文章编号:1000-5862(2017)03-0289-07

多分属性层级结构下引入逻辑约束的理想掌握模式

詹沛达[1],丁树良[2],王立君[3*]

(1.北京师范大学中国基础教育质量监测协同创新中心,北京 100875;2.江西师范大学计算机信息工程学院,江西 南昌 330022;3.浙江师范大学教师教育学院心理系,浙江 金华 321004)

摘要:多分属性比传统的2分属性提供更多更详细的诊断反馈信息,具有广阔的应用前景.在多分属性情境下,当属性之间存在层级结构时,会出现原2分属性情境下不存在的逻辑问题:如果被试仅低程度地掌握了父属性,那么他是否还有可能高程度地掌握子属性?从逻辑上讲,这种"父属性掌握程度低而子属性掌握程度高"的发展情况并不具有普适性.对此,该文首先在多分属性情境下,基于现有的计算理想掌握模式的方法提出了满足"属性掌握水平约束假设"的理想掌握模式计算方法.然后,通过模拟研究说明该逻辑约束的使用方法及忽略该逻辑约束可能对诊断结果带来的危害.

关键词:认知诊断;多分认知属性;多分 Q 矩阵;多分可达矩阵;Q 矩阵;属性层级结构

中图分类号:B 841.7 文献标志码:A DOI:10.16357/j.cnki.issn1000-5862.2017.03.15

0 引言

认知诊断评估(cognitive diagnostic assessment,CDA)是指在心理与教育测量学中对个体认知过程、加工技能或知识结构(统称为属性(attributes))的诊断评估[1].CDA源于人们对测量观念的更新与测验作用的新需求,CDA在测量方法中引入认知心理模型,改变以往评估方法重结果、轻过程的弊端.不仅有助于更深入地研究隐藏在学生总分背后的认知过程和知识结构,还能提供该学生的认知诊断报告和补救性教学建议,对学生个体的发展起到了积极的促进作用.有利于实现《基础教育课程改革纲要(试行)》中"改变课程评价过分强调甄别与选拔的功能,发挥评价促进学生发展、教师提高和改进教学实践的功能"的具体目标.目前,CDA作为心理与教育测量学的前沿领域,已经受到了国内外学者的广泛关注,具有光明的未来.

当前关于CDA的研究绝大部分是基于2分属性(dichotomous attributes,用"0"和"1"分别表示"未掌握"和"掌握")和2分 Q 矩阵(dichotomous Q-matrix,Q_d,用"0"和"1"分别表示"未考查"和"考查")[2-3],但这种"非黑即白"的分类方法过于粗糙[4-5].而在实际教学和测验中更多情况是对知识技能(属性)的多水平要求和考查,比如《全日制义务教育数学课程标准(修改稿)》中就使用了"了解(认识)"、"理解"、"掌握"和"运用"这4个具有程度差异的词汇来描述对不同知识技能的多水平要求,此时基于2分属性的认知诊断方法就显得"力不从心"了.随着评估反馈精细化需求的提升,多分属性(polytomous attributes)和多分 Q 矩阵(polytomous Q-matrix,Q_p)[4]应运而生.多分属性是对2分属性的拓广,其使用顺序类别属性编码(ordered-category attribute coding,OCAC)[4]来对属性的各个水平进行编码(用"0"至"3"分别表示"了解"、"理解"、"掌握"和"运用",或用"0"至"2"分别表示"未掌握"、"勉强掌握"和"掌握很好"),它比2分属性提供更精细的诊断信息[6],更符合我国教育政策与实践应用中对不同知识技能掌握水平的差异化要求,对多分属性的研究值得关注.

在CDA中,为较好地描述不同心理特质的发展顺序,常会假定各属性之间的存在层级结构(hierarchical structure)[7],比如线型(linear)、聚合型(convergent)和发散型(divergent)等以及它们组合而成的复杂网络层

收稿日期:2017-01-26

基金项目:全国教育科学规划教育部重点课题(DBA150236)和国家自然科学基金(31360237,31500909,31300876,31160203,31100756,30860084,11401271)资助项目.

通信作者:王立君(1968-),女,辽宁大连人,副教授,博士,主要从事学科能力测量、青少年社会性发展与积极心理学等方面的研究.E-mail:franrwlj@163.com

级结构. 如图 1 所示,以线型层级结构为例,被试若想掌握属性 3 那必须先依次掌握属性 1 和属性 2,即假设掌握属性 1 是掌握属性 2 的先决条件、掌握属性 2 又是掌握属性 3 的先决条件. 为便于说明,可将先决属性称为父属性,而将后决属性称为子属性,则除了层级结构的起点(线型层级结构中的属性 1) 和终点(线型层级结构中的属性 3) 外,其余属性均同时是父属性又是子属性(线型层级结构中属性 2 既是属性 1 的子属性,又是属性 3 的父属性).

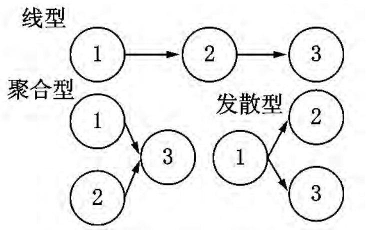

图 1　属性层级结构示例

简化 Q 矩阵是得到满足层级结构的理想掌握模式(ideal mastery pattern,IMP) 的关键,所以建构正确的、符合逻辑的简化 Q 矩阵至关重要. 目前在 CDA 中主要有 2 种建构简化 Q 矩阵的方法: 缩减算法[2-3]和扩张算法[8-9],其中缩减算法是从 2^K-1 种属性模式中删除不符合层级结构模式进而得到简化 Q 矩阵,而扩张算法是基于可达矩阵(reachability matrix,$\boldsymbol{R}$) 采用布尔加来计算简化 Q 矩阵,而且当属性个数较多时扩张算法更具优势. 得到简化 Q 矩阵后,对其引入全 0 模式即可得到 IMP. 文献[10] 在 2 分可达矩阵(dichotomous $\boldsymbol{R}_d$) 建构方法的基础上,给出了多分可达矩阵(polytoumous $\boldsymbol{R}$,$\boldsymbol{R}_p$) 的建构方法[11],丁树良等[12]也对该建构方法进行了重述,并且纠正了文献[11] 中计算线型性 $\boldsymbol{R}_d$ 矩阵的笔误. 由于 $\boldsymbol{R}_p$ 矩阵涉及到多分属性,所以其建构过程比建构 $\boldsymbol{R}_d$ 稍显复杂,具体为:(1) 假设存在 K 个多分属性 $a_1,\cdots,a_k,\cdots,a_K$,且多分属性 a_k 的最高水平数为 $L_k(L_k\geqslant 1)$;(2) 先按照属性层级结构建构出 $K\times K$ 的仅包含 0 和 1 的 $\boldsymbol{R}_d$ 矩阵;(3) 扩元: 将 $\boldsymbol{R}_d$ 矩阵中的第 (k,k) 元(对角线上的元素) 扩充为一个长度为 L_k 的连续向量 $(1,2,\cdots,L_k)$;(4) 用 0 填补空余处,得到 $K\times\sum_{k=1}^{K}L_k$ 的多分矩阵,即 $\boldsymbol{R}_p$ 矩阵.

为更好地说明 $\boldsymbol{R}_p$ 矩阵建构方法,仍以图 1 中聚合型层级结构为例,假设 3 个多分属性的 $L_k=2$,即 $a_k\in\{0,1,2\}$,则可先建构出 $\boldsymbol{R}_d$ 矩阵,扩元为 $\boldsymbol{R}_p$ 矩阵的具体过程为

$$\boldsymbol{R}_d\Rightarrow\begin{bmatrix}(1,2) & 0 & 1\\ 0 & (1,2) & 1\\ 0 & 0 & (1,2)\end{bmatrix}\Rightarrow$$

$$\begin{bmatrix}1\to2 & 0 & 0 & 1 & 1\\ 0 & 0 & 1\to2 & 1 & 1\\ 0 & 0 & 0 & 1\to2\end{bmatrix}\Rightarrow$$

$$\begin{bmatrix}1 & 2 & 0 & 0 & 1 & 1\\ 0 & 0 & 1 & 2 & 1 & 1\\ 0 & 0 & 0 & 0 & 1 & 2\end{bmatrix}=\boldsymbol{R}_p, \qquad (1)$$

至此 $\boldsymbol{R}_p$ 矩阵建构完毕. 之后,丁树良等[12] 在 $\boldsymbol{R}_p$ 矩阵的基础上采用 2 分属性情境下的扩展算法给出了简化 $\boldsymbol{Q}_p$ 矩阵的建构方法,接(1) 式,则有

$$\begin{bmatrix}1 & 2 & 0 & 0 & 1 & 1\\ 0 & 0 & 1 & 2 & 1 & 1\\ 0 & 0 & 0 & 0 & 1 & 2\end{bmatrix}\Rightarrow$$

循环 = 1,　2,　3,4,5,6

$$\left[\begin{array}{cccccc|cc|cccc|cc|c|c}1 & 2 & 0 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 & 2 & 2\\ 0 & 0 & 1 & 2 & 1 & 1 & 1 & 2 & 1 & 2 & 1 & 1 & 2 & 2 & 2 & 2\\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 1 & 2 & 1 & 2 & 1 & 2\end{array}\right]. \qquad (2)$$

值得注意的是,(2) 式仅仅是 $L_k=2$ 时的结果,倘若 L_k 取值更大或多分属性之间的层级结构更松散又或是多分属性数量更多时,简化 Q_p 矩阵包含的列数将远大于简化 Q_d 矩阵的列数. 对(2) 式加上全 0 模式即可得到多分属性情境下的 IMP_p 为

$$IMP_p=\begin{bmatrix}1&2&0&0&1&1&1&1&2&2&2&2&1&1&2&2&0\\0&0&1&2&1&1&1&2&1&2&1&1&2&2&2&2&0\\0&0&0&0&1&2&0&0&0&0&1&2&1&2&1&2&0\end{bmatrix}.$$

为行文方便,下文把上述流程得到的 IMP_p 称为"初始 IMP_p". 易知,初始 IMP_p 隐藏了一个潜在逻辑: 只要被试掌握了父属性(哪怕掌握的水平再低),就有可能高水平地掌握子属性,即在本例中允许如(122) 和(112) 这样的属性掌握模式存在.

属性层级结构本身就是一个强假设,即假设被试掌握某属性的前提是其掌握了该属性的父属性. 但当这一强假设从 2 分属性情境拓广至多分属性情境时(即把属性掌握情况进行更为精细地划分时) 就会出现一个逻辑性问题: 如果被试仅低程度(需要强调的是本研究中"程度"和"水平"是 2 个并不完全相同的概念,"程度"是真实存在的,而"水平"仅仅是对"程度"的一种量化或描述. 真实"程度"低,但可以被划分到一个看似不低的"水平"之中,反之亦然.) 地掌握了父属性,那他是否还有可能高程度地掌握子属性? 仍以图 1 中聚合型层级结构为

例，当被试仅较低程度地掌握了属性 1 和属性 2，那他是否还有可能高程度地掌握属性 3 呢? 比如，如果被试只能较低程度的掌握“加法运算”，那他是否还有可能较高程度地掌握“乘法运算”? 再进而是否还有可能较高程度地掌握“四则运算”? 从事物发展的角度讲，“被试对父属性仅略知一二，而对子属性运用自如”这种倒三角的认知发展过程尽管有存在的可能但不应具有普适性. 这点与属性层级结构假设本身是类似的，即在假设专家制定的属性层级结构建构是正确的、符合事物发展的前提下，那些违反属性层级结构的属性模式也有存在的可能但不应具有普适性. 在认知和发展心理学的相关研究中也不难发现绝大多数能力的发展存在顺序性且发展程度是依发展顺序递减的. 比如，N. C. Jordan 等[13]提出儿童数感(number sense) 在发展上有先后顺序且每一阶段数感的发展是以前一阶段各个成分能力的获得为基础的，即数数(counting) →数知识(number knowledge) →数量转换(number transformation) →数量估计(number estimation) →数型(number patterns) . 在幼儿排序能力发展方面，周欣等[14]指出 4 ~5岁儿童排序能力的发展顺序为: 正排序(由少到多排序) →逆排序(由多到少排序) ; 而戴佳毅等[15]的研究发现 4 ~6 岁儿童对正排序任务的完成率始终高于相应的逆排序任务. 表明样本中大多数儿童对于父属性(正排序) 的掌握程度要高于其对子属性(逆排序) 的掌握程度. 另外，在一般语言能力和心理理论(theory of mind) 方面的研究[16]也表明一般语言能力是(4 岁) 儿童心理理论的显著性预测指标且呈现出一定的因果关系，而已有研究普遍支持在儿童发展的早期，一般语言能力是儿童心理理论发展的前提和基础[17]. 上述几个例子从不同角度支持了能力的发展程度依发展顺序递减这一基本逻辑.

综上所述，初始 IMP_p 隐藏的潜在逻辑(只要被试掌握了父属性(哪怕掌握的水平再低) 就有可能高水平地掌握子属性) 是与能力的发展程度依发展顺序递减这一基本逻辑不相符的，因此需要给予初始 IMP_p 一些额外的逻辑约束. 本文所关注的点就是: 当假设在某认知结构中这种父属性程度低而子属性程度高的情况是不存在或可忽略时，如何对假设本身进行量化，对初始 IMP_p 的计算过程进行约束. 另外，由于多分属性采用了相对更多的“水平”来划分属性，因此一个可尝试的方法是从各“水平”之间关系入手.

1　针对多分属性情境中理想掌握模式的逻辑约束

1.1　属性掌握水平约束假设

从逻辑上讲，若被试掌握某属性的前提是其先要掌握该属性的父属性，那么父属性掌握程度低而子属性掌握程度高的情况是缺少普适性的. 这个逻辑较好理解，将被试的学习过程比作盖大楼，属性就是大楼的每一层，层级结构就好比各个楼层之间的关系. 只有最牢固的地基配上稳固的下层建筑，才有可能盖起更高、更稳固的大楼. 同理，只有被试牢固地掌握了起始属性或父属性，其才有条件去进一步学习和掌握子属性. 但该逻辑问题并没有被之前的研究[11-12]考虑到，因此在模拟研究以及后续参数估计中可能出现被试“掌握”了不符合逻辑的属性模式. 针对该问题，可尝试在计算 IMP_p 的过程中添加逻辑约束.

需要说明的是，尽管上述假设是从“掌握水平”入手的，但本文指出的逻辑问题中的“掌握程度”和多分属性中的“掌握水平”是 2 个并不完全相同的概念. 为使“掌握水平”更好地反应“掌握程度”，且考虑到实际测验中不同属性的最高水平数(L_k) 可能是不同的，先对“掌握水平”进行百分比转化，即 $\delta_{nk} = \alpha_{nk}/L_k$，其中 δ_{nk} 表示被试 n 对属性 k 的掌握水平百分比. 比如，$L_1 = 2$ 且 $\alpha_{n1} = 1$，则表示被试 n 对属性 1 的掌握水平为 50%; 若 $L_2 = 5$ 且 $\alpha_{n2} = 2$，则表示被试 n 对属性 2 的掌握水平为 2/5. 此时说明被试 n 对属性 1 的掌握水平高于对属性 2 的掌握水平，尽管 α_{n1} 本身小于 α_{n2}.

基于此，“属性掌握水平约束假设”可描述为: 假设多分属性间存在层级结构时，被试对父属性的掌握水平百分比大于等于其对子属性的掌握水平百分比，即

$$\alpha_{nk(\mathrm{Father})} / L_{k(\mathrm{Father})} \geqslant \alpha_{nk(\mathrm{Son})} / L_{k(\mathrm{Son})} , \qquad (3)$$

其中 $\alpha_{nk(\mathrm{Father})}$ 和 $\alpha_{nk(\mathrm{Son})}$ 分别表示被试对父属性和子属性的掌握水平，$L_{k(\mathrm{Father})}$ 和 $L_{k(\mathrm{Son})}$ 分别表示父属性和子属性的最高水平数. 以图 1 中的 3 种属性层级结构为例，根据属性掌握水平约束假设有

线型: $a_1/L_1 \geqslant a_2/L_2 \geqslant a_3/L_3$;　　(4)

聚合型: $a_1/L_1 \geqslant a_3/L_3$，$a_2/L_2 \geqslant a_3/L_3$;

发散型: $a_1/L_1 \geqslant a_2/L_2$，$a_1/L_1 \geqslant a_3/L_3$.

为简化研究且不失一般性，下文设定所有属性

的最高水平数一样[6]，此时，(3) 式就等价于 $\alpha_{nk(\mathrm{Father})} \geqslant \alpha_{nk(\mathrm{Son})}$.

需要强调的是，任何“假设”都有其局限性，并不能完全覆盖所有情境. 比如接受度较高的属性层级结构假设，若某测验欲使用该假设，则相当于承认其施测群体中不存在掌握了不满足层级结构的属性模式的学生. 但实际并非如此，其导致的结果就是将属性层级结构假设覆盖之外的少数学生“强行”纳入这个假设之中，进而可能出现对这部分学生的不恰当诊断. 类似，若某测验欲使用属性掌握水平约束假设，也可能出现对部分假设覆盖之外的学生的不恰当诊断. 而至于，承认这些假设和不承认属性层级结构假设和被试掌握水平约束假设之间的区别，笔者认为：承认假设成立相当于“资源集中原则”，即把有限的测验信息量分配给那些真的满足假设的被试，而忽略不满足假设的被试；而不承认假设成立则相当于“资源均分原则”，即在有限的测验信息量情况下，考虑到所有的被试(属性模式). 孰优孰劣，或许就需要使用一些模型数据拟合判断指标(e. g.，AIC、BIC 和 DIC 等) 来判定，不过可以大体推断出：超出假设覆盖面的人越少，则“资源集中原则”更有优势，反之“资源均分原则”更有优势.

1.2 计算有约束的 IMP_p

欲把属性掌握水平约束假设引入计算 IMP_p 的过程中，可以从简化 Q_p 矩阵切入，流程为：(i) 得到 $\boldsymbol{R}_p$ 矩阵[11-12]；(ii) 基于 $\boldsymbol{R}_p$ 矩阵计算简化 Q_p 矩阵，可称为初始简化 Q_p 矩阵；(iii) 把初始简化 Q_p 矩阵中不满足属性掌握水平约束假设的列删除，得到约束简化 Q_p 矩阵；(iv) 对约束简化 Q_p 矩阵加入全 0 模式，即可得到约束 IMP_p.

仍以图 1 中聚合型层级结构为例，把初始简化 Q_p 矩阵转为约束简化 Q_p 矩阵的过程为

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 & 2 & 2 \\ 0 & 0 & 1 & 2 & 1 & 1 & 1 & 2 & 1 & 2 & 1 & 1 & 2 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 1 & 2 & 1 & 2 & 1 & 2 \end{bmatrix}$$

$$\xrightarrow{a_1 \geqslant a_3, a_2 \geqslant a_3}$$

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 & 1 & 2 & 2 \\ 0 & 0 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 \end{bmatrix}, \quad (5)$$

再对(5) 式添加全 0 模式可得到约束 IMP_p 为

约束 IMP_p =

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 & 1 & 2 & 2 & 0 \\ 0 & 0 & 1 & 2 & 1 & 1 & 2 & 1 & 2 & 1 & 2 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 0 \end{bmatrix}.$$

约束 IMP_p 比初始 IMP_p(见(4) 式) 少 3 种不符合逻辑的属性模式，且随着 L_k 取值增大或属性数量更多时，约束 IMP_p 比初始 IMP_p 之间的列数差将会更大.

最后需要强调的是，属性掌握水平约束假设仅适用于被试对属性的掌握情况，而不适用于题目对属性的考查，也就是说，题目对父属性的考查水平可以低于其对子属性的考查水平. 这很好理解，本文不再解释.

2 模拟研究

2.1 研究目的与设计

2.1.1 研究目的与设计　模拟研究目的是为探究当忽视“父属性程度低而子属性程度高”这一逻辑问题时对测验结果带来的危害. 当然，该研究是以属性掌握水平约束假设为正确的前提，即被试的“真实”属性掌握模式中不存在“父属性程度低而子属性程度高”的情况. 为呈现忽视该假设可能带来的危害，分别采用包含该假设的模型(真实模型) 和忽视该假设模型进行参数估计，之后对比两者属性判准率的之间的差异(题目参数返真性暂不考虑).

为简化研究且不失一般性，假设有 $K = 3$ 个最高水平数为 $L_k = 3$ 的多分属性，并涉及图 1 中 3 种属性层级结构. 对于聚合型层级结构而言，初始 IMP_p 为 3×42 的矩阵，引入属性掌握水平约束假设($\alpha_1 \geqslant \alpha_3$ 且 $\alpha_2 \geqslant \alpha_3$) 后，可得到 3×31 的约束 IMP_p 为

约束 IMP_p =

$$\begin{bmatrix} 1 & 2 & 3 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \cdots & 1 & 2 & 3 & 3 & 3 & 2 & 3 & 3 & 0 \\ 0 & 0 & 0 & 1 & 2 & 3 & 1 & 1 & 2 & 3 & \cdots & 3 & 3 & 1 & 2 & 3 & 3 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 0 \end{bmatrix}. \quad (6)$$

对于线型层级结构而言，初始 IMP_p 为 3×41 的矩阵，引入属性掌握水平约束假设($a_1 \geqslant a_2 \geqslant a_3$) 后，可得到 3×20 的约束 IMP_p 为

约束 IMP_p =

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 3 & 2 & 3 & 3 & 1 & 2 & 3 & 2 & 3 & 3 & 2 & 3 & 3 & 3 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 3 & 1 & 1 & 1 & 2 & 2 & 3 & 2 & 2 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 3 & 0 \end{bmatrix}. \quad (7)$$

对于发散型层级结构而言，初始 IMP_p 为 3×46 的矩阵，引入属性掌握水平约束假设($\alpha_1 \geqslant \alpha_2$ 且 $\alpha_1 \geqslant \alpha_3$) 后，可得到 3×29 的约束 IMP_p 为

约束 IMP_p =

$$\begin{bmatrix} 1&2&3&1&1&2&2&2&2&3&\cdots&2&2&3&3&3&3&3&3&0 \\ 0&0&0&1&0&1&2&0&0&1&\cdots&2&2&2&2&2&3&3&3&0 \\ 0&0&0&0&1&0&0&1&2&0&\cdots&1&2&1&2&3&1&2&3&0 \end{bmatrix}. \tag{8}$$

2.1.2　**模型选择**　模拟作答数据生成采用重参数化多分属性 DINA（reparameterized polytomous attributes DINA，RPa-DINA）模型[5]，该模型可描述为

$$P_{ni1} = (1 - s_i - g_i)\eta_{ni} + g_i,$$

$$\eta_{ni} = \prod_{k=1}^{K} \omega_{nik}^{q_{ik}^*},$$

$$\omega_{nik} = I\{\alpha_{nk} \geqslant q_{ik}\}, q_{ik}^* = I\{q_{ik} > 0\},$$

其中 P_{ni1} 表示被试 n 答对题目 i 的概率，s_i 为失误参数，g_i 为猜测参数，K 为属性总数，$q_{ik} \in \{0,1,\cdots,L\}$ 为 Q_p 矩阵中的元素，$\alpha_{nk} \in \{0,1,\cdots,L\}$ 为被试 n 对属性 k 的掌握情况，$\eta_{ni} \in \{0,1\}$ 为被试 n 对题目 i 的理想作答（ideal response），$q_{ik}^* \in \{0,1\}$ 为坍缩 Q 矩阵（collapsed $\boldsymbol{Q}$ matrix，将 Q_p 矩阵中大于 0 的元素坍缩为 1，用于表示题目是否考查了该属性，即与 Q_p 矩阵对应的 Q_d 矩阵）中的元素，$\omega_{nik} \in \{0,1\}$ 可被视为被试 n 对题目 i 中属性 k 的潜在作答，它是一个条件变量，即假设当被试 n 对第 k 个属性的掌握水平大于等于题目 i 对第 k 个属性的考查水平时才有 $\omega_{nik} = 1$. 进一步根据连接缩合规则（conjunctive condensation rule）可知，仅当被试 n 对题目 i 的所有潜在作答均为 1 时（$\sum_{k=1}^{K} \omega_{nik} = K$），被试的理想作答 η_{ni} 才等于 1. RPa-DINA 的待估计参数为 $2I + NP - 1$ 个，NP 为所有可能的属性模式数量，即 IMP 的数量. 当不考虑属性层级结构假设也不考虑属性掌握水平约束假设时 $NP = \prod_{k=1}^{K}(L_k + 1) = 64$. 当考虑层级结构假设或属性掌握水平约束假设时可将不符合假设的属性模式从待估计的属性模式参数中删除，对聚合型层级结构而言，当仅考虑属性层级假设时 $NP = 42$，而 2 个假设都考虑时 $NP = 31$，对线性层级结构而言，当仅考虑属性层级假设时 $NP = 41$，而 2 个假设都考虑时 $NP = 20$，对于发散型层级结构而言，当仅考虑属性层级结构假设时 $NP = 46$，而 2 个假设都考虑时 $NP = 29$.

2.1.3　**参数设定、模拟作答与评价指标**　设定被试量 $N = 1\ 000$，被试的“真实”属性掌握模式从（6）~（8）式中随机抽取；题目数 $I = 30$，且所有题目的猜测参数 $g_i = 0.1$，失误参数 $s_i = 0.1$；测验 Q_p 矩阵基于（6）~（8）式（删除全 0 模式）随机生成. 模拟作答时，首先根据各参数“真值”和 RPa-DINA 计算被试 n 在项目 i 上的正确作答概率 P_{ni1}，则被试作答服从伯努利分布：$Y_{ni1} \sim \mathrm{Bernoulli}(P_{ni1})$. 为探究属性返真性，同时采用精准分类（exact classification）和加权分类（weighted classification）[6] 判别法，2 方法均基于属性判准率 ACCR 和属性模式判准率 PCCR 为

$$ACCR(\alpha_k) = \left(\sum_{n=1}^{N} W_{nk}\right) / N,$$

$$PCCR(\alpha) = \left(\sum_{n=1}^{N} \prod_{k=1}^{K} W_{nk}\right) / N;$$

对于精准分类

$$W_{nk} = \begin{cases} 1, \text{if } \alpha_{nk} = \hat{\alpha}_{nk}, \\ 0, \text{otherwise}; \end{cases}$$

而对于加权分类，

$$W_{nk} = \begin{cases} 1/2^{|\alpha_{nk} - \hat{\alpha}_{nk}|}, & \text{if } |\alpha_{nk} - \hat{\alpha}_{nk}| < L_k, \\ 0, & \text{otherwise}, \end{cases}$$

其中 α_{nk} 和 $\hat{\alpha}_{nk}$ 分别表示真实属性水平和估计属性水平. 加权分类法要比精准分类法更为宽松，但也符合常理，比如被试某属性的真实水平是 3，那么估计值为 2 总比估计值是 1 来得好. 模型-数据拟合指标采用 −2LL、AIC、BIC 和 DIC 指标，指标值越小表明相对拟合越好.

2.2　研究结果与结论

表 1 列出了考虑掌握水平约束假设和忽略掌握水平约束假设这 2 种情况下模型-数据拟合结果. 可以看到 4 个相对拟合指标均表明考虑掌握水平约束假设时，模型对数据的拟合程度更高. 表 2 列出了考虑假设和忽视假设 2 种情况下多分属性判准率指标值，可以看到：1）无论是 ACCR 还是 PCCR，无论是精准分类还是加权分类，考虑假设时的判准率均高于忽略假设时的判准率，这点符合“资源集中原则”，因为在模拟研究中有没有被试超出属性约束水平假设覆盖范围；2）加权分类比精准分类的判准率要高，这符合预期也非本文关注点. 本文的关注点是精准分类比加权分类下的判准率要低多少. 通过表 2 括号中的数据可看到，当考虑到假设时精准分类与加权分类之间的判准率差值要小于忽略假设时两者之间的差值. 这说明考虑假设时没有完全估计准确的属性数量要相对更少，即这种处于中间的不确定态相对更少.

总之，根据研究结果可知当受测群体中不存在“父属性程度低而子属性程度高”这种不符合逻辑的被试时，若分析数据时考虑属性掌握水平约束假设则能够得到相对更为准确的属性判准率，反之得

到更低的判准率.

表 1 模型-数据拟合程度

属性层级结构	分析模型	−2LL	AIC	BIC	DIC
线型	RPa-DINA(忽略掌握水平约束假设)	19 633.31	19 833.31	20 324.09	22 275.31
	RPa-DINA(考虑掌握水平约束假设)	19 546.90	19 704.90	20 092.61	21 228.44
聚合型	RPa-DINA(忽略掌握水平约束假设)	19 509.93	19 711.93	20 207.61	21 435.24
	RPa-DINA(考虑掌握水平约束假设)	19 438.16	19 618.16	20 059.86	20 887.76
发散型	RPa-DINA(忽略掌握水平约束假设)	19 509.49	19 719.49	20 234.80	21 599.51
	RPa-DINA(考虑掌握水平约束假设)	19 442.75	19 618.75	20 050.63	21 270.90

表 2 多分属性判准率(括号内为加权分类与精准分类之间差值)

属性层级结构	是否考虑假设	判准方法	ACCR			PCCR
			α_1	α_2	α_3	
线型	忽略假设	精准分类	0.908	0.896	0.917	0.769
		加权分类	0.955(0.047)	0.949(0.053)	0.960(0.043)	0.876(0.107)
	考虑假设	精准分类	0.921	0.949	0.957	0.853
		加权分类	0.962(0.041)	0.975(0.026)	0.979(0.022)	0.921(0.068)
聚合型	忽略假设	精准分类	0.931	0.924	0.916	0.801
		加权分类	0.967(0.036)	0.963(0.039)	0.954(0.038)	0.896(0.095)
	考虑假设	精准分类	0.932	0.925	0.959	0.838
		加权分类	0.967(0.035)	0.963(0.038)	0.980(0.021)	0.915(0.077)
发散型	忽略假设	精准分类	0.916	0.915	0.928	0.800
		加权分类	0.959(0.043)	0.959(0.044)	0.965(0.037)	0.892(0.092)
	考虑假设	精准分类	0.930	0.945	0.948	0.848
		加权分类	0.966(0.036)	0.973(0.028)	0.974(0.026)	0.919(0.071)

3 总结

目前,关于多分属性的研究尚属于国际前沿,仅有少许相关研究暂涉及,包括多分属性的设定[4]、多分可达矩阵的建构[11-12]、多分属性认知诊断模型与方法探讨[5-6,11]和多分 Q 矩阵理论[12]等. 为了解决“只要被试掌握了父属性(哪怕掌握的水平再低),就有可能高水平地掌握子属性”这一逻辑问题,本文尝试对计算 IMP_p 过程中引入属性掌握水平约束假设,该假设限制被试对父属性的掌握水平大于等于其对子属性的掌握水平. 当然上文也有提到,这一逻辑约束仅仅适用于被试对属性的掌握,而并不影响题目对属性的考查. 另外,本文还通过模拟研究展示了属性掌握水平约束假设的用法和忽略属性掌握水平约束假设对测验结果带来的危害. 最后,需要强调的是本文提出的是一种逻辑性约束,还需要后续实证研究的验证.

4 参考文献

[1] Yang Xiangdong, Embretson S E. Construct validity and cognitive diagnostic assessment [A]. Gierl J P L M. Cognitive diagnostic assessment for education: theory and applications [C]. Cambridge, UK: Cambridge University Press, 2007: 119-145.

[2] Tatsuoka K K. Rule space: an approach for dealing with misconceptions based on item response theory [J]. Journal of Educational Measurement, 1983(20): 345-354.

[3] Tatsuoka K K. A probabilistic model for dianosing misconceptions by the pattern classification approach [J]. Journal of Educational and Behavioral Acquisition, 1985(10): 453-488.

[4] Karelitz T M. Ordered category attribute coding framework for cognitive assessments [D]. Urbana: University of Illinois at Urbana-Champaign, 2004: 44-67.

[5] 詹沛达,边玉芳,王立君. 重参数化的多分属性诊断分类模型及其判准率影响因素 [J]. 心理学报, 2016(48): 318-330.

[6] Chen Jinsong, de la Torre J. A general cognitive diagnosis model for expert-defined polytoumous attributes [J]. Applied Psychological Measurement, 2013(37): 417-437.

[7] Leighton J P, Gierl M J, Hunka S M. The attribute hierarchy method for cognitive assessment: a variation on Tatsuoka´s rule-space approach [J]. Journal of Educational Measurement, 2004(41): 205-237.

[8] 杨淑群，蔡声镇，丁树良，等. 求解简化 Q 矩阵的扩张算法[J]. 兰州大学学报：自然科学版，2008(3)：87-91.

[9] Ding Shuliang，Luo Fen，Cai Yan，et al. Complement to tatsuoka's Q matrix Theory[C]. Tokyo：Universal Academy Press，2008：417-424.

[10] 涂冬波，蔡艳，丁树良. 认知诊断理论、方法与应用[M]. 北京：北京师范大学出版社，2012：1-30.

[11] Sun Jianan，Xin Tao，Zhang Shumei，et al. A polytomous extension of the generalized distance discriminating method[J]. Applied Psychological Measurement，2013(37)：503-521.

[12] 丁树良，罗芬，汪文义，等. 0-1 和多值可达矩阵的性质及应用[J]. 江西师范大学学报：自然科学版，2015，39(1)：64-68.

[13] Jordan N C，Kaplan D，Oláh L，et al. Number sense growth in kindergarten：a longitudinal investigation of children at risk for mathematics difficulties[J]. Child Development，2006(77)：153-175.

[14] 周欣，王滨. 4～5 岁儿童对书面数符号的表征和理解能力的发展[J]. 心理科学，2004，27(5)：1132-1136.

[15] 戴佳毅，王滨. 4～6 岁幼儿排序能力发展特点的初步研究[J]. 幼儿教育：教育科学，2007(10)：37-40.

[16] 张丽锦，吴南. 4、5 岁儿童一般语言能力和心理理论关系的纵向研究[J]. 心理学报，2004，42(12)：1166-1174.

[17] Tardif T，Wing-Chen So C，Kaciroti N. Language and false belief：evidence for general，not specific，effects in Cantonese-speaking preschoolers[J]. Developmental Psychology，2007，43(2)：318-340.

The Ideal Mastery Pattern for Polytomous Attributes with Hierarchical Structure Incorporating Mastery Level Restriction

ZHAN Peida[1]，DING Shuliang[2]，WANG Lijun[3*]

(1. Collaborative Innovation Center of Assessment Toward Basic Education Quality，Beijing Normal University，Beijing 100875，China；

2. College of Computer Information Engineering，Jiangxi Normal University，Nanchang Jiangxi 330022，China；

3. Department of Psychology，Zhejiang Normal University，Jinhua Zhejiang 321004，China)

Abstract: The polytomous attributes，particularly those defined as part of the test development process，can provide additional diagnostic information. When polytomous attributes follow a hierarchical structure，a latent logical problem will emerge，which is if a student has only acquired the low level of a father attribute(i. e.，pre-requisite attribute)，will he acquire the high level of the son attribute? This situation is uncommon in reality. In terms of logical，the process of learning generally proceeds sequentially，so a good mastery of one attribute must base on a good enough pattern prepared for the pre-requisite attribute. Ideal mastery pattern(IMP) included the all possible mastery pattern for students within an assessment. Unfortunately，the existing calculation for IMP from the polytoumous reachability matrix and the polytomous reduced Q matrix ignored the logical problem above-mentioned. Then there will be some students be classified into illogical attribute pattern，such as the(122) and(112) in linear hierarchical structure. Aimed at this problem，a logical restraint of IMP for polytomous attributes is proposed，i. e. restrict the mastery level of father attribute is higher or equal to the son attribute. A simulation study was given to demonstrate applications and implications of the mastery level restriction. Results show that ignoring the mastery level restraint would result in worse model-data fit and worse attribute(pattern) correct classification rate，when the response data was generated from true attributes that followed the mastery level restraint.

Key words: cognitive diagnosis; polytomous attributes; polytomous Q matrix; polytomous reachability matrix; Q matrix; attributes hierarchical structure

(责任编辑：冉小晓)