中线性型和发散型对阈值的敏感性较低, 而收敛型和无结构型对于阈值的敏感性较高.

关键词: 贝叶斯网络结构学习算法; 属性层级结构; K2 算法

中图分类号: B 841 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2021.04.09

0 引言

在心理和教育测量领域中, 认知诊断(Cognitive Diagnosis, CD)^[1-4]是一种对个体认知能力、知识状态诊断评估的过程. 以认知心理学和心理测量学作为理论基础, 认知诊断在多个领域中得到广泛的应用, 如教育领域、临床心理领域等.

认知诊断将认知理论与统计模型相结合, 推断出个体完成任务所需的知识、技能、策略等, 这些“知识”“技能”“策略”也称为认知属性. 在教育测量中, 可以将属性更一般地视为对知识点的理解、对细粒度能力的掌握情况等. 如在小学数学测验中, 属性可以对应加、减、乘、除的计算能力. J. P. Leighton 等^[5]认为认知属性之间不是相互独立的, 而是以类似网络的形式相互联系, 认知属性间存在一定的心理顺序、逻辑顺序或者层级关系, 被称为属性层级结构(Attribute Hierarchy, AH), 由此提出了属性层级方法(Attribute Hierarchy Method, AHM), 并用属性层级关系图来表征相关任务的认知模型. 但是, 属性间正确的层级关系难以准确定义, 原因主要有: (i) 使用的方法不恰当, 有学者^[6]发现 K. K. Tatsuoka^[7]

提出的由测验项目提取属性层级关系的方法是不可靠的; (ii) 各个学科的专家依凭自己的经验来确定属性层级结构的方法有可能存在偏差. 属性层级结构之间的关系定义的不准确会给诊断分类造成负面的影响, 得出错误的诊断结果, 导致诊断测验的信度和效度下降.

目前, 关于属性层级结构的研究多数是以属性层级结构作为已知条件来研究在不同的属性层级结构下诊断模型、估计算法、选题策略等的表现^[8-9]. 对于获取属性层级结构的研究相对较少, 有一些学者提出了估计属性层级结构的方法. 如 K. K. Tatsuoka^[7]提出行的逐对比较算法, 该算法可由测验 Q 阵得到属性层级结构. K. K. Tatsuoka^[7]认为, Q 矩阵行之间的包含关系与属性的先决关系等价. 若 2 个属性之间存在层级关系, 并且题目测试了某个属性, 则该题必定会包含它的先决属性. 于是她通过比较 Q 矩阵的任意的 2 行, 得到两两属性之间的先决关系, 从而导出所有属性之间的层级关系.

然而, 行的逐对比较算法在使用上有较大的限制, 即所使用 Q 阵必须是包含可达矩阵(R 阵)的 Q 矩阵^[6], 否则所估计出的属性层级结构会有较大的偏差. 对此, 丁树良等^[10]开发出了缩减算法. 所谓缩

收稿日期: 2020-09-21

基金项目: 国家自然科学基金(31660279, 31600909), 教育部人文社会科学研究(17YJC190029), 江西省高等学校教学改革研究(JXJG-19-2-43, JXJG-19-23-2), 江西省教育厅科学技术研究(GJJ191691, GJJ191128)和江西省教育厅人文社科课题(XL20202)资助项目.

作者简介: 喻晓锋(1980—), 男, 安徽太湖人, 副教授, 博士, 主要从事心理统计与测量方面的研究. E-mail: xyu6@jxnu.edu.cn

减算法是指 Q 矩阵的某一列若能够被 Q 矩阵中其他列的布尔和表示,则将这一列从 Q 矩阵中删除,重复这一步骤,直到 Q 矩阵的每一列都不能由其他列的布尔和表示为止。

前面提到的这2种算法都只是对 Q 矩阵本身进行了分析,忽视了在考生的作答信息中所蕴含的属性间的关系。实际上,属性间的依赖关系是蕴含在考生的作答数据中的。更准确地说,属性间的依赖关系(包括拓扑顺序)是蕴含在考生的属性掌握模式中的,因此考虑从考生的作答数据中获取属性层级关系值得探索。

J. Templin 等^[11]提出了层级诊断分类模型(Hierarchical Diagnostic Classification Model, HDCM)来确定属性层级结构。Xu Gongjun 等^[12]提出了潜在变量选择(Latent Variable Selection, LVS)的方法来估计 Q 矩阵。Wang Chun 等^[13]将 LVS 与 HDCM 相结合,认为属性层级结构会体现在 LCDM 的主效应参数 δ 上。Wang Chun 等^[13]考虑的另一种方法是正则化的潜在类别分析法。

Hu Bo 等^[14]根据 LCDM 和 HDCM 之间的关系构建了饱和的贝叶斯网络(用 K 表示测验考察的属性个数,饱和的网络结构包含 $2^K - 1$ 个参数)以及缩减的贝叶斯网络(存在属性层级结构)。使用似然比检验来比较饱和的贝叶斯网络和缩减的贝叶斯网络。若检验结果表明这2种模型并没有显著的区别,则所构建的属性层级结构就是合理的。

上述估计属性层级结构的方法大多是基于模型的方法,一次只能检验一种结构,效率较低。而行的逐对比较算法对 Q 阵的要求较高,RLCM 方法的正确率比较低。贝叶斯网络的优点在于知识或者能力的结构(即属性层级结构)可以直观地用有向无环图(Directed Acyclic Graph, DAG)来表示。更重要的是,贝叶斯网络可以通过结构学习算法使用数据驱动的方式从数据中得到属性层级结构^[15]。

基于以上分析,本文选择经典的贝叶斯网络结构学习算法——K2 算法,考虑从学生对属性的掌握数据中推导出属性间的层级关系,关注 K2 算法对不同评价标准(阈值)的敏感性,比较不同的属性层级结构在不同阈值下对于属性层级结构估计的准确性。一方面,为大规模开展诊断测验提供更实际可行的统计方法,另一方面,以此来减少人为判断所造成的主观性,并且在人为界定出现不一致时提供客观的补充证据。

1 理论背景

1.1 贝叶斯网络

贝叶斯网(Bayesian Network, BN)^[16]是一种表示因果推理的概率图模型,它由2个部分构成:(i)有向无环图,它表示随机变量(被称为节点)之间的关系;(ii)图结构的条件概率参数集,该参数集是由每个节点的条件概率分布(Conditional Probability Distribution, CPD)组成。其中有向无环图描述贝叶斯网络中每个节点之间的定性(依赖或独立)关系,条件概率分布描述贝叶斯网络节点之间的定量关系。本文的网络结构中的节点特指属性,在没有特别说明的情况下,本文的节点和属性所表示的含义相同。

在知识学习上,属性间的拓扑顺序是指它们在学习和教学过程中的先后顺序。根据是否需要给定节点间的拓扑顺序作为条件,可以将贝叶斯网络结构学习的算法分为2类,即需要拓扑顺序的结构学习算法和不需要拓扑顺序的结构学习算法。

1.2 需要拓扑顺序的贝叶斯网络结构学习算法

K2 算法^[17]属于考虑拓扑排序的结构学习算法。作为在搜索评分算法中的经典算法,K2 算法按照给定的拓扑顺序和评分函数并结合贪心搜索算法找到评分最优的网络结构。在确定评分函数和拓扑顺序之后,K2 算法的具体过程如下:

(i) 按照拓扑顺序,对于第 i 个节点(i 从 1 开始)根据选定的评分函数计算当前网络结构的得分,并在排在节点前面的节点集(记为 o_{rder_i})中寻找其父节点(拓扑顺序的第 1 个节点没有父节点);

(ii) 根据 (i) 从 o_{rder_i} 中选择可能的父节点结构,计算此时网络结构的评分,当评分函数达到最大时加入该节点,作为备选节点;

(iii) 若加入备选节点后网络结构的评分达到预先设置的评分阈值,则加入备选的节点就是其父节点,并加入当前网络结构中,否则不作为父节点,并结束搜索当前节点的父节点;

(iv) 重复 (i) ~ (iii),直到父节点的个数大于或等于设定的最大父节点个数;

(v) 搜索过程一直进行到拓扑顺序的最后一个节点。

在 K2 算法中的评分函数被称为 K2 评分函数,其计算公式为

$$g(i, \pi_i) = \sum_{j=1}^{q_i} (\log((r_i - 1)! / (N_{ij} + r_i - 1)!)) + \sum_{k=1}^{r_i} \log(N_{ijk}!),$$

其中 i 为节点的编号 j 为节点的取值编号 π_i 为节点 X_i 的父节点 q_i 为节点 π_i 的取值数量 r_i 为节点 X_i 的取值数量 N_{ijk} 为当 X_i 的父节点取第 j 个值、 X_i

取第 k 个值时的样本数量 $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

2 模拟研究

2.1 实验目的

本文研究不同属性层级结构对 K2 算法步骤 (iii) 中设置的评分阈值、不同题目质量、样本量、测验考察的属性个数与估计成功率之间的关系,设计如下的蒙特卡罗模拟研究.

2.2 实验设计

考虑可能影响属性层级结构学习效果的 5 个主要因素分别为样本量、属性个数、属性层级结构、题目质量以及评分阈值. 其中样本量有 3 个水平, 分别为 500 人、1 000 人、2 000 人; 属性个数有 4 个水平, 分别为 3、4、5、6; 属性层级结构有 4 个水平, 分别为线性型 (Linear)、收敛型 (Convergent)、发散型 (Divergent)、无结构型 (Unstructured), 其结构如图 1 所示 (图 1 只给出了 4 属性时的结构图). 基于 DINA 模型生成数据, 题目质量有 2 个水平 (高质量题目和低质量题目), 题目参数 (猜测参数 g 和失误参数 s) 服从均匀分布. 评分标准有 6 个水平, 由阈值来体现. 在 K2 算法中, 新结构的评分需要高于原有结构的评分, 备选节点才会作为当前节点的父节点加入. 由于新旧得分之间的差异往往不大, 并且不同属性层级结构对于评分标准的敏感度不同, 因此这里采用的评分标准为 $(g_{old} - g_{new}) / g_{old} > c_{rit}$, 其中 g_{new} 和 g_{old} 分别对应新旧结构的评分, c_{rit} 为阈值, 一共有 6 个水平 (0.01, 0.02, 0.03, 0.04, 0.05). c_{rit} 取 0 表示当新结构的得分高于原有结构的得分时以新结构作为当前的估计结果, 这是最宽松的条件. 阈值越大表明评分标准越严格.

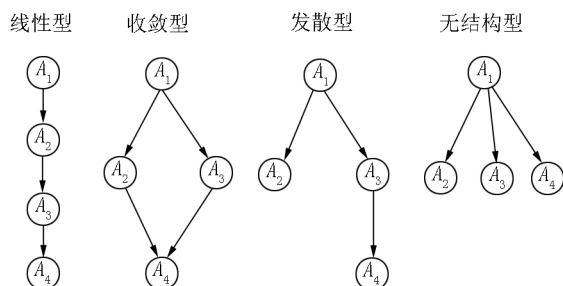


图 1 4 种属性层级结构

2.3 研究流程

模拟研究流程如图 2 所示.

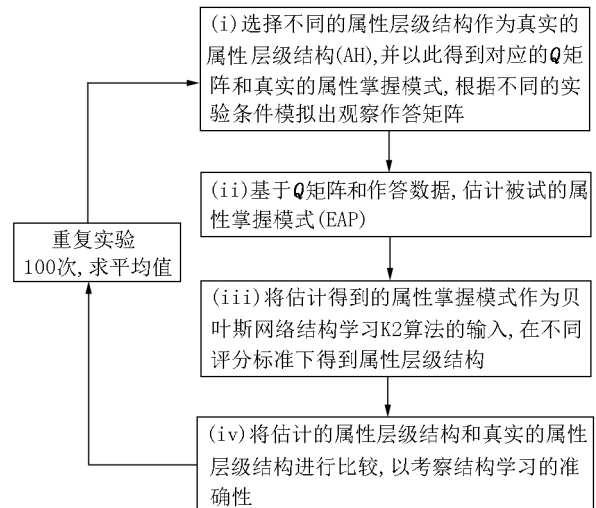


图 2 模拟研究流程

2.4 数据生成

数据生成分为 2 个步骤: (i) 模拟学生的作答数据. 模拟研究设置参见文献 [18], 即在测验考察 3 或 4 个属性时, 包含 20 道题目; 在测验考察 5 或 6 个属性时, 包含 30 道题目. 使用 R 语言来模拟学生作答数据, 0 表示答错这道题, 1 表示答对这道题. (ii) 基于学生作答数据和 Q 矩阵, 估计学生属性掌握模式, 1 表示掌握该属性, 0 表示没掌握该属性.

2.5 结构学习算法输入设置

学生的属性掌握模式为数据集 D , 是 K2 算法的输入数据. 其他输入参数分别设置如下: (i) 节点的拓扑排序; (ii) 每个节点的最大父节点数; (iii) 每个节点的取值个数为 2 (用于表示该属性是否掌握, 只能取 0 或 1).

2.6 评价指标

在此次模拟研究中使用了 5 个评价指标, 分别为正确率、平均多弧数量、平均少弧数量、平均逆向弧的数量以及模式判准率 $P_{CCR}^{[18]}$.

(i) 正确率 (Correct Recovery Rate, CRR):

$$C_{RR} = \frac{\sum_{c=1}^C s_{ucc_c}}{C},$$

其中 C 为重复实验的次数 (即 100 次), s_{ucc} 是每次实验学习得到属性层级结构和真实的属性层级结构是否完全相符, 完全相符为 1, 反之为 0.

(ii) 平均多弧数量 (Average Wrong Connection Edges, AWCE):

$$A_{WCE} = \frac{\sum_{c=1}^C W_{CE_c}}{C},$$

其中 W_{CE} 是每次实验学习得到属性层级结构相比较于真实的属性层级结构所多出弧的数量.

(iii) 平均少弧数量 (Average Missing Edges, AME):

$$A_{ME} = \sum_{c=1}^C M_{E_c} / C,$$

其中 M_E 是每次实验学习得到属性层级结构相比较于真实的属性层级结构所少学习到的弧的数量.

(iv) 平均逆向弧的数量 (Average Wrong Orientation Edges, AWOE):

$$A_{WOE} = \sum_{c=1}^C W_{OE_c} / C,$$

其中 W_{OE} 是每次实验学习得到属性层级结构相比较于真实的属性层级结构方向相反弧的数量.

在这 4 个评价指标中, C_{RR} 的取值区间是 $[0, 1]$, 值越大表明结果越好 (见表 1); 而剩下的 3 个指标的取值区间是 $[0, K]$, 值越小表明结果越好.

2.7 实验结果

由于本文结果所涉及的表和图较多,在不影响理解的情况下,正文只给出部分的表和图,余下的结果部分可以向笔者索取.由表 2 可知,从总体来看,当题目质量较低时,K2 算法在合适的评分标准的条件下有较高的成功率学习到正确的属性层级结构.当真实属性层级结构是线性型结构时,K2 算法在 6 种阈值下的学习结果都与真实的结构完全相符.发散型结构在 3 个属性和 4 个属性时的学习正确率达到 90% 以上.在属性较多的情况下,比如在 6 个属性的条件下,发散结构的正确率低于在属性较少时的正确率,但随着样本量的增多逐渐提高,并且随着阈值的升高,正确率也升高,这表明当属性较多时,发散型在更加严格的评分标准下估计正确率更高.收敛型的学习结果正确率总体低于线性型和发散型的正确率.属性越多,正确率越低,因为属性越多,属性层级结构就越复杂,而样本量对于其正确率的影响不大.阈值的大小对收敛型结果影响较大,阈值越大,正确率越高,即评分标准越严格,估计的结构更接近真实的结构.无结构型在 4 个属性的条件下基本正确,在 5~6 个属性条件下,正确率整体偏低.阈值的大小对无结构型结果影响较大,随着阈值升高,正确率先升高后降低,在阈值为 0.01 和 0.02 时,正确率达到峰值.

表 1 基于 K2 算法结构学习的 C_{RR}

属性个数	人数	阈值	线性型	收敛型	发散型	无结构型
3	500	0	1.00	0.13	0.98	/
		0.01	1.00	0.50	1.00	/
		0.02	1.00	0.84	1.00	/
		0.03	1.00	0.92	1.00	/
		0.04	1.00	0.98	1.00	/
		0.05	1.00	1.00	1.00	/
	1 000	0	1.00	0.01	0.99	/
		0.01	1.00	0.32	1.00	/
		0.02	1.00	0.73	1.00	/
		0.03	1.00	0.97	1.00	/
		0.04	1.00	1.00	1.00	/
4	2 000	0	1.00	0	0.99	/
		0.01	1.00	0.05	1.00	/
		0.02	1.00	0.71	1.00	/
		0.03	1.00	0.97	1.00	/
		0.04	1.00	1.00	1.00	/
	500	0	1.00	0.25	0.97	0.94
		0.01	1.00	0.48	0.98	0.98
		0.02	1.00	0.68	1.00	1.00
		0.03	1.00	0.81	1.00	1.00
		0.04	1.00	0.93	1.00	1.00
4	1 000	0	1.00	0.05	0.91	0.95
		0.01	1.00	0.33	1.00	1.00
		0.02	1.00	0.61	1.00	1.00
		0.03	1.00	0.80	1.00	1.00
		0.04	1.00	0.92	1.00	1.00
	2 000	0	1.00	0	0.92	0.92
		0.01	1.00	0.15	1.00	1.00
		0.02	1.00	0.47	1.00	1.00
		0.03	1.00	0.74	1.00	1.00
		0.04	1.00	0.92	1.00	1.00

注: 测验长度 20, 属性个数为 3 和 4, 题目质量较低.

由表 2 可知, K2 算法在高质量题目的条件下学习到正确的属性层级结构的可能性更高, 其正确率整体略高于在低质量题目的条件下的正确率, 并且在不同条件下正确率的特征变化也基本一致. 当真实属性层级结构是线性型结构时, K2 算法的学习结果与真实的结构完全相符; 发散型结构的学习结果也基本正确. 同样, 属性越多, 正确率越低 (见图 3).

样本量对于收敛型和无结构型的正确率的影响不大, 但是对于在 6 个属性条件下的发散型, 样本量

越大,其正确率越高. 阈值越大,收敛型和发散型的正确率越高,即评分标准越严格,估计的结构更接近真实的结构. 同样,阈值的大小对无结构型结果影响较大. 随着阈值升高,正确率先升高后降低,当阈值为0.01和0.02时,正确率达到峰值(见图4). 但是当属性较多时,高质量题目的正确率高于低质量题目的正确率.

表2 基于 K2 算法结构学习的 C_{RR}

属性个数	人数	阈值	线性型	收敛型	发散型	无结构型
3	500	0	1.00	0.06	0.98	/
		0.01	1.00	0.50	1.00	/
		0.02	1.00	0.81	1.00	/
		0.03	1.00	0.91	1.00	/
		0.04	1.00	0.99	1.00	/
	1 000	0	1.00	0.01	0.99	/
		0.01	1.00	0.25	1.00	/
		0.02	1.00	0.68	1.00	/
		0.03	1.00	0.93	1.00	/
		0.04	1.00	0.99	1.00	/
	2 000	0	1.00	0	1.00	/
		0.01	1.00	0.07	1.00	/
		0.02	1.00	0.69	1.00	/
		0.03	1.00	1.00	1.00	/
		0.04	1.00	1.00	1.00	/
4	500	0	1.00	0.27	1.00	0.98
		0.01	1.00	0.58	1.00	1.00
		0.02	1.00	0.81	1.00	1.00
		0.03	1.00	0.92	1.00	1.00
		0.04	1.00	0.98	1.00	1.00
	1 000	0	1.00	0.05	0.99	0.97
		0.01	1.00	0.40	1.00	1.00
		0.02	1.00	0.71	1.00	1.00
		0.03	1.00	0.97	1.00	1.00
		0.04	1.00	1.00	1.00	1.00
	2 000	0	1.00	0	1.00	0.96
		0.01	1.00	0.13	1.00	1.00
		0.02	1.00	0.71	1.00	1.00
		0.03	1.00	0.98	1.00	1.00
		0.04	1.00	1.00	1.00	1.00

注: 测验长度为20,属性个数为3和4,题目质量较高.

由表3可知,K2算法学习所得到的5个属性的收敛型的属性层级结构相较于真实的属性层级结构

有更多的弧,阈值越低,其评分标准越宽松,平均弧数量越多. 因为评分标准越宽松,备选节点越容易被识别为父节点,所以易出现更多的弧(见图5(a)). K2算法学习所得到的5个属性的无结构型的属性层级结构相较于真实的属性层级结构通常会出现少弧,阈值越高,评分标准越严格,平均弧数量越少. 因为评分标准越严格,备选节点越不容易被识别为父节点,所以易出现少弧(见图5(b)).

表3 在高质量题目条件下收敛型和无结构型的错误类型

属性个数	AH	人数	阈值	A_{WCE}	A_{ME}	A_{WOE}
5	收敛型	500	0	0.64	0	0
			0.01	0.40	0	0
			0.02	0.23	0	0
			0.03	0.08	0	0
			0.04	0.04	0	0
		1 000	0.05	0.01	0	0
			0	0.94	0	0
			0.01	0.67	0	0
			0.02	0.22	0	0
			0.03	0.06	0	0
	无结构型	2 000	0.04	0	0	0
			0.05	0	0	0
			0	1.00	0	0
			0.01	0.87	0	0
			0.02	0.24	0	0
		500	0.03	0.02	0	0
			0.04	0.01	0	0
			0.05	0	0	0
			0	0.09	0	0
			0.01	0.01	0	0
	收敛型	1 000	0.02	0	0.06	0
			0.03	0	0.18	0
			0.04	0	0.77	0
			0.05	0	2.09	0
		2 000	0	0.13	0	0
			0.01	0	0	0
			0.02	0	0	0
			0.03	0	0.04	0
			0.04	0	0.38	0
			0.05	0	1.67	0
	无结构型	500	0	0.26	0	0
			0.01	0	0	0
			0.02	0	0	0
			0.03	0	0	0
			0.04	0	0.21	0
		1 000	0.05	0	1.51	0

注: AH 表示属性层级结构(Attribute Hierarchy).

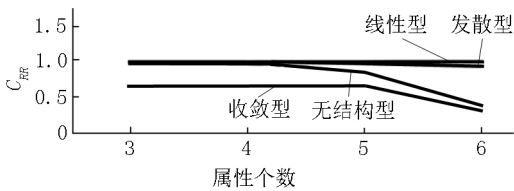


图 3 属性个数对不同属性层级结构 CRR 的影响

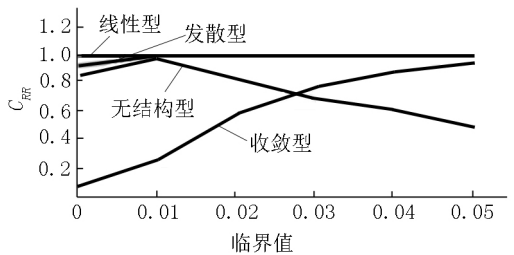


图 4 阈值对不同属性层级结构 CRR 的影响

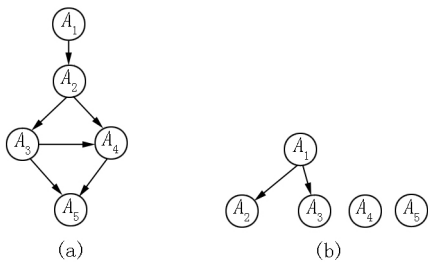


图 5 收敛型、无结构型错误类型举例

由此可以看出, K2 算法对于属性层级结构的估计有较高的正确率, 其中线性型完全正确, 发散型基本正确, 收敛型和无结构型对于阈值有不同的敏感度, 收敛型在高阈值的条件下正确率高, 无结构型在低阈值的条件下正确率高, 且这 2 种结构属性越多, 正确率降低越明显.

3 实证数据研究

为了进一步探究贝叶斯网络结构学习算法在实证数据中的应用, 并比较 3 种结构学习算法的效果, 本文拟使用结构学习算法对一批实证数据进行属性层级结构的估计.

3.1 数据描述

本研究选用图宾根大学的一项教学实验第 1 部分的 504 份作答数据^[19-20]. 参与者需要回答 12 道基础的概率论题目. 例如, “一个箱子中装有 30 个弹珠, 其中, 红色 8 颗, 黑色 10 颗, 黄色 12 颗, 那么随机拿一个弹珠, 它是黄色的概率是多少?”. 该测验一共测量了 4 个属性: 一个事件的经典概率(A_1), 对立事件发生的概率(A_2)、互斥事件发生概率之和(A_3)、相互独立事件发生的概率(A_4). 测验的 Q 矩

阵设定如表 4 所示, 前 4 道题只测量了单个属性, 第 5 ~ 10 题均测量了 2 个属性, 第 11 题和第 12 题测量了 3 个属性.

表 4 实证数据测验 Q 阵

题目	属性											
	1	2	3	4	5	6	7	8	9	10	11	12
A_1	1	0	0	0	1	1	1	1	1	0	1	1
A_2	0	1	0	0	1	1	0	0	0	1	1	0
A_3	0	0	1	0	0	0	1	1	0	0	0	1
A_4	0	0	0	1	0	0	0	0	1	1	1	1

3.2 研究结果与分析

依据测验 Q 矩阵, 利用 DINA 模型对作答数据进行参数估计, 得到参与者的属性掌握模式. 将估计的属性掌握模式作为 K2 算法的输入, 对其拓扑排序的结果为 A_1 、 A_2 、 A_3 、 A_4 .

K2 算法得到属性层级结构如图 6 所示, 在严格评分标准下(阈值为 0.05, 0.04, 0.03) 结果如图 6(a) 所示, 在宽松评分标准下(阈值为 0.02, 0.01, 0) 结果如图 6(b) 所示. A_1 是其他属性的父属性, 这因为 A_1 是“一个事件的经典概率”, 是掌握其他拓展概率论概念的基础. 而 A_2 是 A_3 的父属性, 因为 A_2 是“对立事件发生的概率”, A_3 是“互斥事件发生概率之和”, 对立事件是互斥事件的一种特殊情况, 所以 A_2 是 A_3 的父属性也符合逻辑. 多了 A_2 到 A_4 的弧, A_4 是“相互独立事件发生的概率”, 与“对立事件发生的概率”这一属性关系并不大. 在严格评分标准下的结果更符合逻辑. 并且对照模拟研究的结果, 4 个属性, 500 左右样本, 不同评分标准结果有差异. 因为可能真实结构更符合收敛型, 所以在严格评分标准下的结果可能更接近真实属性层级结构.

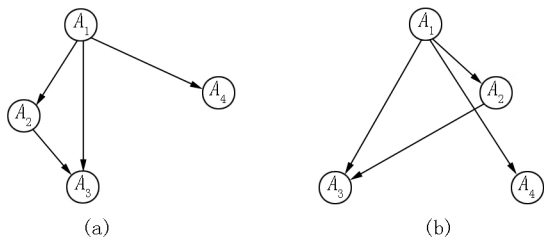


图 6 K2 算法的结果

4 总结与讨论

认知诊断测验的属性之间往往可能存在层级结

构,一方面在实际的测验中,这些属性层级结构往往依据专家经验获得,容易出现错误或意见不一致的情况;另一方面,在学生的属性掌握模式中蕴含了属性间的层级结构.因此,本研究系统考察了基于数据驱动的贝叶斯网络结构学习方法获取属性层级结构,为获取属性层级结构提供了一种补充方法和新的视角.

模拟研究结果表明:K2 算法整体上对属性层级结构的学习有较高的准确率.对线性型结构可以完全准确估计,对发散型结构的估计有较高的正确率.因为收敛型和无结构型本身结构的复杂性,属性个数越多,可能的属性层级结构越多,其估计的正确率越低.题目质量对于正确率有略微的影响,题目质量越好,正确率越高.样本量对正确率并没有较大的影响.收敛型和无结构型对阈值有不一样的敏感度,阈值越高,其评分标准越严格,收敛型正确率越高;而无结构型在阈值越高和在评分标准更严格的条件下正确率更高.而对实证数据的分析表明:在严格的评分标准下得到的属性层级结构更加符合逻辑,也符合模拟研究的结果.

本研究虽然取得了一些有意义的结果,但也存在一些不足和一些值得进一步深入研究的地方:首先,获得属性间的拓扑排序是 K2 算法的前提条件,而在有些情况下,拓扑排序并不容易获得,因此未来的研究可以考虑使用不需要拓扑排序作为前提的结构学习算法.另外,贝叶斯网络结构学习算法主要分为 3 种:基于搜索评分的算法、基于条件独立性检测的算法、混合算法.本研究使用的 K2 算法属于基于搜索评分的算法.在贝叶斯网络结构学习算法中,还有其他优秀的结构学习算法,如 HC 算法、MMHC 算法等.在未来的研究中,也可以考虑使用其他算法学习属性层级结构.其次,在本研究中只考虑了 4 种属性层级结构,在未来的研究中可以考虑学习离散型、混合型等属性层级结构.最后,模拟研究的实验条件考虑了属性个数、题目质量、样本量、阈值,未来可以增加实验条件,如考虑多分属性,进一步探索贝叶斯网络结构学习算法对属性层级结构估计的效果.

5 结论

K2 算法对于属性层级结构的估计有较高的正

确率,其中线性型完全正确,发散型基本正确,收敛型和无结构型对阈值有不同的敏感度,收敛型在高阈值的条件下正确率高,无结构型在低阈值的条件下正确率高,而且这 2 种结构属性越多,其正确率降低越明显.

6 参考文献

- [1] 罗照盛. 认知诊断评价理论基础 [M]. 北京: 北京师范大学出版社, 2019.
- [2] Leighton J, Gierl M. Cognitive diagnostic assessment for education: theory and applications [M]. Cambridge: Cambridge University Press, 2007.
- [3] Templin J, Henson R A. Diagnostic measurement: theory, methods and applications [M]. Guilford: Guilford Press, 2010.
- [4] von Davier M, Lee Y S. Handbook of diagnostic classification models [M]. Cham: Springer International Publishing, 2019.
- [5] Leighton J P, Gierl M J, Hunka S M. The attribute hierarchy method for cognitive assessment: a variation on Tatsuo's rule: space approach [J]. Journal of Educational Measurement, 2004, 41(3): 205-237.
- [6] 王晓庆, 丁树良, 罗芬. 认知诊断中的 Q 矩阵及其作用 [J]. 心理科学, 2019, 42(3): 739-746.
- [7] Tatsuo K K. Architecture of knowledge structures and cognitive diagnosis: a statistical pattern recognition and classification approach [EB/OL]. [2020-03-19]. <https://psycnet.apa.org/record/1995-97594-013>.
- [8] 涂冬波, 蔡艳, 戴海琦. 几种常用非补偿型认知诊断模型的比较与选用: 基于属性层级关系的考量 [J]. 心理学报, 2013, 45(2): 243-252.
- [9] Liu Ren, Huggins M A C, Bradshaw L. The impact of Q -matrix designs on diagnostic classification accuracy in the presence of attribute hierarchies [J]. Educational and Psychological Measurement, 2017, 77(2): 220-240.
- [10] 丁树良, 罗芬. 由偏序关系的可达阵导出 Hasse 图的有效算法: 兼谈其在认知诊断中的作用 [J]. 江西师范大学学报: 自然科学版, 2013, 37(5): 441-444.
- [11] Templin J, Bradshaw L. Hierarchical diagnostic classification models: a family of models for estimating and testing attribute hierarchies [J]. Psychometrika, 2014, 79(2): 317-339.
- [12] Xu Gongjun, Shang Zhuoran. Identifying latent structures

- in restricted latent class models [J]. *Journal of the American Statistical Association* 2018 ,113(523) : 1284-1295.
- [13] Wang Chun ,Lu Jing. Learning attribute hierarchies from data: two exploratory approaches [J]. *Journal of Educational and Behavioral Statistics* 2020 ,46(1) : 58-84.
- [14] Hu Bo ,Templin J. Using diagnostic classification models to validate attribute hierarchies and evaluate model fit in Bayesian networks [J]. *Multivariate Behavioral Research* , 2020 ,55(2) : 300-311.
- [15] 喻晓锋,丁树良,秦春影 等. 贝叶斯网在认知诊断属性层级结构确定中的应用 [J]. *心理学报* 2011 ,43(3) : 338-346.
- [16] Pearl J. On evidential reasoning in a hierarchy of hypotheses [J]. *Artificial Intelligence* ,1986 ,28(1) : 9-15.
- [17] Cooper G F ,Herskovits E. A Bayesian method for the induction of probabilistic networks from data [J]. *Machine Learning* ,1992 ,9(4) : 309-347.
- [18] Yu Xiaofeng ,Cheng Ying. Data-driven Q -matrix validation using a residual: based statistic in cognitive diagnostic assessment [J]. *British Journal of Mathematical and Statistical Psychology* 2020 ,73(S1) : 145-179.
- [19] Heller J ,Wickelmaier F. Minimum discrepancy estimation in probabilistic knowledge structures [J]. *Electronic Notes in Discrete Mathematics* 2013 ,42: 49-56.
- [20] Philipp M ,Strobl C ,de la Torre J et al. On the estimation of standard errors in cognitive diagnosis models [J]. *Journal of Educational and Behavioral Statistics* 2018 ,43(1) : 88-115.

The Attribute Hierarchical Structure Learning Based on K2 Algorithm

YU Xiaofeng¹ ,MA Yifan¹ ,LUO Zhaosheng¹ ,QIN Chunying²

(1. School of Psychology ,Jiangxi Normal University ,Nanchang Jiangxi 330022 ,China;

2. Department of Mathematics and Computer Science ,Nanchang Normal University ,Nanchang Jiangxi 330032 ,China)

Abstract: There is often some hierarchical relationship among the attributes measured by the diagnostic test ,but attribute hierarchical relationships are often difficult to obtain. Experience-based attribute-relationships are prone to errors. The attribute mastering model is taken as input ,and the accuracy of the attribute hierarchical structure learned by the K2 algorithm under different conditions is evaluated. Simulation research and empirical data analysis results show that the K2 algorithm has a high accuracy rate for the learning of the attribute hierarchy ,and the K2 algorithm has different sensitivities to the four basic hierarchical structures ,among which the linear and divergent types are sensitive to the threshold ,while the convergence type and unstructured type have higher sensitivity to the threshold.

Key words: Bayesian network structure learning algorithm; attribute hierarchical structure; K2 algorithm

(责任编辑: 冉小晓)