

文章编号:1000-5862(2020)02-0136-06

带噪音预处理的改进探索性 Q 矩阵标定方法

汪文义¹,高朋¹,宋丽红²,汪腾¹

(1. 江西师范大学计算机信息工程学院,江西 南昌 330022;2. 江西师范大学初等教育学院,江西 南昌 330022)

摘要:考虑到在实际应用中学生在做题时的猜测和失误(统称为噪音)会影响探索性因素分析法所使用的四分相关矩阵的质量,该文提出四分相关矩阵的一种噪音修正方法,并将其应用于 Q 矩阵标定.模拟研究表明:猜测和失误这2种噪音会对 Q 矩阵的标定产生不利的影响;基于修正后的四分相关矩阵的探索性因素分析法,在样本量较大和噪音较大等情况下,均能有效地提高 Q 矩阵标定的准确率.

关键词:认知诊断; Q 矩阵;探索性因素分析方法;四分相关系数;数据预处理

中图分类号:B 841 **文献标志码:**A **DOI:**10.16357/j.cnki.issn1000-5862.2020.02.04

0 引言

随着认知心理学和心理测量学的飞速发展,认知诊断由于其能够在微观角度方面详细地描述学生的知识掌握状态而表现出巨大的发展潜力.不过,认知诊断是一项复杂的工程,它至少应包括“ Q 矩阵理论”和“诊断分类”2大部分^[1]. Q 矩阵是对测验项目与认知属性关系的描述,是诊断分类的基础,所以 Q 矩阵的合理标定是认知诊断的重要环节.根据以往的研究, Q 矩阵的标定方法一般有3种思路:项目的简单检查法、多评分者法和基于项目参数的迭代方法^[2].

项目的简单检查法和多评分者法的主要依据是测验编制者或领域专家的主观经验,虽然保证了标定的 Q 矩阵的可解释性,但由于其主观性,对同一测验的 Q 矩阵的标定就必然存在争议.典型的例子是国外的许多研究者对K. K. Tatsuoka的分数减法测验的属性界定就争论了20多年,至今仍无定论.基于项目参数的迭代方法是先粗略地标定 Q 矩阵,再根据项目参数估计是否收敛或异常,学生属性掌握模式的分类准确率是否太低,模型数据是否拟合等信息来评估初始 Q 矩阵的质量,然后对初始 Q 矩阵进行修正,最终得到较客观的 Q 矩阵.但该方法存在模型参数的估计严重依赖初始界定的 Q 矩阵、

错误的 Q 矩阵对诊断模型的拟合也比较敏感等问题.

不论一个测验所考察的具体属性是什么,其 Q 矩阵与项目参数信息都隐藏在学生的作答反应中,因此可以尝试从学生的作答反应入手来估计项目参数和初始 Q 矩阵.汪文义等^[3]提出了基于探索性因素分析的 Q 矩阵标定方法,在只有作答数据且对 Q 矩阵一无所知时,结合社会科学中广泛使用的因素分析法,在0-1评分的基础上基于作答数据探索初始 Q 矩阵,所得的初始 Q 矩阵标定准确率较高,再通过验证性方法修正初始 Q 矩阵,进一步提高其标定准确率.本文在该方法的基础上,考虑到在实际应用中学生在做题时的噪音会影响探索性因素分析所使用的四分相关矩阵的质量,提出四分相关矩阵的一种噪音修正方法,并将其应用于基于探索性因素分析的 Q 矩阵标定方法,以提高原方法 Q 矩阵标定的准确率.

1 研究方法

1.1 认知诊断模型

认知诊断模型是用于根据被试在认知诊断测验上作答数据来推断其在多个属性上的掌握情况的一类心理计量模型.常见的认知诊断模型有:确定性输

收稿日期:2019-04-25
基金项目:国家自然科学基金(61967009,31500909,31360237,31160203,30860084),全国教育科学规划教育部重点课题(DHA150285),江西省社会科学规划(17JY10)和江西师范大学教学改革研究(JXSDJG1848)资助项目.
作者简介:汪文义(1983-),男,湖南衡山人,副教授,博士,主要从事教育测量与信息处理的研究. E-mail:wenyiwang@jxnu.edu.cn

入噪音与门(deterministic inputs, noisy “and” gate, DINA)模型、确定性输入噪音或门(deterministic inputs, noisy “or” gate, DINO)模型、简化的重新参数化统一模型(reduced reparameterized unified model, R-RUM)等. 根据属性间的关系, DINA 与 DINO 分别属于非补偿与补偿模型, 而 R-RUM 介于补偿与非补偿模型之间. 本文关注于噪音对 Q 矩阵标定的影响, 主要考察新方法在 DINA 模型、DINO 模型和 R-RUM 下标定 Q 矩阵的返真率.

设所考察的属性个数为 K , 第 K 个属性用 A_K 表示. 由这 K 个属性的不尽相同的组合编制了 J 个题目, 以一个 $J \times K$ 的矩阵表示题目对属性的考察情况. 矩阵的第 j 行反映了第 j 题对第 k 个属性的考察情况, 当矩阵元素 $q_{jk} = 1$ 时, 第 j 题考察了第 k 个属性, 否则即为未考察该属性. 第 j 个题目的猜测参数为 g_j , 失误参数为 s_j , 设有 I 个学生参加了该测验, 第 i ($i = 1, 2, \dots, I$) 个学生对 K 个属性的掌握情况用 α_i 表示. 本文还涉及以下相关模型.

DINA 模型是一个参数定义在题目水平上的非补偿模型, 当 2 值潜在变量或理想反应 $\varepsilon_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$ 取值为 1 时, 第 i 个学生掌握了第 j 个题目所考察的所有属性, 若取值为 0 则表示未完全掌握. 该题相应的正确作答概率分别为 $1 - s_j$ 和 g_j 2 种. DINA 模型的公式^[4] 为

$$P(X_{ij} = 1 \mid \alpha_i) = (1 - s_j)^{\varepsilon_{ij}} g_j^{1 - \varepsilon_{ij}}.$$

DINO 模型是参数定义在题目水平上的补偿模型, 适用于学生只要掌握了题目所考察的任意一个属性就有较高的作答概率的测验情景, DINO 模型通过 2 值潜在变量或理想反应 $\omega_{ij} = 1 - \prod_{k=1}^K (1 - \alpha_{ik})^{q_{jk}}$ 的取值来表示学生是否掌握了至少 1 个属性, DINO 模型的公式^[5] 为

$$P(X_{ij} = 1 \mid \alpha_i) = (1 - s_j)^{\omega_{ij}} g_j^{1 - \omega_{ij}}.$$

R-RUM 是基于参数识别的目的对统一模型进行参数重构而提出的模型. 模型参数 $\pi_j^* = \prod_{k=1}^K P(Y_{ijk} = 1 \mid \alpha_{ik} = 1)^{q_{jk}}$ 为定义在题目水平上的参数, 取值均在 0 ~ 1 之间, π_j^* 的值越大表示题目越简单. 模型参数 $r_{jk}^* = r_{jk} / \pi_j^* = P(Y_{ijk} = 1 \mid \alpha_{ik} = 0) / (P(Y_{ijk} = 1 \mid \alpha_{jk} = 1))$ 是定义在题目和属性交叉的水平上的参数, 取值均在 0 ~ 1 之间, r_{jk}^* 的个数等于题目 j 所考察的属性个数 K_j^* . r_{jk}^* 是学生没有掌握属性 k 但靠

猜测在题目 j 上正确应用该属性的概率与掌握了属性 k 的学生能在题目 j 上正确应用该属性的概率之比, 因此 r_{jk}^* 反映了属性 k 在第 j 题上的重要程度. R-RUM 的模型公式^[6] 为

$$P(X_{ij} = 1 \mid \alpha_i) = \pi_j^* \prod_{k=1}^K r_{jk}^{*(1 - \alpha_{jk}) q_{jk}}.$$

1.2 带噪音预处理的四分相关矩阵估计

在实际中, 学生作答会受不同因素的影响. 若在本应作答错误的题目上作答正确或在本应作答正确的题目上作答错误, 则被称为发生了猜测或失误. 事实上, 猜测和失误是作答数据中噪音的主要来源. 因此, 通过实测获得的作答反应是潜在反应受猜测和失误这 2 种噪音影响的结果. 在进行 Q 矩阵估计时, 为了得到更精确的 Q 矩阵, 就需要对噪音进行预处理.

要对噪音进行预处理, 首先需从作答数据中估计出每道题目的失误和猜测. F. B. Baker^[7] 认为, 当样本量介于 2 000 ~ 3 000 之间时, 可将能力值低于 - 3.5 的被试的正确作答比率看作猜测, 若样本量小于 2 000, 能力值可放宽到 - 2.5. 猜测率和失误率分别可视为正确作答概率的下渐近线和上渐近线, 因此猜测和失误参数均可按此方法类似估计. 本研究的样本量在 1 000 以下, 能力值放宽到 - 1.645 ~ 1.645, 假设能力服从标准正态分布, 能力低于 - 1.645 或高于 1.645 的比例均为 5%. 如果对接受测试的一组学生按照总分进行排序, 从得分最低的学生开始取 5% 的学生称为低分组, 从得分最高的学生开始取 5% 的学生称为高分组. 若样本量足够, 则基本上可以认为, 低分组的学生未掌握该测试考察的任何属性, 高分组的学生掌握了该测试所考察的所有属性. 因此, 可以将低分组学生在某道题上的作答正确的比率近似认为是该题的猜测概率, 高分组学生在某道题上作答错误的比率近似认为是该题的失误概率. 在认知诊断测验中, 若测验测量 5 个属性且各属性间相互独立, 则在知识状态分布为均匀分布情况下, 约为 3% 的知识状态为一个属性没掌握或所有属性全部掌握的知识状态. 这也可能较好地解释为何选择 5% 的高分组和低分组的学生的作答反应来估计失误率和猜测率.

对于题目 J_1 和题目 J_2 , A 、 B 、 C 和 D 分别表示在这 2 道题上的得分模式为 (0, 0)、(0, 1)、(1, 0) 和 (1, 1) 的观察频数或频率, a 、 b 、 c 和 d 分别代表对应潜在反应或理想反应模式(无噪音)为 (0, 0)、(0, 1)、(1, 0) 和 (1, 1) 的理论频数或频率, s_1 、 s_2 和 g_1 、 g_2 分别代表题目 J_1 和 J_2 的失误与猜测概率. 观察频

率、理论频率和项目参数满足如下公式:

$$\begin{cases} A = (1 - g_1)(1 - g_2)a + (1 - g_1)s_2b + \\ \quad s_1(1 - g_2)c + s_1s_2d, \\ B = (1 - g_1)g_2a + (1 - g_1)(1 - s_2)b + \\ \quad s_1g_2c + s_1(1 - s_2)d, \\ C = g_1(1 - g_2)a + g_1s_2b + (1 - s_1)(1 - \\ \quad g_2)c + (1 - s_1)s_2d, \\ D = g_1g_2a + g_1(1 - s_2)b + (1 - s_1)g_2c + \\ \quad (1 - s_1)(1 - s_2)d. \end{cases}$$

将这个方程组中 4 个等式两端相加可知, $A + B + C + D = a + b + c + d$. 这个方程组在完全理想情况下, 并且项目反应函数或项目参数满足或服从单调性假设, 即 $(1 - s_1) > g_1$ 且 $(1 - s_2) > g_2$, 可以由 A 、 B 、 C 和 D 和真实的 s_1 、 s_2 和 g_1 、 g_2 直接解出唯一解 a 、 b 、 c 和 d . 但是, 无论是实际数据还是模拟作答数据, 样本量的限制或模拟作答数据的随机性等都会造成 A 、 B 、 C 和 D 必然存在一定的误差, 并且采用上面方法估计的猜测和失误参数也存在一定的误差. 因此, 在下面的模拟研究中将在可行解区域内以最小二乘法来求 a 、 b 、 c 和 d 的最佳解, 其中 a 、 b 、 c 和 d 的可行解须满足各个频率位于 $0 \sim 1$ 之间.

1988 年, 有研究者^[8] 在全息项目因素分析中采用类似的思想校准作答数据中的猜测, 然后估计题目间作答数据的四分相关系数. 该方法是同时校准作答数据中的猜测和失误, 再估计题目间作答数据的四分相关系数. 四分相关的估计方法最早是由 K. Pearson^[9] 提出的, 在其提出的四分相关系数估计方法中, 公式简单且准确度较高, 应用较为广泛的一种形式是

$$Q_3 = \cos(\pi/(1 + \sqrt{ad/bc})).$$

当四分相关矩阵非正定时, 可使用最小的正特征根替换小于或等于 0 的特征根, 然后得到平滑后正定的四分相关矩阵.

2 模拟研究

2.1 研究设计

考虑可能影响 Q 矩阵标定效果的 7 个主要因素: 样本量(3 个水平)、 Q 矩阵(2 个水平)、认知诊断模型(3 个水平)、项目参数(2 个水平)、四分相关矩阵(3 个水平)、离散化方法(3 个水平)、因子旋转方法(2 个水平). 共计 $3 \times 2 \times 3 \times 2 \times 3 \times 3 \times 2 = 648$ 种影响因素或方法组合. 后面 3 个因素均是方法内因素, 将重点考察四分相关矩阵校准对 Q 矩阵标定

的影响. 将前面 4 个因素作为不同实验条件, 各种实验条件重复 30 次. 考虑 $K = 5$ 个相互独立属性, 知识状态分布采用离散均匀分布. 在此种知识状态分布下, 被试的模拟样本量分别为 300、500 和 1 000.

模拟 2 个测验 Q 矩阵, 其中一个为约简 Q 矩阵, 设为 Q_1 . Q_1 包含 1 个可达阵, 共含 31 个项目. Q_1 穷举了所有可能属性向量, 可以全面反映 Q 矩阵标定方法对不同属性向量的综合返真率. 取长度为 5 的 0-1 向量的 32 种排列组合, 去掉全零行得到的矩阵即为 Q_1 . 另一个 Q 矩阵共 20 个项目, 与文献[11] 中的 Q_3 相同, 仍记为 Q_3 . Q_3 中含有 2 个单位阵和测量 5 个属性中 2 个属性的 10 列属性向量. 整体而言, 2 种类型的测验 Q 矩阵, Q_1 较复杂而 Q_3 较简单.

在模拟研究中, 考虑 3 个认知诊断模型, 分别是 DINA 模型、DINO 模型和 R-RUM. 根据项目参数大小, 将项目质量分为高质量和低质量 2 个水平. DINA 模型和 DINO 模型下项目参数 ε 服从均匀分布, 即 $\varepsilon \sim U(0.05, 0.25)$ 或 $\varepsilon \sim U(0.05, 0.40)$, 它们分别代表这 2 个模型的高质量项目参数和低质量项目参数水平. R-RUM 下项目参数 $\pi^* \sim U(0.80, 0.98)$ 和 $r^* \sim U(0.10, 0.60)$, 或者 $\pi^* \sim U(0.75, 0.95)$ 和 $r^* \sim U(0.20, 0.95)$, 分别代表 R-RUM 的高质量项目参数和低质量项目参数水平. 根据以上模拟的知识状态、 Q 矩阵、项目参数, 可由 DINA 模型、DINO 模型和 R-RUM 分别模拟得分数据. 为了重复进行 30 次, 需要含 4 个因素的不同实验条件, 每种条件下重复模拟 30 个得分矩阵.

2.2 实验步骤

根据研究设计中的各种实验条件下所模拟的得分矩阵, 可以采用基于探索性 Q 矩阵的标定方法标定或估计 Q 矩阵. 根据得分矩阵可计算任意 2 个题目的四分相关, 由此得到四分相关矩阵. 为了考查四分相关矩阵校准对 Q 矩阵标定的影响, 四分相关矩阵分为 3 种: (i) 真实参数校准, 即采用真实的项目参数校准观察的频数再计算四分相关; (ii) 估计参数校准, 即采用估计的项目参数校准观察的频数再计算四分相关, 即采用本文 1.2 节中的方法; (iii) 不校准, 即直接采用观察的频数 A 、 B 、 C 和 D 计算 2 个题目的四分相关. 然后对以上一种类型的四分相关矩阵, 分别采用因素分析方法估计负荷矩阵. 因素负荷矩阵可分别采用因子旋转方法中斜交旋转和正交旋转得到旋转后的因素负荷矩阵, 最后分别使用负荷矩阵的行均值、列均值或全元素均值对旋转后的因素负荷矩阵进行离散化, 得出估计 Q 矩阵. 比较估计 Q 矩阵和模拟的真实 Q 矩阵, 可计算估计 Q 矩阵

的元素的平均返真率,用于评价各种实验条件下各方法所标定或估计 Q 矩阵的表现.

2.3 实验结果

表 1、表 2 和表 3 分别给出了 DINA 模型、DINO 模型和 R-RUM 中各因素或方法下 Q 矩阵元素平均返真率. 整体而言,作答数据中的噪音预处理能有效地提高了 Q 矩阵标定的返真率,新方法能够提高原探索性因素分析法的效率. 因为 DINA 和 DINO 模型中各因素或方法下 Q 矩阵元素平均返真率的变化趋势类似. 下面先分析这 2 个模型下的结果:四分相关无论是采用项目参数真值或估计值,还是未做任何校准,标定或估计的 Q 矩阵的元素返真率均随着样

本量增加而提高;随着样本量增加,估计校准的四分相关所得的返真率比未校准的返真率增幅更大;从项目质量或项目参数的影响来看,项目质量越低或猜测和失误参数越大, Q 矩阵的返真率越低,作答数据中噪音越大,校准方法的 Q 阵返真率增幅越大;从测验 Q 矩阵的影响来看, Q 矩阵越简单, Q 矩阵的返真率越高, Q 矩阵越复杂,校准方法的 Q 阵返真率增幅越大;从离散化方法的影响来看,3 种离散化方法的效果基本相当,列均值方法稍优;从因子旋转方法的影响来看,2 种旋转方法的效果基本相当,斜交方法略优;在相同样本量情况下,采用真值或估计校准的四分相关所得的返真率比未做校准更高.

表 1 DINA 模型中各因素或方法下 Q 矩阵元素平均返真率

四分相关	样本量	项目质量		测验 Q 阵		离散化方法			旋转方法		均值
		高	低	Q_1	Q_3	行均值	列均值	全均值	斜交	正交	
真值校准	300	0.930	0.831	0.792	0.969	0.877	0.884	0.881	0.884	0.877	0.880
	500	0.959	0.874	0.844	0.989	0.912	0.921	0.916	0.919	0.914	0.916
	1 000	0.983	0.939	0.924	0.998	0.954	0.966	0.963	0.961	0.962	0.961
估计校准	300	0.902	0.824	0.758	0.969	0.863	0.864	0.863	0.868	0.859	0.863
	500	0.939	0.867	0.817	0.989	0.901	0.903	0.905	0.907	0.899	0.903
	1 000	0.972	0.923	0.896	0.999	0.946	0.949	0.948	0.950	0.945	0.948
未校准	300	0.903	0.817	0.751	0.970	0.860	0.863	0.857	0.861	0.860	0.860
	500	0.921	0.835	0.772	0.984	0.881	0.879	0.874	0.874	0.882	0.878
	1 000	0.952	0.867	0.826	0.993	0.911	0.912	0.906	0.904	0.915	0.910
均值		0.940	0.864	0.820	0.985	0.901	0.905	0.901	0.903	0.901	0.902

表 2 DINO 模型中各因素或方法下 Q 矩阵元素平均返真率

四分相关	样本量	项目质量		测验 Q 阵		离散化方法			旋转方法		均值
		高	低	Q_1	Q_3	行均值	列均值	全均值	斜交	正交	
真值校准	300	0.935	0.833	0.799	0.969	0.881	0.887	0.883	0.886	0.882	0.884
	500	0.964	0.870	0.848	0.987	0.912	0.921	0.918	0.920	0.915	0.917
	1 000	0.985	0.920	0.907	0.998	0.946	0.958	0.954	0.954	0.952	0.953
估计校准	300	0.912	0.814	0.760	0.966	0.864	0.865	0.861	0.868	0.858	0.863
	500	0.945	0.868	0.827	0.986	0.906	0.906	0.907	0.909	0.904	0.906
	1 000	0.968	0.927	0.896	0.999	0.946	0.949	0.947	0.950	0.945	0.947
未校准	300	0.912	0.798	0.752	0.958	0.857	0.856	0.852	0.854	0.856	0.855
	500	0.932	0.832	0.784	0.980	0.881	0.884	0.882	0.879	0.886	0.882
	1 000	0.960	0.863	0.830	0.993	0.914	0.911	0.909	0.907	0.916	0.911
均值		0.946	0.858	0.823	0.982	0.901	0.904	0.901	0.903	0.901	0.902

下面先分析 R-RUM 中各因素或方法下 Q 矩阵元素平均返真率的结果:四分相关无论是采用项目参数真值或估计值,还是未做任何校准,标定或估计的 Q 矩阵的元素返真率均随着样本量增加而提高;在样本量较大时,估计校准的四分相关所得的返真率比未校准的返真率稍高;从项目质量或项目参数的影响来看,项目质量越低或猜测和失误参数越大, Q 矩阵的返真率越低,作答数据中噪音越大,校准方法的 Q 阵返真率增幅越大;从测验 Q 矩阵的影响来

看, Q 矩阵越简单, Q 矩阵的返真率越高, Q 矩阵越复杂,校准方法的 Q 阵返真率增幅越大;从离散化方法的影响来看,3 种离散化方法的效果相当,列均值方法稍优;从因子旋转方法的影响来看,2 种旋转方法的效果基本相当,斜交方法略优;在相同样本量情况下,采用估计校准的四分相关所得的返真率比未做校准越高.

R-RUM 下的结果与 DINA 和 DINO 模型下的结果有所差异,特别是真值校准所得的 Q 矩阵返真率

有时并不比估计校准或未校准的结果好. 这主要是因为:在 R-RUM 中,采用了完全未掌握项目所测的任何属性的正确作答概率 $\pi_j^* \prod_{k=1}^K r_{jk}^{*q_{jk}}$ 作为猜测的真值,失误采用了 $(1 - \pi_j^*)$ 作为失误的真值,而 R-RUM 的噪音不仅含有真实或所估计的猜测和失

误,而且真值校准方法对于介于完全未掌握和完全掌握之间的属性掌握模式还存在一定的噪音未处理;而估计校准方法,因为采用高低分组的试估计作答数据中的噪音在一定程度上处理了介于完全未掌握和完全掌握之间的属性掌握模式上作答反应中的噪音.

表 3 R-RUM 下各因素或方法下 Q 矩阵元素平均返真率

四分相关	样本量	项目质量		测验 Q 阵		离散化方法			旋转方法		均值
		高	低	Q ₁	Q ₃	行均值	列均值	全均值	斜交	正交	
真值校准	300	0.925	0.827	0.869	0.883	0.870	0.881	0.876	0.879	0.873	0.876
	500	0.951	0.859	0.892	0.918	0.896	0.912	0.907	0.907	0.903	0.905
	1 000	0.969	0.892	0.911	0.950	0.923	0.935	0.933	0.932	0.929	0.930
估计校准	300	0.922	0.842	0.863	0.901	0.872	0.890	0.885	0.884	0.881	0.882
	500	0.950	0.887	0.890	0.948	0.908	0.925	0.923	0.920	0.918	0.919
	1 000	0.970	0.912	0.911	0.971	0.931	0.947	0.944	0.942	0.940	0.941
未校准	300	0.925	0.845	0.867	0.903	0.874	0.891	0.889	0.882	0.887	0.885
	500	0.952	0.884	0.889	0.946	0.905	0.925	0.922	0.916	0.919	0.918
	1 000	0.968	0.903	0.905	0.966	0.926	0.942	0.938	0.936	0.934	0.935
均值		0.948	0.872	0.888	0.932	0.901	0.917	0.913	0.911	0.909	0.910

3 讨论

考虑到在实际应用中学生在做题时的猜测和失误(噪音)会影响探索性因素分析法所使用的四分相关矩阵的质量,本文先提出四分相关矩阵的一种噪音修正方法,并将其应用于 Q 矩阵标定. 模拟研究表明:猜测和失误这 2 种噪音会对 Q 矩阵的标定产生不利的影响;基于修正后的四分相关矩阵的探索性因素分析法,在样本量较大的情况下,可以有效地提高 Q 矩阵标定的返真率;特别地,在作答数据中噪音较大或测验 Q 矩阵越复杂时,基于修正后的四分相关矩阵的探索性因素分析法比未修正方法的 Q 矩阵标定的返真率幅度更明显. 在初始 Q 矩阵未知的情况下,原探索性因素分析法在 Q 矩阵的标定上存在一定的不足与提升空间,通过噪音预处理能有效地提高 Q 矩阵标定的准确率. 另外,还得到了一些与前人研究一致的结论:样本量越大,标定效果越好;Q 矩阵越简单在校准时的返真率越高,Q 矩阵越复杂返真率越低;对负荷矩阵离散化的 3 种方法的对比显示,3 种离散化方法的效果十分接近,列均值比较法稍优.

本研究仍存在许多不足之处,有待完善. 本文仅对 DINA 模型、DINO 模型和 R-RUM 产生的作答数据进行分析,有待对一般化的认知诊断模型^[12-14]产生的作答数据进行分析. 研究中提出的预处理中所用到的猜测和失误只是粗略的估计,而猜测和失误

的估计值直接影响到预处理的效果. 对于 R-RUM 或更一般化的认知诊断模型,怎样更有效而准确地通过学生的作答反应来估计作答数据中的噪音,这有待于进一步探讨. 原方法在多级评分上的推广,本研究的预处理思想仍然适用,但校准时的方程如何设立,仍需探讨. 噪音预处理的方法是否可用于已有的 Q 矩阵修正方法^[15-19],有待探讨. 使用全息项目因素分析方法、其他四分相关矩阵估计方法、负荷矩阵的其他离散化方法,结果如何,值得进一步深入研究. 对实测数据分析效果如何,也有待探讨.

4 参考文献

[1] Tatsuoka K K. Architecture of knowledge structure and cognitive diagnosis: a statistical pattern recognition and classification approach [C] //Nichols P D, Chipman S F, Brennan R L. Cognitively diagnostic assessment. Hillsdale, NJ: Erlbaum, 1995: 327-361.

[2] 刘永, 涂东波. 认知诊断测验 Q 矩阵估计方法比较 [J]. 中国考试, 2015(5): 53-63.

[3] 汪文义, 宋丽红, 丁树良. 基于探索性因素分析的 Q 矩阵标定方法 [J]. 江西师范大学学报: 自然科学版, 2015, 39(2): 138-144, 170.

[4] Junker B W, Sijtsma K. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory [J]. Applied Psychological Measurement, 2001, 25(3): 258-272.

[5] Templin J L, Henson R A. Measurement of psychological

disorders using cognitive diagnosis models [J]. Psychological Methods,2006,11(3):287-305.

[6] Hartz S M. A Bayesian framework for the unified model for assessing cognitive abilities: blending theory with practicality [D]. Champaign: University of Illinois at Urbana-Champaign,2002.

[7] Baker F B. Item response theory: parameter estimation techniques [M]. New York; Marcel Dekker, 1992: 109-110.

[8] Bock R D, Gibbons R, Muraki E. Full-information item factor analysis [J]. Applied Psychological Measurement, 1988,12(3):261-280.

[9] Kang Chunhua. Q -matrix refinement based on item fit statistic RMSEA [J]. Applied Psychological Measurement, 2019,43(7):527-542.

[10] Castellan N J. On the estimation of tetrachoric correlation coefficient [J]. Psychometrika,1966,31(1):67-73.

[11] Liu Jingchen, Xu Gongjun, Ying Zhiliang. Data-driven learning of Q -matrix [J]. Applied Psychological Measurement,2012,36(7):548-564.

[12] de la Torre J. The generalized DINA model framework [J]. Psychometrika,2011,76(2):179-199.

[13] Henson R A, Templin J L, Willse J T. Defining a family of cognitive diagnosis models using log-linear models with latent variables [J]. Psychometrika,2009,74(2):191-210.

[14] von Davier M. A general diagnostic model applied to language testing data [J]. British Journal of Mathematical and Statistical Psychology,2008,61(2):287-307.

[15] 丁树良, 罗芬,汪文义,等. Q 矩阵标定的一种简便方法 [J]. 江西师范大学学报:自然科学版,2018,42(2):130-133.

[16] 汪大勋,高旭亮,韩雨婷,等. 一种简单有效的 Q 矩阵估计方法开发:基于非参数化方法视角 [J]. 心理科学,2018,41(1):180-188.

[17] 汪大勋,高旭亮,蔡艳,等. 一种非参数化的 Q 矩阵估计方法:ICC-IR 方法开发 [J]. 心理科学,2018,41(2):466-474.

[18] 汪文义,宋丽红,丁树良. 基于可达阵的一种 Q 矩阵标定方法 [J]. 心理科学,2018,41(4):968-975.

[19] Wang Wenyi, Song Lihong, Ding Shuliang, et al. An EM-based method for Q -matrix validation [J]. Applied Psychological Measurement,2018,42(6):446-459.

The Improved Exploratory Method of Q -Matrix Specification with Noise Preprocessing

WANG Wenyi¹, GAO Peng¹, SONG Lihong², WANG Teng¹

(1. College of Computer and Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China;
2. Elementary Education Collage, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: Considering that the noises of guessing and slip that students answer problems can influence the quality of the tetrachoric correlation coefficient matrix for the exploratory factor analysis method in practical application, a noise preprocessing method of the tetrachoric correlation coefficient matrix is proposed and is used for Q -matrix specification. Simulation studies show that the noises of guessing and slip have an adverse impact on the calibration of the Q -matrix. The exploratory factor analysis method based on the modified tetrachoric correlation coefficient matrix can effectively improve the accuracy of Q -matrix specification when the sample size is relatively large and the noise is relatively high.

Key words: cognitive diagnosis; Q -matrix; exploratory factor analysis method; tetrachoric correlation correlation; data preprocessing

(责任编辑:冉小晓)