

文章编号: 1000-5862(2015)02-0138-07

基于探索性因素分析的 Q 矩阵标定方法

汪文义¹, 宋丽红², 丁树良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 330027)

摘要: 标定 Q 矩阵是认知诊断评估中最基本也是最为关键的一步。如今 Q 矩阵标定的统计方法, 多数为验证性方法, 即验证或修正已有 Q 矩阵中元素的方法。在常见的 Q 矩阵未知和已有作答数据情形下, 提出将探索性因素分析方法和验证性方法相结合的 Q 矩阵标定方法, 并采用模型整体拟合指标、分类准确性指标等, 综合确定属性数和 Q 矩阵。模拟研究表明: 新方法可较好标定 Q 矩阵。

关键词: Q 矩阵; 探索性方法; 验证性方法; 模型整体拟合指标; 分类准确性指标

中图分类号: B 841.7; TP 301.6 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2015.02.05

0 引言

教育认知诊断评估(CDA)旨在测量学生特定的知识和加工技能(简称为属性), 为学生提供认知强项和认知弱项^[1], 即报告学生掌握了哪些属性, 未掌握哪些属性^[2]。要报告学生测验分数背后所隐含的属性, 这就需要确定测验项目所考察的属性。这相当于进行试验之前要准备仪器设备, 并且要知道相应的仪器设备的基本用途, 比如要测运动员跑步的速度, 就要带计时的秒表甚至米尺等, 并且设备的确定直接决定测量的精准确度。因此, 确定测验项目所考察的属性直接影响对学生属性推断的正确性。

在认知诊断中, 测验项目所考察的属性, 用项目与属性的关联矩阵, 即用测验 $Q = (q_{jk})$ 矩阵^[3-4]表示, 其中行表示项目, 列表示属性, 如果项目 j 考察了属性 k , 则 $q_{jk} = 1$, 否则 $q_{jk} = 0$ 。大多数CDA以 Q 矩阵为基础^[5], 可以说 Q 矩阵标定对CDA是最基本也是最为关键的一步^[6-7]。

根据已有文献 Q 矩阵的构建过程大致可归纳如下3个过程: (i) 根据文献调查结果和已有相关理论研究, 由学科专家得出初始属性集。然后, 使用定性方法得出测验 Q 矩阵, 如通过任务分析, 编码项目特征并建立其与属性的关系; (ii) 结合学生口语报

告, 分析学生作答项目时所用到的属性; (iii) 学科专家对属性进行评分, 选择正确作答项目要求的属性等方法。最后, 使用定量方法验证测验 Q 矩阵。

教育认知诊断评估要真正走进课堂评估, 走近老师和学生的视野, 过多依靠专家标定 Q 矩阵不太现实: (i) 有的情况下很难由专家完成 Q 矩阵标定, 如由于年级、学科和测试的多样性, 要标定的 Q 矩阵太多, 专家标定 Q 矩阵成本更无法想象; (ii) 即使时间和经济等条件允许, 由于属性粒度、属性数和特定项目所测的属性很难把握, 学科专家正确标识测验 Q 矩阵比较难^[7-13]。

针对上述第2种情况, 目前已有研究者提出了不少统计方法, 通过对专家 Q 矩阵进行验证或修正, 以得到质量更高的 Q 矩阵^[9, 11, 14-16]。然而在第一种情况, 即几乎无法获得专家标定 Q 矩阵时, 关于获得质量较高的 Q 矩阵的方法研究相对薄弱。虽然Liu Jingchen等^[15, 17]提出了基于项目属性向量穷取法估计 Q 矩阵的标定方法, 但是此方法计算量相当大^[9]。

对于0-1评分数据, 本研究提出一种综合了探索与验证的 Q 矩阵标定方法。该方法先采用社会科学研究中广泛使用的因素分析方法, 基于作答数据探索初始 Q 矩阵, 再对初始 Q 矩阵进行验证或修正, 同时使用模型整体拟合指标、分类结果的效度指标

收稿日期: 2014-12-06

基金项目: 国家自然科学基金(31360237, 31300876, 31160203, 31100756, 30860084), 教育部人文社会科学研究青年基金(13YJC880060)和江西省教育科学2013年度一般课题(13YB032)资助项目。

作者简介: 汪文义(1983-), 男, 湖南衡山人, 讲师, 博士, 主要从事教育和统计测量方面的研究。

等综合确定属性个数并获得质量较高的Q矩阵。为了检查新方法的有效性,重点讨论以下3个方面:(i)检查因素分析方法探索到的初始Q矩阵的正确率。伴随着这个问题,有必要探讨影响因素分析效果的主要因素。如相关系数矩阵类型、因子旋转方法和负荷矩阵的离散化方法;(ii)检查并比较2种Q矩阵修正方法对初始Q矩阵标定正确率的提高幅度;(iii)检查2个指标确定属性个数和评价Q矩阵质量的效果。

1 研究方法

1.1 结合因素分析的Q矩阵标定方法

1.1.1 基于因素分析的初始Q矩阵探索 俞宗火等^[18]提到对于0-1评分数据下的因素分析方法,最好使用全息项目因素分析方法(FIFA)。FIFA是一种直接以项目反应理论为基础的探索性因素分析方法,可使用软件TESTFACT4.0估计因子负荷矩阵。但是,本文使用基于积差相关矩阵或4分相关矩阵的经典线性因素分析方法,主要原因有:(i)FIFA由于涉及多重数值积分,当潜在因子数较多时,如因子数大于7,收敛较慢;(ii)笔者通过对同一批模拟数据,使用基于经典线性因素分析方法和FIFA估计因子负荷矩阵,结果发现负荷相关达到0.8甚至更高,说明两者估计有着类似的线性变化趋势;(iii)在使用因素分析方法探索初始Q矩阵时,只是进行项目内各潜在因子负荷之间相对大小的比较,对负荷的精度要求不会太高,所以使用收敛较快的经典线性因素分析方法。

先介绍因素分析方法探索初始Q矩阵的步骤:

(i)由得分数据计算积差相关矩阵或4分相关矩阵。2个项目之间的4分相关系数的估计可由公式 $r_{tetra} = \cos(\pi^\circ / (1 + \sqrt{(AD)/(BC)}))$ 得到,其中A、B、C、D表示得分数据中所有被试在2个项目上得分模式为(0,0)、(0,1)、(1,0)、(1,1)的频数或频率。当4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,得到平滑后的4分相关矩阵,详情可参见TESTFACT4.0使用手册。

(ii)使用极大似然法估计因素负荷矩阵,并获得旋转后的因素负荷矩阵^[19]。若能得到更多数值较大的负荷和接近零的负荷,以拉开负荷之间的距离,可方便负荷矩阵离散化。因此,在使用极大似然法估计因素负荷矩阵之后,再分别采用广泛使用的方差

极大正交旋转法(varimax)或斜交旋转法(promax)^[19]获取旋转后的因素负荷矩阵。记负荷矩阵为 $A = (a_{jk})$,其中 a_{jk} 表示项目j上因素或属性k的负荷。

(iii)离散化旋转后的因素负荷矩阵,得到初始测验Q矩阵。项目所考察的属性越多,有的属性所分摊的负荷会有所减小。根据探索性结构方程模型的观点,忽略探索性因素分析中小于0.3负荷的做法并不好^[20]。因此本文提出基于项目内因素负荷相对比较的离散化方法,至少有以下3种:均值比较法,

即若 $a_{jk} \geq \sum_{k=1}^K a_{jk} / K$ (K为因子数或属性数),则令 $q_{jk} = 1$,否则 $q_{jk} = 0$;最大值法,即若 $a_{jk} = \max(a_{j1}, a_{j2}, \dots, a_{jK})$,则令 $q_{jk} = 1$,否则 $q_{jk} = 0$;大于0法,即若 $a_{jk} > 0$,则令 $q_{jk} = 1$,否则 $q_{jk} = 0$ 。

1.1.2 初始Q矩阵的验证与修正 在因素分析探索初始Q矩阵过程中,因其只基于相关系数矩阵信息,且离散化方法仍具有不确定性,因此导出的Q矩阵质量可能仍有提高的空间。为了获得更为准确的Q矩阵,有待利用被试作答矩阵对初始Q矩阵进一步验证与修正。本文选用2种表现较好的Q矩阵修正方法^[21-22]:最大似然估计(MLE)方法和交差方法。这2种方法原本是较为理想的属性在线标定方法^[8,12-13,23]。下面简要介绍用于测验Q矩阵修正的MLE方法和交差方法。

MLE方法和交差方法均可看成是对属性向量的一种条件估计方法。记 $\hat{\alpha}_i^{(t)}$ 和 $\hat{\beta}_j^{(t)}$ 为第t次EM算法估计得到的被试知识状态估计和项目参数估计值。 Q_i 表示约简Q阵,表示所有可能的属性向量。MLE方法估计项目j的 $\hat{q}_j^{(t)}$ 为

$$\hat{q}_j^{(t)} = \arg \max_{q_h \in Q_j} \prod_{i=1}^N P_{jh}(\hat{\alpha}_i^{(t)}) x_{ij} (1 - P_{jh}(\hat{\alpha}_i^{(t)}))^{1-x_{ij}},$$

其中 $P_{jh}(\hat{\alpha}_i^{(t)})$ 为特定认知诊断模型下被试i在项目j上的正确作答概率。

交差方法主要基于2点假设,在无失误和猜测情况下,项目j上作答正确的被试所掌握的属性必包含项目j所考查的属性;而项目j上作答错误的被试的知识状态及其子集必定不是 q_j 。由于样本量、失误和猜测的影响,某些知识状态答对与答错的比率可能处于伯仲之间,交差方法将优先选择答对与答错的比率或答错与答对的比率大的作交或差运算,当得到的属性向量的候选空间仅剩1个时,就不再考虑其他知识状态;当候选空间永远不止1个元素

的情况,只是简单使用约简 Q 阵中第 1 列.

MLE 和交差方法的实现只需在 EM 算法增加修正 Q 阵步骤,具体步骤如下: (i) 根据初始或更新后的测验 Q 阵,使用 EM 算法估计项目参数和被试知识状态; (ii) 对每个项目 j ($j = 1, 2, \dots, m$), 使用 MLE 方法或交差方法估计,得到更新的测验 Q 矩阵; (iii) 重复上 2 步,直到收敛准则满足为止. 收敛的终止准则采用边际对数似然函数的相对变化率小于 0.001.

1.1.3 Q 矩阵质量评价指标 为了获得最终的 Q 矩阵,主要采用 2 类指标评价新方法所得的 Q 矩阵质量. 考虑到好的 Q 矩阵既要保证拟合数据,又要保证分类诊断结果的有效性和丰富性. 因此本研究主要采用模型整体相对拟合指标和分类准确性指标. 除此之外,由于是模拟研究实验,真实的 Q 矩阵已知,还将计算测验 Q 矩阵元素的返真率. 在计算返真率时,由于 Q 矩阵列的位置会发生置换,将依次找出最匹配的列,然后计算正确标定元素的比率.

(i) 模型整体相对拟合指标: 这一类拟合指标常用于对 2 个或多个模型进行比较,从而选择 1 个拟合较优或最优的模型. 本文使用偏差 ($-2L_L$)、 A_{IC} 和 B_{IC} ^[24] 作为整体相对拟合指标,这些指标均是越小越好.

(ii) 模式或属性分类准确性指标: 在真实测验情景下,被试知识状态未知,难于获得模式或属性判断率. 因此,本文采用模式或属性分类准确性指标对其进行估计^[25],指标越大越准确.

模式分类准确性指标是基于被试知识状态后验分布计算得到. 若知识状态采用最大后验估计方法,则测验的模式分类准确性指标:

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^N \max_{c=1,2,\dots,T} (P(\alpha_c | X_i)),$$

其中 $P(\alpha_c | X_i) \propto L(X_i | \beta, \alpha_c) p(\alpha_c)$ 表示被试 i 在各知识状态 α_c 上的后验概率.

属性分类准确性指标是基于各属性的边际分布

$$\hat{p}_{ik} = \sum_{c=1}^T I(\alpha_{ck} = 1) P(\alpha_c | X_i)$$

计算得到,其中 $I(\alpha_{ck} = 1)$ 为示性函数,当 $\alpha_{ck} = 1$ 时,示性函数取值为 1,否则取值为 0. 若根据 \hat{p}_{ik} 大于或小于划界分数 0.5 估计属性掌握与否,则测验的属性分类准确性指标: $\hat{\tau}_k = \frac{1}{N} \sum_{i=1}^N \max(\hat{p}_{ik}, 1 - \hat{p}_{ik})$.

1.2 模拟设计

1.2.1 研究设计 考虑可能影响新方法效果的 8

个主要因素: 相关矩阵(2 个水平)、因子旋转方法(2 个水平)、离散化方法(3 个水平)、认知诊断模型(2 个水平)、测验 Q 矩阵(2 个水平)、项目参数水平(2 个水平)、样本量(3 个水平)、修正方法(2 个水平). 实验条件组合数共 $2 \times 2 \times 3 \times 2 \times 2 \times 2 \times 3 \times 2 = 576$.

1.2.2 数据模拟 假设 5 个相互独立的属性,知识状态分布采用离散均匀分布,分别模拟 300, 500 和 1 000 个被试. 模拟 2 个测验 Q 矩阵,1 个为约简 Q 阵,包含 1 个可达阵(单位阵),共含 31 个项目,另一个含 2 个可达阵和约简 Q 阵中考察 2 个属性的 10 个项目,共 20 个项目,后者与文献[15]中 Q_3 相同. 用于模拟得分数据的 2 个认知诊断模型,分别是确定性输入噪音与门模型(DINA 模型)和简化的再参数化统一模型(R-RUM)^[26]. DINA 模型下项目参数服从均匀分布 $U(0.05, 0.25)$ 或 $U(0.05, 0.40)$. R-RUM 下项目参数 $\pi \sim U(0.8, 0.98)$ 和 $r^* \sim U(0.1, 0.6)$, 或者 $\pi \sim U(0.75, 0.95)$ 和 $r^* \sim U(0.2, 0.95)$. 然后分别由 DINA 模型或 R-RUM 模拟生成得分矩阵. 共 24 个数据模拟条件,每条件下重复模拟 20 个得分矩阵,共生成 $24 \times 20 = 480$ 个得分矩阵.

1.2.3 分析步骤 首先由得分矩阵计算相关矩阵(积差相关矩阵或 4 分相关矩阵),设定因子数分别为 1, 2, ..., 7(为节省篇幅,部分结果中仅列出 3-7),采用极大似然法估计负荷矩阵,使用因子旋转方法(varimax 和 promax)进行因子旋转,分别使用均值比较法、最大值法和大于 0 法对旋转后的负荷矩阵进行离散化,得出初始 Q 矩阵. 然后使用带 MLE 和交差方法的 EM 算法对初始 Q 矩阵进行验证或修正,得到最终测验 Q 矩阵. 同时,计算初始 Q 矩阵的返真率,并在 EM 算法结束后计算各评价指标.

2 结果与分析

表 1 给出了因素分析不同条件下初始或最终 Q 矩阵元素返真率. 结果显示,基于积差相关矩阵、promax 斜旋转法和均值比较法组合条件下,导出的初始测验 Q 矩阵返真率最高(0.916). 为节省篇幅,下面仅对此组合条件下的实验结果作进一步分析.

表 1 因素分析不同条件下初始或最终 Q 矩阵元素返真率

| 相关阵 | 旋转法 | 离散化方法 | 初始 Q 矩阵元素返真率 | | | | 最终 Q 矩阵元素返真率 | | | |
|-------|-----|--------|----------------|-------|-------|-------|----------------|-------|-------|-------|
| | | | M | 最小值 | 最大值 | 方差 | M | 最小值 | 最大值 | 方差 |
| 积差相关 | 斜旋转 | 均值比较法 | 0.916 | 0.639 | 1.000 | 0.076 | 0.937 | 0.690 | 1.000 | 0.069 |
| | | 最大值法 | 0.788 | 0.619 | 0.900 | 0.110 | 0.897 | 0.652 | 1.000 | 0.091 |
| | | 大于 0 法 | 0.830 | 0.610 | 0.980 | 0.076 | 0.914 | 0.677 | 1.000 | 0.071 |
| | 正交法 | 均值比较法 | 0.915 | 0.645 | 1.000 | 0.078 | 0.937 | 0.671 | 1.000 | 0.070 |
| | | 最大值法 | 0.788 | 0.619 | 0.900 | 0.110 | 0.896 | 0.639 | 1.000 | 0.091 |
| | | 大于 0 法 | 0.697 | 0.490 | 0.832 | 0.068 | 0.860 | 0.590 | 1.000 | 0.085 |
| 4 分相关 | 斜旋转 | 均值比较法 | 0.910 | 0.658 | 1.000 | 0.080 | 0.935 | 0.671 | 1.000 | 0.071 |
| | | 最大值法 | 0.787 | 0.606 | 0.900 | 0.111 | 0.896 | 0.645 | 1.000 | 0.090 |
| | | 大于 0 法 | 0.818 | 0.620 | 0.981 | 0.081 | 0.908 | 0.703 | 1.000 | 0.074 |
| | 正交法 | 均值比较法 | 0.910 | 0.658 | 1.000 | 0.080 | 0.935 | 0.690 | 1.000 | 0.071 |
| | | 最大值法 | 0.787 | 0.606 | 0.900 | 0.110 | 0.897 | 0.652 | 1.000 | 0.090 |
| | | 大于 0 法 | 0.690 | 0.490 | 0.826 | 0.065 | 0.859 | 0.610 | 1.000 | 0.085 |

表 2 列出了在重复 20 次中, 各个因子上 A_{ic} 指标最小的频数($-2L_L$ 和 B_{ic} 表现稍逊于 A_{ic}). 真实属性数 5 下频数越接近 20 表示识别属性个数准确率越高. 当样本量达到 1 000 时, 能非常准确地识别出正确的属性个数. DINA 模型下, 含 2 个可达阵比含一个可达阵的 Q 矩阵更易确定正确属性个数, 而 R-RUM 模型下两者差异不太明显. 项目参数越优质, 如 DINA 猜测和失误小或 R-RUM 下 π 大或 r^* 小, 结果越准确, 对样本量要求也相应更小.

表 3 给出了 DINA 模型下各属性数下 Q 矩阵返真率. 以正确的属性个数(5 个) 的结果为例: 含一个可达阵条件下初始 Q 矩阵元素标定的返真率的均值达到 84%, MLE 和交差方法修正后的返真率均值分别为 89% 和 92%; 含 2 个可达阵条件下初始 Q 矩阵元素标定的返真率的均值达到 99%, MLE 和交差方法修正后的返真率均值几乎都为 100%.

对于 R-RUM 下属性数为 5 的 Q 矩阵返真率(结果未列出), 含一个可达阵条件下初始 Q 矩阵元素的返真率均值达到 88%, MLE 和交差方法修正后的返真率均值为 92% 和 86%. 含两个可达阵条件下初始 Q 矩阵元素标定的返真率的均值达到 95%, 修正后的返真率为 96% 和 93%. 含 2 个可达阵比含一个可达阵的 Q 矩阵, 初始 Q 矩阵返真率更高, 修正空间已经比较小. 而对于含一个可达阵的 Q 矩阵, 初始 Q 矩阵返真率相对要低一点, 这时 Q 矩阵修正方法就具有较大的修正空间.

从表 3 的 DINA 模型的分类准确性来看(R-RUM 结果类似) 相同属性个数条件下, 分类准

确性越大, 相应的 Q 矩阵返真率越高. 随着因子数或属性数增加, 分类准确性呈现逐渐下降的趋势. 在真实属性数加 1 处, 出现较大幅度的下降. 特别是在样本量为 1 000 或项目参数越优质, 下降十分明显. 有点像因子分析中碎石检验法, 根据因子对应的特征根构成的碎石图形状来判断保留因子的个数, 曲线由陡峭变平缓的前一个点被认为是提取的最大因子数.

3 讨论

3.1 结论

本文针对 Q 矩阵未知情形下, 提出结合探索性因素分析方法的 Q 矩阵标定方法. 通过模拟研究结果分析, 可以得出如下结论:

- 1) 在完全没有专家标定 Q 矩阵情形下, 因素分析方法用于探索 Q 矩阵效果较好. 在积差相关矩阵、promax 斜交旋转、均值比较法离散化组合条件下, 效果最为理想.
- 2) MLE 和交差方法对因素分析导出的初始测验 Q 矩阵进行修正, 效果较好. 其中 MLE 方法好于交差方法. 当测验 Q 矩阵含有 1 个可达阵时, 效果更明显.
- 3) 模型整体相对拟合指标可较好识别真实的潜在属性个数和正确 Q 矩阵的元素. 特别是在样本量达到 1 00 时, 表现相当不错. 分类准确性指标显示标定的测验 Q 矩阵质量不错.

表2 重复 20 次条件下 A_{IC} 指标最小的频数

| 模型 | 项目参数 | Q 阵 | N | 因子数(MLE) | | | | | | | 因子数(交叉方法) | | | | | | |
|-------|-----------------|-------|-------|-----------|---|---|----|----|---|---|------------|---|---|----|----|---|---|
| | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| DINA | $U(0.05, 0.25)$ | 1R | 300 | 0 | 0 | 0 | 4 | 13 | 2 | 1 | 0 | 0 | 0 | 0 | 18 | 2 | 0 |
| | | | 500 | 0 | 0 | 0 | 1 | 14 | 5 | 0 | 0 | 0 | 0 | 0 | 17 | 3 | 0 |
| | | | 1 000 | 0 | 0 | 0 | 0 | 18 | 2 | 0 | 0 | 0 | 0 | 0 | 19 | 1 | 0 |
| | | 2R | 300 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 0 | 0 |
| | | | 500 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 0 | 0 |
| | | | 1 000 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 0 | 0 |
| | $U(0.05, 0.40)$ | 1R | 300 | 0 | 0 | 3 | 8 | 8 | 1 | 0 | 0 | 0 | 1 | 5 | 13 | 1 | 0 |
| | | | 500 | 0 | 0 | 1 | 7 | 10 | 2 | 0 | 0 | 0 | 0 | 6 | 13 | 1 | 0 |
| | | | 1 000 | 0 | 0 | 0 | 4 | 12 | 4 | 0 | 0 | 0 | 0 | 4 | 14 | 2 | 0 |
| | | 2R | 300 | 0 | 0 | 0 | 2 | 18 | 0 | 0 | 0 | 0 | 0 | 3 | 17 | 0 | 0 |
| | | | 500 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 1 | 19 | 0 | 0 |
| | | | 1 000 | 0 | 0 | 0 | 1 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 0 | 0 |
| R-RUM | $U(0.8, 0.98)$ | 1R | 300 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 2 | 18 | 0 | 0 |
| | | | 500 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 1 | 19 | 0 | 0 |
| | | | 1 000 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 0 | 0 |
| | | 2R | 300 | 0 | 0 | 0 | 1 | 19 | 0 | 0 | 0 | 0 | 0 | 11 | 9 | 0 | 0 |
| | | | 500 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 5 | 15 | 0 | 0 |
| | | | 1 000 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 1 | 18 | 1 | 0 |
| | $U(0.75, 0.95)$ | 1R | 300 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 4 | 16 | 0 | 0 |
| | | | 500 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 1 | 19 | 0 | 0 |
| | | | 1 000 | 0 | 0 | 0 | 0 | 19 | 1 | 0 | 0 | 0 | 0 | 0 | 19 | 1 | 0 |
| | | 2R | 300 | 0 | 0 | 0 | 12 | 8 | 0 | 0 | 0 | 0 | 3 | 14 | 3 | 0 | 0 |
| | | | 500 | 0 | 0 | 0 | 1 | 19 | 0 | 0 | 0 | 0 | 0 | 9 | 11 | 0 | 0 |
| | | | 1 000 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 1 | 19 | 0 | 0 |

注: 1R 或 2R 分别表示含 1 个或 2 个可达阵的 Q 阵, N 表示样本量。

3.2 讨论

因为测验 Q 矩阵的质量高是分类准确性指标高的必要条件, 分类准确性指标可以用来衡量测验 Q 矩阵的质量。仿照碎石检验法思想, 分类准确性指标不仅可辅助用于确定属性数并选择相应的测验 Q 矩阵, 还可用于评价诊断结果的可用性。比如, 属性分类准确性指标可以供专家用来评判测验中属性的效度是否满足应用的要求, 以对属性进行相应的取舍。模型整体拟合很好, 但是分类结果不能提供充分而有效的诊断信息, 对补救和教学也是无益的。如 2 参数逻辑斯蒂克模型仍可能较好拟合作答数据, 仅提供潜在能力分数, 对补救和教学的作用也是有限的。这说明, 仅使用模型整体拟合指标并不够充分。

本研究仅对充分必要测验 Q 矩阵进行研究。在实际应用中, 怎么保证测验 Q 矩阵是充分必要的呢? 当然, 如果最后分析出来的测验 Q 矩阵含有单位阵或者通过挖掘得到可达阵, 并且由这个可达阵获得的属性层级关系能够正确表达理论的属性层级关系, 可以说 Q 矩阵是充分必要的。对于未严格按照

诊断测验设计原则进行设计的测验, 如翻新 (retrofitting) 的测验^[27], 测验 Q 矩阵要满足此要求非常难。这时, 验证性方法的表现会受到一定的影响^[28], 虽不能保证因素分析方法还能获得整体较高的返真率, 但是在某些题目上仍可能获得较高的返真率。这是因为, 对于含一定数量项目的测验 Q 矩阵, 尽管并非充分必要 Q 矩阵, 即不可能一一区分所有知识状态^[29-30], 但是, 这种测验至少仍可以较好区分其中一些知识状态, 即可保证其中一些知识状态具有较高的效度。这就可以用来确定测验中一些项目所考察的属性。知识状态效度可通过模式分类准确性指标进行评价。报告具有较高效度的被试模式, 对于诊断和补救仍然是有一定的意义。

本研究仍存在诸多不足之处, 有待完善。若采用该方法分析原本不为诊断目的而设计的标准化测验, 一定会有部分项目不符合诊断要求。如 Jang 在探讨像托福 (TOFEL) 这样的标准化测验是否可用于诊断目的时, 发现标准化测验的心理计量本质与诊断评估特征存在明显冲突。经验证据显示, 多于

20% 的测验项目不能提供诊断信息(但是多数项目仍可提供诊断信息),并且使用这些项目还会对诊断结果造成负面的影响^[31]。这时,就需要结合项目拟合绝对指标对测验项目进行筛选,如必要还可请专家对试题进行修改等。有待更多的模拟研究和实证研究去探讨这个问题。

本研究在修正方法中使用模型整体相对拟合指标和分类准确性指标对 Q 矩阵进行评价,如果不使用修正方法,可使用交叉验证的方法对初始 Q 矩阵进行评价,效果如何,值得进一步深入研究。本研究仅对 DINA 模型和 R-M 产生的作答数据进行分析,

有待对其他认知诊断模型产生的作答数据进行分析。本研究提出的方法,可推广到多级评分数据情形,因素分析方法仍可导出初始 Q 矩阵,但验证性方法得使用多级评分认知诊断模型下的相应方法,或者转化成 0-1 分数数据处理。对于符合属性之间具有层级结构关系或补偿关系,本文提出的方法是否适用,值得探讨。使用全息项目因素分析方法、其他因素旋转方法和其他负荷矩阵的离散化方法,结果如何,值得进一步深入研究。对实测数据分析效果如何,有待探讨。

表 3 DINA 模型下 Q 矩阵返真率和分类准确性

| | | | 包含 1 个可达阵的 Q 阵 | | | | | | | | 包含 2 个可达阵的 Q 阵 | | | | | | | |
|-----------------|-------|-----|------------------|-------|-------|----------|-------|-------|-------|----------|------------------|-------|-------|----------|-------|-------|-------|----------|
| | | | MLE | | | | 交差方法 | | | | MLE | | | | 交差方法 | | | |
| 项目参数 | N | 因子数 | C_1 | C_2 | P_A | M_{PA} | C_1 | C_2 | P_A | M_{PA} | C_1 | C_2 | P_A | M_{PA} | C_1 | C_2 | P_A | M_{PA} |
| $U(0.05, 0.25)$ | 300 | 3 | 0.80 | 0.86 | 0.77 | 0.91 | 0.80 | 0.82 | 0.77 | 0.91 | 0.88 | 0.87 | 0.84 | 0.94 | 0.88 | 0.81 | 0.82 | 0.93 |
| | | 4 | 0.82 | 0.89 | 0.71 | 0.90 | 0.82 | 0.93 | 0.74 | 0.91 | 0.94 | 0.94 | 0.82 | 0.95 | 0.94 | 0.93 | 0.81 | 0.94 |
| | | 5 | 0.84 | 0.92 | 0.65 | 0.89 | 0.84 | 0.98 | 0.73 | 0.92 | 1.00 | 1.00 | 0.81 | 0.95 | 1.00 | 1.00 | 0.81 | 0.95 |
| | | 6 | 0.80 | 0.86 | 0.50 | 0.85 | 0.80 | 0.92 | 0.56 | 0.88 | 0.98 | 0.99 | 0.62 | 0.92 | 0.98 | 0.98 | 0.60 | 0.91 |
| | | 7 | 0.78 | 0.84 | 0.40 | 0.82 | 0.78 | 0.90 | 0.46 | 0.86 | 0.95 | 0.98 | 0.50 | 0.89 | 0.95 | 0.98 | 0.49 | 0.89 |
| | 500 | 3 | 0.83 | 0.89 | 0.77 | 0.91 | 0.83 | 0.84 | 0.75 | 0.90 | 0.87 | 0.88 | 0.86 | 0.95 | 0.87 | 0.81 | 0.83 | 0.93 |
| | | 4 | 0.85 | 0.92 | 0.72 | 0.90 | 0.85 | 0.94 | 0.73 | 0.91 | 0.94 | 0.95 | 0.83 | 0.95 | 0.94 | 0.94 | 0.82 | 0.95 |
| | | 5 | 0.88 | 0.96 | 0.69 | 0.91 | 0.88 | 0.98 | 0.72 | 0.92 | 1.00 | 1.00 | 0.81 | 0.95 | 1.00 | 1.00 | 0.81 | 0.95 |
| | | 6 | 0.85 | 0.93 | 0.56 | 0.88 | 0.85 | 0.94 | 0.56 | 0.88 | 0.98 | 0.99 | 0.63 | 0.92 | 0.98 | 0.98 | 0.63 | 0.92 |
| | | 7 | 0.81 | 0.87 | 0.41 | 0.83 | 0.81 | 0.90 | 0.44 | 0.86 | 0.96 | 0.98 | 0.48 | 0.89 | 0.96 | 0.98 | 0.48 | 0.89 |
| | 1 000 | 3 | 0.84 | 0.90 | 0.77 | 0.91 | 0.84 | 0.83 | 0.76 | 0.91 | 0.88 | 0.89 | 0.85 | 0.95 | 0.88 | 0.82 | 0.82 | 0.93 |
| | | 4 | 0.87 | 0.93 | 0.73 | 0.91 | 0.87 | 0.94 | 0.75 | 0.92 | 0.95 | 0.95 | 0.82 | 0.95 | 0.95 | 0.94 | 0.81 | 0.95 |
| | | 5 | 0.92 | 0.98 | 0.73 | 0.92 | 0.92 | 0.99 | 0.74 | 0.93 | 1.00 | 1.00 | 0.81 | 0.95 | 1.00 | 1.00 | 0.81 | 0.95 |
| | | 6 | 0.87 | 0.94 | 0.56 | 0.88 | 0.87 | 0.95 | 0.57 | 0.89 | 0.99 | 0.99 | 0.60 | 0.91 | 0.99 | 0.99 | 0.59 | 0.91 |
| | | 7 | 0.81 | 0.87 | 0.40 | 0.82 | 0.81 | 0.90 | 0.43 | 0.85 | 0.97 | 0.98 | 0.48 | 0.89 | 0.97 | 0.98 | 0.45 | 0.88 |
| $U(0.05, 0.40)$ | 300 | 3 | 0.74 | 0.80 | 0.68 | 0.87 | 0.74 | 0.78 | 0.66 | 0.85 | 0.86 | 0.86 | 0.80 | 0.93 | 0.86 | 0.71 | 0.76 | 0.91 |
| | | 4 | 0.75 | 0.79 | 0.55 | 0.84 | 0.75 | 0.84 | 0.56 | 0.84 | 0.92 | 0.93 | 0.73 | 0.92 | 0.92 | 0.87 | 0.72 | 0.91 |
| | | 5 | 0.75 | 0.78 | 0.46 | 0.82 | 0.75 | 0.85 | 0.50 | 0.83 | 0.97 | 0.98 | 0.64 | 0.91 | 0.97 | 0.98 | 0.64 | 0.90 |
| | | 6 | 0.74 | 0.77 | 0.36 | 0.79 | 0.74 | 0.80 | 0.40 | 0.81 | 0.94 | 0.96 | 0.48 | 0.87 | 0.94 | 0.95 | 0.47 | 0.87 |
| | | 7 | 0.74 | 0.76 | 0.27 | 0.76 | 0.74 | 0.78 | 0.32 | 0.79 | 0.91 | 0.93 | 0.35 | 0.83 | 0.91 | 0.94 | 0.35 | 0.83 |
| | 500 | 3 | 0.77 | 0.81 | 0.67 | 0.86 | 0.77 | 0.79 | 0.66 | 0.85 | 0.87 | 0.87 | 0.80 | 0.93 | 0.87 | 0.70 | 0.76 | 0.91 |
| | | 4 | 0.78 | 0.82 | 0.56 | 0.84 | 0.78 | 0.86 | 0.57 | 0.84 | 0.94 | 0.94 | 0.74 | 0.92 | 0.94 | 0.87 | 0.72 | 0.92 |
| | | 5 | 0.79 | 0.83 | 0.47 | 0.82 | 0.79 | 0.90 | 0.52 | 0.84 | 0.99 | 1.00 | 0.65 | 0.91 | 0.99 | 1.00 | 0.65 | 0.91 |
| | | 6 | 0.77 | 0.80 | 0.35 | 0.79 | 0.77 | 0.84 | 0.38 | 0.80 | 0.97 | 0.98 | 0.49 | 0.87 | 0.97 | 0.97 | 0.47 | 0.87 |
| | | 7 | 0.75 | 0.76 | 0.25 | 0.75 | 0.75 | 0.79 | 0.29 | 0.77 | 0.94 | 0.96 | 0.34 | 0.83 | 0.94 | 0.95 | 0.34 | 0.83 |
| | 1 000 | 3 | 0.79 | 0.83 | 0.66 | 0.86 | 0.79 | 0.79 | 0.66 | 0.85 | 0.87 | 0.87 | 0.80 | 0.93 | 0.87 | 0.69 | 0.76 | 0.91 |
| | | 4 | 0.81 | 0.86 | 0.56 | 0.84 | 0.81 | 0.89 | 0.60 | 0.85 | 0.94 | 0.95 | 0.74 | 0.92 | 0.94 | 0.87 | 0.73 | 0.92 |
| | | 5 | 0.84 | 0.89 | 0.50 | 0.83 | 0.84 | 0.94 | 0.54 | 0.85 | 1.00 | 1.00 | 0.65 | 0.91 | 1.00 | 1.00 | 0.65 | 0.91 |
| | | 6 | 0.81 | 0.84 | 0.35 | 0.78 | 0.81 | 0.88 | 0.39 | 0.80 | 0.98 | 0.99 | 0.48 | 0.87 | 0.98 | 0.99 | 0.47 | 0.87 |
| | | 7 | 0.78 | 0.80 | 0.26 | 0.75 | 0.78 | 0.80 | 0.28 | 0.77 | 0.95 | 0.96 | 0.34 | 0.84 | 0.95 | 0.95 | 0.34 | 0.84 |

注: N 表示样本量, C_1 和 C_2 分别表示初始 Q 矩阵和修正后 Q 矩阵元素返真率, P_A 和 M_{PA} 分别表示模式和属性分类准确性。为方便比较,4 列 C_1 为重复制。

4 参考文献

- [1] 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(3): 205-237.
- [2] 戴海琦, 罗照盛. 心理测量学 [M]. 北京: 高等教育出版社 2010.
- [3] Embretson S E. A general latent trait model for response processes [J]. Psychometrika 1984 49(2): 175-186.
- [4] Tatsuoka K K. Toward an integration of item-response theory and cognitive error diagnosis [C]// Frederiksen N, Glaser R L, Lesgold A M, et al. Diagnostic Monitoring of Skill and Knowledge Acquisition [A]. NJ: Erlbaum, 1990: 453 - 488.
- [5] Huebner A. An overview of recent developments in cognitive diagnostic computer adaptive assessments [J]. Practical Assessment, Research & Evaluation 2010 15(3): 1-7.
- [6] Im S. Statistical consequences of attribute misspecification of the rule space model [D]. New York: Columbia University 2007.
- [7] McGlohen M K, Chang Huahua. Combining computer adaptive testing technology with cognitively diagnostic assessment [J]. Behavior Research Methods 2008 40(3): 808-821.
- [8] Chen Ping, Xin Tao, Wang Chun, et al. Online calibration methods for the DINA model with independent attributes in CA-CAT [J]. Psychometrika 2012 77(2): 201-222.
- [9] Chiu C Y. Statistical Refinement of the Q-Matrix in Cognitive Diagnosis [J]. Applied Psychological Measurement, 2013 37(8): 598-618.
- [10] DeCarlo L T. On the analysis of fraction subtraction data: the DINA model, classification, latent class sizes, and the Q-matrix [J]. Applied Psychological Measurement 2011, 35(1): 8-26.
- [11] DeCarlo L T. Recognizing uncertainty in the Q-matrix via a Bayesian extension of the DINA model [J]. Applied Psychological Measurement 2012 36(6): 447-468.
- [12] 陈平, 辛涛. 认知诊断计算机化自适应测验中的项目增补 [J]. 心理学报 2011 43(7): 836-850.
- [13] 汪文义, 丁树良, 游晓锋. 计算机化自适应诊断测验中原题的属性标定 [J]. 心理学报 2011 43(8): 964-976.
- [14] de la Torre J. An empirically based method of Q-matrix validation for the DINA model: development and applications [J]. Journal of Educational Measurement 2008 45(4): 343-362.
- [15] Liu Jingchen, Xu Gongjun, Ying Zhiliang. Data-driven Learning of Q-matrix [J]. Applied Psychological Measurement 2012 36(7): 548-564.
- [16] 涂冬波, 蔡艳, 戴海琦. 基于 DINA 模型的 Q 矩阵修正方法 [J]. 心理学报 2012 44(4): 558-568.
- [17] Liu Jingchen, Xu Gongjun, Ying Zhiliang. Theory of self-learning Q-matrix [J]. Bernoulli 2013 19(5A): 1790-1817.
- [18] 俞宗火, 戴海琦, 唐小娟. 全息项目因素分析在心理学研究中的应用 [J]. 心理与行为研究 2006 4(4): 306-311.
- [19] 王权. 现代因素分析 [M]. 杭州: 杭州大学出版社, 1993.
- [20] Asparouhov T, Muthén B, Muthén M. Exploratory structural equation modeling [J]. Structural Equation Modeling: A Multidisciplinary Journal 2009 16(3): 397-438.
- [21] Wang Wenyi, Ding Shuliang, Song Lihong. New Q-matrix validation methods and their sensitivity under the DINA model [C]. San Francisco: CA 2013.
- [22] 丁树良, 罗芬, 汪文义. 认知诊断分类中心的确定 [J]. 心理学探新 2013(05): 396-401.
- [23] 陈平, 辛涛. 认知诊断计算机化自适应测验中在线标定方法的开发 [J]. 心理学报 2011 43(6): 710-724.
- [24] Chen J S, de la Torre J, Zhang Zao. Relative and absolute fit evaluation in cognitive diagnosis modeling [J]. Journal of Educational Measurement 2013 50(2): 123-140.
- [25] 汪文义, 宋丽红, 丁树良等. 认知诊断测验的属性分类一致性和分类准确性指标 [EB/OL]. [2014-06-01]. [http://www. paper. edu. cn/html/releasepaper/2014/06/177/](http://www.paper.edu.cn/html/releasepaper/2014/06/177/).
- [26] Feng Yuling, Habing B T, Huebner A. Parameter estimation of the reduced RUM Using the EM algorithm [J]. Applied Psychological Measurement 2014 38(2): 137-150.
- [27] Leighton J P, Gierl M J. Cognitive diagnostic assessment for education: theory and applications [M]. New York: Cambridge University Press 2007.
- [28] 汪文义, 丁树良. 题库结构对原始题在线属性标定准确性之影响研究 [J]. 心理科学 2012 35(2): 452-456.
- [29] 丁树良, 杨淑群, 汪文义. 可达矩阵在认知诊断测验编制中的重要作用 [J]. 江西师范大学学报: 自然科学版 2010 34(5): 490-495.
- [30] 丁树良, 汪文义, 杨淑群. 认知诊断测验蓝图的设计 [J]. 心理科学 2011 34(2): 258-265.
- [31] Jang E E. A validity narrative Effects of reading skills diagnosis on teaching and learning [D]. Urbana: University of Illinois at Urbana-Champaign 2005.

(下转第 170 页)

The Cloning and Bioinformatics Analysis of Der f14 Gene from the *Dermatophagoide Farinae*

WAN Qian¹ ZHONG Zhengrong¹ LIU Zhigang^{2*}

(1. Department of Clinical Laboratory ,The First Affiliated hospital of BengbuMedical College ,Bengbu Anhui 233000 ,China;

2. Allergy and Immunology Insitute ,Shenzhen University School of Medicine ,Shenzhen Guangdong 518600 ,China)

Abstract: Genes sequence which have been synthesized from the pure culture of dust mites according to the earlier stage of our research group. The Der f14 gene fragment was amplified and cloned into pUC57 vector and then sub-cloned into the expression vector pET with restriction enzyme. The recombinant plasmid pET28a-Der f14 was constructed. Then the sequencing result was analyzed by software in ExPaSy ,and NCBI web. Results show that the opening reading frame(ORF) of Der f14 was obtained and was 5013 bp in full length. The protein encoded by this sequence was deduced to be consisted with 1 666 amino acids. Similarity of the amino acid sequence of the group 14 allergens were 95% between *dermatophagoides farinae* and *dermatophagoides pteronyssinu*. one signal peptide from 1 to 18 position. The secondary structure was composed of alpha helix(35. 83%) extended strand(17. 17%) and random coil(47%) . The gene of target protein encode a kind of vitellogenin.

Key words: *dermatophagoides farinae*; der f14; bioinformatics; clone; evolution tree

(责任编辑: 刘显亮)

(上接第 144 页)

The Statistical Specification of the Q - Matrix ——an Integration of EFA and Q - Matrix Validation Method

WANG Wenyi¹ SONG Lihong² DING Shuliang¹

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

2. Elementary Educational College ,Jiangxi Normal University ,Nanchang Jiangxi 330027 ,China)

Abstract: The vast majority of the existing statistical methods to specify the Q -matrix do rely on the draft Q -matrix constructed by the subject matter experts or the researches. Based on previous related research ,the authors consider the estimation problem of the Q -matrix under the condition of without having the draft Q -matrix. In particular ,they introduce a Q -matrix identification approach which integrates exploratory factor analysis (EFA) with Q -matrix validation method. The results of a simulation study show that the initial Q -matrix can be correctly explored with high recovery rate using the EFA and the validation methods.

Key words: the Q -matrix; exploratory method; confirmatory method; model fit index; classification accuracy index

(责任编辑: 冉小晓)