

文章编号: 1000-5862(2014)06-0600-05

项目属性标错时可达阵补救作用的研究

甘朝红, 汪文义, 丁树良

(江西师范大学计算机信息工程学院, 江西 南昌 330022)

摘要: 以 Q 矩阵理论为基础, 从分析专家给测验项目标定属性时可能出现的差错入手, 设计不同的测验 Q 矩阵和模拟实验, 分析比较在不同情况下的模式判准率, 考察可达阵对测验 Q 矩阵标定出现差错时的补救作用。

关键词: Q 矩阵; 可达阵; 模式判准率; 补救作用

中图分类号: B 841.7; TP 301.6

文献标志码: A

0 引言和记号

在认知诊断过程中, 题目属性的标定非常重要, 即 Q 矩阵的标定非常重要^[1-2]。当题目属性标定出现差错时, 可能导致属性层级关系的混乱, 且属性标定出现差错在所难免。为了提高认知诊断的准确率, 应尽量避免差错, 或设计当出错时可以弥补这些差错的预案。

当题目属性标定出现错误时, 有研究^[3-5]讨论了如何修正 Q 矩阵。本文和以往研究问题的角度不同, 不是去修改 Q 阵, 而是建立弥补的预案。已有理论和实验中表明, 可达阵 R 在认知诊断测验编制中起着重要作用^[6]。受此启发, 本文用 Matlab 作为程序设计语言编程进行模拟实验, 研究可达阵 R 对项目属性标定出错后是否有补救作用。本文从分析专家给测验项目标定属性时, 可能出现的差错入手, 研究设计不同的测验 Q 矩阵对题目属性标定出现差错的容忍能力。用 R 表示可达矩阵, Q_i 表示测验 Q 矩阵, 即测验蓝图^[7]。基于可达矩阵 R 通过扩张算法^[10]得到的矩阵称为潜在 Q 矩阵^[8-9], 记为 Q_p 。

设欲诊断的问题包含 n 个属性, 将把穷举 n 个属性的所有 0-1 列得 Q_{all} , Q_{all} 的列数为 2^n ; 记不符合属性层级关系的列构成的矩阵为 Q_{cuo} , 从 Q_{all} 的 2^n 列中去掉 0 列, 再去掉 Q_p 的所有列, 即构成错误列矩阵 Q_{cuo} ; 而在 Q_p 矩阵中由可达阵 R 扩张出的列称为扩张矩阵, 记为 Q_{kuo} 。考虑 2 类包含错误标定题目

属性的 Q_i : 1) Q_i 含可达阵; 2) Q_i 不含可达阵。测验项目标定属性时可能出现的差错, 不外乎以下 2 种情况: (i) 在专家标注的 Q_i 中, 有题目属性层级关系出错, 这种错误在属性层级关系紧密时却假设其是独立结构时容易发生; (ii) 专家标定的 Q_i 的列虽然符合属性层级关系, 但出现了误指, 即 Q_i 中的第 j 列应为 q_j , 但专家错标为 q_i , $q_j \neq q_i$, 而 q_j 和 q_i 均为 Q_p 中的列。

本文按以上 2 种差错情况进行实验, 前提假设是专家界定的属性及其层级关系正确, 只是在对具体测验项目标定属性及其层级关系时出错。

1 研究思路和研究方法

由于认知诊断测验蓝图的设计中可达矩阵有重要作用^[6], 因此在研究中分别设计项目数相同和错误列数相同的测验矩阵 Q_i , 又分为包含整个可达阵、仅仅包含可达阵某些列以及不包含可达阵任何列的情况进行比较, 考察可达阵是否能够减轻其他项目属性标注不准确引起的不良后果。再考虑错误的列数不同情况, 分别就属性层级关系类型的不同进行一一比较。

考虑 5 种基本属性层级关系(线型、收敛型、发散型、无结构型和独立型), 设置不同的 Q_i , 通过插入或者取代可达阵的若干列对诊断准确率的差异来验证可达阵对 Q_i 出现错误的弥补作用。以下考虑 5 个属性, 且属性及其层级关系见图 1。

收稿日期: 2014-09-16

基金项目: 国家自然科学基金(30860084, 31160203, 31100756, 31360237), 国家社会科学基金(12BYY055, 13BYY087) 和江西省教育厅科技计划(GJJ13208, GJJ1332006, GJJ13227, GJJ13208, GJJ13209) 资助项目。

作者简介: 甘朝红(1973-), 女, 江西丰城人, 讲师, 主要从事计算机辅助教学和计算机应用等方面的研究。

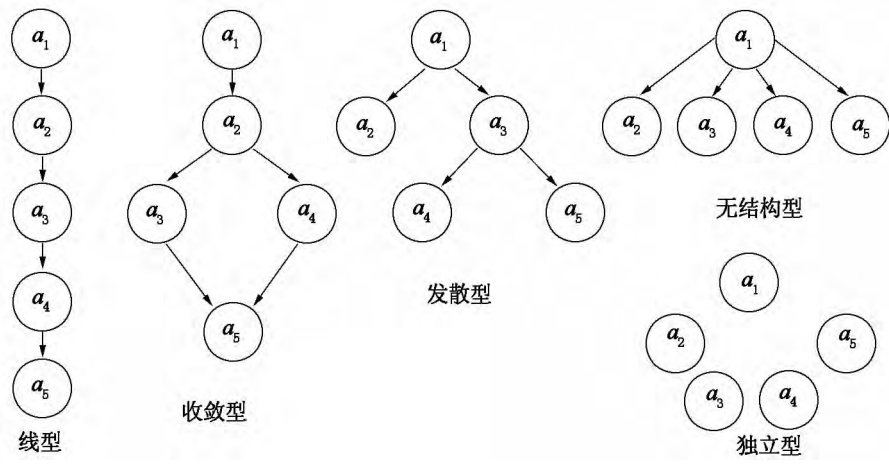


图 1 属性及其层级关系图

以下实验以 DINA 模型^[11-14]为认知诊断模型,模拟失误和猜测参数(服从 $[0.05, 0.25]$ 上的均匀分布).在每一个条件下模拟 1 000 个被试,重复做多次实验求模式判准率的平均值进行比较.整个实验过程均用 Matlab 语言编程实现.

2 出现第 1 种差错的实验及结果分析

2.1 实验设计

出现第 1 种差错情况下,设计 3 类测验矩阵 Q_i : (I) Q_i 中均为 Q_{cno} 中随机抽取的列; (II) 随机抽取 Q_{kno} 中的 1 2 3, ... 列逐步取代 (I) 中 Q_i 的第 1 2 3, ... 列; (III) 用可达阵 R 中的 1 2 3, ... 列逐步取代 (I) 中 Q_i 的第 1 2 3, ... 列,并且在同一次实验中 (I)、(II) 和 (III) 相对固定.比较 3 类测验矩阵下的模式判准率,如果情况 (III) 的模式判准率高于 (I) 和 (II),说明可达阵对项目属性标定出错后有补救作用.由于试验中存在随机行为,为减少实验误差,采用重复实验 30 次取判准率的平均值进行比较,并且求其均方差以了解各次实验结果的变动程度.

考虑到以下 2 种情况: 1) 在线型结构中, $Q_p =$

R , 即在该层级关系下 Q_{kno} 有 0 列,因而无法构造出不含可达阵的列,从而也就无法反衬出可达阵的弥补作用,所以这里不讨论线型结构; 2) 在独立结构中 Q_p 有 $2^n - 1$ 列,属性与属性之间没有先决关系,也就无法构造不符合属性层级关系的错误列,所以独立结构在第 1 种情况下不需要讨论.因此,对发生了第 1 种属性标错的情况,只讨论收敛型、发散型和无结构型 3 种层级关系下的各种情况.

2.2 实验结果及分析

2.2.1 收敛型实验结果及分析 由图 1 和扩张算法,可知 Q_p 有 6 列, Q_{kno} 只有 1 列, Q_{cno} 为 25 列,根据上述方法构造的测验 Q_i 有 8 种,记为 $Q_i(k)$ $k = 1, 2, \dots, 8$.其中 $Q_i(1)$: 从 Q_{cno} 的 25 列随机选 6 列; 用 Q_{kno} 的 1 列随机取代 $Q_i(1)$ 的 1 列,得 $Q_i(2)$; 用 R 的 h 列随机取代 $Q_i(1)$ 的 h 列,得到 $Q_i(2+h)$, $h = 1, 2, 3, 4, 5$; $Q_i(8)$: 用 Q_p 做测验项目矩阵.

表 1 为在收敛型结构下,上述 8 个测验项目矩阵分别实验 30 次时的平均模式判准率和均方差.以表 1 中平均值为数据绘制折线图如图 2 中带菱形节点的折线.

表 1 收敛型结构下的平均判准率和均方差

Q_i 分类	$Q_i(1)$	$Q_i(2)$	$Q_i(3)$	$Q_i(4)$	$Q_i(5)$	$Q_i(6)$	$Q_i(7)$	$Q_i(8)$
平均值	0.521	0.541	0.587	0.631	0.668	0.718	0.731	0.734
均方差	0.092	0.074	0.077	0.077	0.083	0.057	0.062	0.035

在表 1 中,测验矩阵 $Q_i(1)$ 均由属性层级关系有误的列构成时,平均判准率较低,当以扩展列取代其中 1 列后,平均判准率有所提升.随着 Q_i 中可达阵列数的增加,其平均判准率不断上升.图 2 中带菱形节点的折线也清楚地表明这一点.特别地,从表 1

中 $Q_i(2)$ 和 $Q_i(3)$ 看, $Q_i(3)$ 的平均判准率更高,说明测验矩阵中其他相同的情况下,含可达阵列的比含扩展列的平均判准率要高.即当出现第 1 种差错的情况时,测验项目中包含可达阵的模式判准率比测验项目中不包含或不完全包含可达阵的模式判

准确率更高. 均方差表明各次实验结果的波动差异性不大.

2.2.2 发散型实验结果 由图 1 和扩张算法, 可知 Q_p 有 10 列, 则 Q_{kuo} 有 5 列, Q_{cuo} 的列数为 21 列, 根据上述方法构造的 Q_i 有 12 种, 记为 $Q_i(k)$ $k=1, 2, \dots, 12$. 其中 $Q_i(1)$: 从 Q_{cuo} 的 21 列随机选 7 列;

用 Q_{kuo} 的 h 列随机取代 $Q_i(1)$ 的 h 列, 得 $Q_i(1+h)$, $h=1, 2, 3, 4, 5$; 用 R 的 h 列随机取代 $Q_i(1)$ 的 h 列, 得到 $Q_i(6+h)$ $h=1, 2, 3, 4, 5$; $Q_i(12)$: 用 Q_p 作 Q_i . 实验结果如表 2 所示, 以表 2 中平均值为数据绘制折线图如图 2 中带矩形节点的折线.

表 2 发散型结构下平均判准率和均方差

Q_i 分类	$Q_i(1)$	$Q_i(2)$	$Q_i(3)$	$Q_i(4)$	$Q_i(5)$	$Q_i(6)$
平均值	0.544	0.518	0.539	0.536	0.533	0.514
均方差	0.092	0.078	0.072	0.072	0.078	0.062
Q_i 分类	$Q_i(7)$	$Q_i(8)$	$Q_i(9)$	$Q_i(10)$	$Q_i(11)$	$Q_i(12)$
平均值	0.549	0.585	0.630	0.644	0.705	0.757
均方差	0.077	0.090	0.100	0.109	0.055	0.035

比较表 2 中的平均值, 其中 $Q_i(7) > Q_i(2)$, $Q_i(8) > Q_i(3)$, $Q_i(9) > Q_i(4)$, $Q_i(10) > Q_i(5)$, $Q_i(11) > Q_i(6)$, 说明相同情况下扩展列的弥补作用不如可达阵列的, 而且用可达阵的第 1 列, 2 列, \dots 去替换, 图 2 中带矩形节点的折线上扬, 表明平均判准率呈上升趋势, 可见可达阵的弥补作用明显. 均方差结果表明各次实验结果的差异性不大.

2.2.3 无结构实验结果 由图 1 和扩张算法, 可知 Q_p 有 16 列, 则 Q_{kuo} 有 11 列, Q_{cuo} 的列数为 15 列, 根据上述方法构造的 Q_i 有 12 种, 记为 $Q_i(k)$ $k=1, 2, \dots, 12$, 其中 $Q_i(1)$: 从 Q_{cuo} 的 15 列随机选 7 列; 而从 $Q_i(2)$ 到 $Q_i(12)$ 的定义和发散型中相同. 实验结果如表 3 所示, 以表 3 中平均值为数据绘制折线图如图 2 中带三角形节点的折线.

表 3 无结构平均判准率和均方差

Q_i 分类	$Q_i(1)$	$Q_i(2)$	$Q_i(3)$	$Q_i(4)$	$Q_i(5)$	$Q_i(6)$
平均	0.449	0.436	0.434	0.430	0.371	0.427
均方差	0.061	0.065	0.068	0.076	0.066	0.052
Q_i 分类	$Q_i(7)$	$Q_i(8)$	$Q_i(9)$	$Q_i(10)$	$Q_i(11)$	$Q_i(12)$
平均	0.424	0.410	0.432	0.476	0.625	0.775
均方差	0.056	0.061	0.083	0.088	0.086	0.040

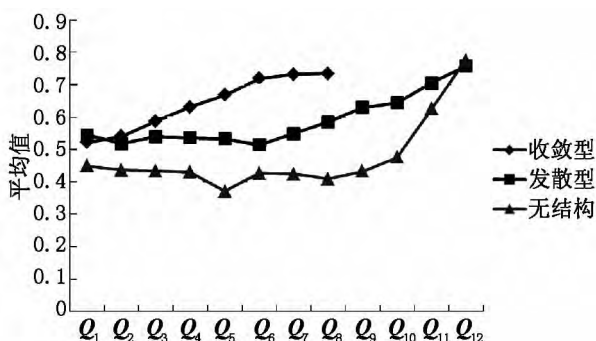


图 2 发生第 1 种差错下 Q_i 不同时平均判准率比较

从表 4、表 5 和图 2 中带三角形节点的折线看出, 属性层级关系有误的列作测验项目矩阵的情况下, 平均判准率较低, 说明出现这种差错会导致判准率下降, 用扩展列去替换其弥补作用不明显, 而用可达阵的列替换, 开始不明显, 当达到 4 列以上, 其弥补作用就比较明显. 均方差结果表明各次实验结果

的差异性不大.

综上所述: 当出现第 1 种差错的情况时, 测验项目中包含可达阵的模式判准率比测验项目中不包含或不完全包含可达阵的模式判准率更高.

3 出现第 2 种差错的实验及结果分析

3.1 实验设计

出现第 2 种差错时, 测验项目矩阵 Q_i 中的第 j 列应为 q_j , 但错标为 q_i , 其中 $i \neq j$. 而 q_j 和 q_i 都是简化关联矩阵 Q_p 中的列. 显然被试仍然按 q_j 作答, 而期望反应模式的计算却以 q_i 为准, 这必造成误判, 使得判准率不高. 为了验证可达阵是否对错误列造成的判准率下降有弥补作用, 以可达矩阵 R 为基础, 分别添加 q_j 或者 q_i , 其中添加 q_j 的测验矩阵记为 Q_{to} , 添加 q_i 的测验矩阵记为 Q_i . 先求基于 Q_{to} 的

模式判准率 PMR_5 , 而后从扩展矩阵 Q_{kuo} 中抽取 h 列, 分别取代上述 Q_{i0} 和 Q_i 中可达矩阵 R 部分的 h 列 $h=1, 2, \dots, 5$, 求各种取代后的模式判准率记为 $PMR_4, PMR_3, PMR_2, PMR_1$ 和 PMR_0 . 观察 6 种情况下判准率的变化, 随着 Q_{i0} 中可达矩阵 R 部分中的列数减少, 如果判准率呈下降趋势的话, 说明可达阵对出现的差错有弥补作用.

特别地, 为了保证实验的完备性, 首先 q_j 和 q_i 应该在简化关联矩阵 Q_p 随机取不同的 2 列; 其次用来取代 Q_{i0} 和 Q_i 的 R 中的列应该在 Q_{kuo} 中随机选取. 因此, 应该实验多次求其平均进行比较.

具体而言, 实验设计如下: 1) 从简化关联矩阵 Q_p 随机取出不同的 2 列作为 q_j 和 q_i ; 2) 令 $Q_i = [R | q_i]$ $Q_{i0} = [R | q_j]$, 求模式判准率 PMR_5 ; 3) 从 Q_{kuo}

中随机取 1 列取代 2) 中 Q_{i0} 和 Q_i 的 R 中的 1 列, 求模式判准率 PMR_{5-h} ; $h=1, 2, 3, 4, 5$; 4) 重复以上各步骤 100 次, 分别求 $PMR_5, PMR_4, PMR_3, PMR_2, PMR_1$ 和 PMR_0 的平均值进行比较.

3.2 实验及结果分析

从图 1 的各种层级关系的可达阵和简化关联矩阵来看, 线型结构没有扩张列, 收敛型结构扩张列只有 1 列, 无法按上面的实验设计思路进行实验, 可以采用其他方法实验(比如采用逐渐增加可达阵的个数), 对此本文不讨论, 而只讨论发散型、无结构型和独立型 3 种结构.

对上述 6 种情况, 每种做 100 次实验. 第 2 种差错下, 3 种结构下平均判准率的实验结果如表 4 所示, 以表 4 为数据生成图, 如图 3 所示.

表 4 发生第 2 种差错时模式判准率比较(100 次实验的平均值)

	PMR_5	PMR_4	PMR_3	PMR_2	PMR_1	PMR_0
发散型	0.593	0.556	0.510	0.493	0.446	0.397
无结构型	0.524	0.520	0.400	0.339	0.318	0.309
独立结构	0.461	0.321	0.242	0.206	0.191	0.185

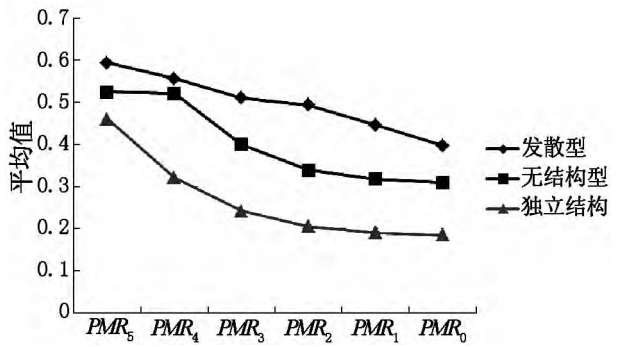


图 3 发生第 2 种差错下平均判准率比较

从表 4 和图 3 看出, 在用扩张列替代可达阵的过程中, 随着扩张列数增加, 可达阵的列数随之减少, 3 种结构下判准率都呈下降趋势. 说明 5 个属性, 错 1 列的情况下, 对项目标定出现第 2 种差错时, 含可达阵列数越多其弥补作用越明显.

3.3 进一步验证

假定测验项目矩阵 Q_i 中的第 j 列本来应该是 q_j , 但错标为 q_i , 设计 4 类 Q_i 以进一步验证可达阵对其他项目中属性错误标定的补求作用: 1) 可达阵之外再加上未标错属性的 1 列; 2) 可达阵之外再加上标错属性的 1 列; 3) 扩张部分 5 列再加上未标错属性的 1 列; 4) 扩张部分 5 列再加上标错属性的 1 列. 对上述 4 类 Q_i 分别实验求模式判准率, 对应地记为 PMR_1, PMR_2, PMR_3 和 PMR_4 . 定义 1 次差值为 D_1 和 D_2 , 其中 $D_1 = PMR_1 - PMR_2$, $D_2 = PMR_3 -$

PMR_4 . D_1 和 D_2 分别表达可达矩阵和非可达矩阵(即扩张部分)对于其他题目的属性的错误标定的补偿作用. 考虑到对试验结果的预期是可达矩阵对其他题目属性的错误标定有比较大的补偿功能, 所以 D_1 的值应该比较小, 因此再定义 2 次差值为 D , 其中 $D = D_2 - D_1$. 如果 $D > 0$, 则说明可达矩阵对于其他题目属性的错误标定有补偿功能.

实验只考虑发散型结构、无结构型和独立结构 3 种类型, 原因同上述. 以 DINA 模型为认知诊断模型, 模拟 1 000 被试, 模拟失误和猜测参数, 对上述 4 种 Q_i 分别做实验, 每次实验从简化关联矩阵 Q_p 随机取出不同的 2 列作为 q_j 和 q_i , 从 Q_{kuo} 中随机取 5 列取代可达阵构造不同的 Q_i , 一次实验求得 4 个模式判准率 PMR_1, PMR_2, PMR_3 和 PMR_4 , 继而求出模式判准率的差值 D_1 和 D_2 , 重复实验 100 次求 100 个差值的平均. 整个实验过程均用 Matlab 语言编程实现. 实验结果如表 5 所示.

表 5 3 种结构下 1 次差值和 2 次差值(100 次实验的平均值)

结构类型	D_1	D_2	$D = D_2 - D_1$
发散型	0.010 5	0.027 9	0.017 4
无结构型	0.009 9	0.019 6	0.009 7
独立结构	0.005 7	0.021 9	0.016 2

由表 5 看出, 3 种结构下, D_1 和 D_2 的值均大于 0, 说明在给项目标注属性时, 标错属性会导致模式判准率降低; 而每种结构下 D 大于 0, 即 D_1 小于 D_2 ,

说明测验矩阵中含可达阵时,属性标错时的判准率更接近正确标注时的判准率.可见可达阵在项目属性标注出错时确有弥补作用.

4 结论和展望

通过对 5 个属性 5 种属性层级关系结构的模拟实验得出:测验项目中包含可达阵的模式判准率比测验项目中不包含或不完全包含可达阵的模式判准率更高,而且可达阵对其他项目属性标定中出现的差错有一定的补救作用.这进一步验证了可达阵的在测验编制中的重要作用.如果在测验项目矩阵中加入可达阵,使各种知识状态对应不同的理想反应模式,哪怕其他项目质量稍差一些,都可以最大限度地提高诊断准确率.这表明在实际工作中,命题专家对于可达阵的列对应的题目应该精雕细刻,因为它们可以使 Q_i 有一定的稳健性.

5 参考文献

- [1] Tatsuka K K. Rule space: an approach for dealing with misconceptions based on item response theory [J]. Journal of Educational Measurement, 1983, 20(4): 345-354.
- [2] 丁树良,汪文义,罗芬.认知诊断中 Q 矩阵和 Q 矩阵理论 [J]. 江西师范大学学报:自然科学版, 2012, 36(5): 441-445.
- [3] Leighton J P, Gierl M J. Cognitive diagnostic assessment for education: theory and applications [M]. Cambridge: Cambridge University Press, 2007: 242-274.
- [4] 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): 346-362.
- [5] 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.
- [6] 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): 490-494.
- [7] 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.
- [8] 祝玉芳,丁树良.规则空间模型理论基础的改进 [J]. 江西师范大学学报:自然科学版, 2008, 32(1): 69-72.
- [9] 丁树良,祝玉芳,林海菁,等. Tatsuoka Q 矩阵理论的修正 [J]. 心理学报, 2009, 41(2): 175-181.
- [10] 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, 44(3): 87-91, 96.
- [11] Rupp A A, Templin J. The effects of Q -matrix misspecification on parameter estimates and classification accuracy in the DINA model [J]. Educational and Psychological Measurement, 2008, 68(1): 78-96.
- [12] de la Torre J. DINA model and parameter estimation: a didactic [J]. Journal of Educational and Behavioral Statistics, 2009, 34(1): 115-130.
- [13] 丁树良,汪文义,罗芬.多级评分认知诊断测验蓝图的设计:根树型结构 [J]. 江西师范大学学报:自然科学版, 2014, 38(2): 111-118.
- [14] 丁树良,罗芬,汪文义.多级评分认知诊断测验蓝图的设计:独立型和收敛型结构 [J]. 江西师范大学学报:自然科学版, 2014, 38(3): 265-269.

The Research on the Remedial Effects of Reachability Matrix When Identifying an Item Attribute Incorrectly

GAN Chao-hong, WANG Wen-yi, DING Shu-liang

(College of Computer Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: On the basis of the Q matrix theory, through analyzing the possible mistakes that the experts identify the attributes of items, different test item matrices are designed and lots of simulation experiments are made to compare the classification accuracies, in order to study if the reachability matrix has remedial effect to the Q matrix misspecification.

Key words: Q matrix; reachability matrix; classification accuracy; remedial effect

(责任编辑:冉小晓)